Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3131284 C 2023/09/19

(11)(21) 3 131 284

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 2014/08/28

(41) Mise à la disp. pub./Open to Public Insp.: 2015/03/05

(45) Date de délivrance/Issue Date: 2023/09/19

(62) Demande originale/Original Application: 2 920 899

(30) Priorité/Priority: 2013/08/28 (US61/871,219)

(51) Cl.Int./Int.Cl. *C07K 19/00* (2006.01), *C07K 14/47* (2006.01), *C12N 15/00* (2006.01), *C12N 15/09* (2006.01), *C12N 9/22* (2006.01)

(72) Inventeurs/Inventors:
PASCHON, DAVID, US;
ZHANG, LEI, US

(73) Propriétaire/Owner:
SANGAMO THERAPEUTICS, INC., US

(74) Agent: KIRBY EADES GALE BAKER

(54) Titre : COMPOSITIONS DE LIAISON DE DOMAINES DE LIAISON A L'ADN ET DE DOMAINES DE CLIVAGE
(54) Title: COMPOSITIONS FOR LINKING DNA-BINDING DOMAINS AND CLEAVAGE DOMAINS

(57) Abrégé/Abstract:
Disclosed herein are compositions for linking DNA-binding domains and cleavage domains to form nucleases, for example nucleases with altered target site separation (gap) preferences as compared to conventional linkers. Also described are fusion proteins comprising these linkers. The disclosure also provides methods of using these fusion proteins and compositions thereof for targeted cleavage of cellular chromatin in a region of interest and/or homologous recombination at a predetermined region of interest in cells. Thus, in one aspect, described herein are amino acid sequences linking a DNA-binding domain to a cleavage domain. The linker may replace one or more residues of the DNA-binding domain and/or cleavage domain.

ABSTRACT

Disclosed herein are compositions for linking DNA-binding domains and cleavage domains to form nucleases, for example nucleases with altered target site separation (gap) preferences as compared to conventional linkers. Also described are fusion proteins comprising these linkers. The disclosure also provides methods of using these fusion proteins and compositions thereof for targeted cleavage of cellular chromatin in a region of interest and/or homologous recombination at a predetermined region of interest in cells. Thus, in one aspect, described herein are amino acid sequences linking a DNA-binding domain to a cleavage domain. The linker may replace one or more residues of the DNA-binding domain and/or cleavage domain.

Date Reçue/Date Received 2021-09-17

COMPOSITIONS FOR LINKING DNA-BINDING DOMAINS AND CLEAVAGE DOMAINS

[0001] This application is a divisional application divided from Canadian Patent Application 2,920,899, which is the national phase application from International Patent Application PCT/US2014/053170 filed internationally on August 28, 2014 and published as WO/2015/031619 on March 5, 2015.

[0002] Not applicable.

TECHNICAL FIELD

[0003] The present disclosure is in the fields of genome and protein engineering.

BACKGROUND

[0004] Artificial nucleases, such as engineered zinc finger nucleases (ZFN), transcription-activator like effector nucleases (TALENs), the CRISPR/Cas system with an engineered crRNA/tracr RNA ('single guide RNA') and/or nucleases based on the Argonaute system (*e.g.*, from *T. thermophilus*, known as 'TtAgo', (Swarts *et al* (2014) *Nature* 507(7491): 258-261), comprising DNA binding domains (nucleotide or polypeptide) operably linked to cleavage domains have been used for targeted alteration of genomic sequences. For example, zinc finger nucleases have been used to insert exogenous sequences, inactivate one or more endogenous genes, create organisms (*e.g.*, crops) and cell lines with altered gene expression patterns, and the like. *See, e.g.*, U.S. Patent Nos. 8,586,526; 8,329,986; 8,399,218; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,067,317; 7,262,054; 7,888,121; 7,972,854; 7,914,796; 7,951,925; 8,110,379; 8,409,861; U.S. Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060063231; 20080159996; 201000218264; 20120017290; 20110265198; 20130137104; 20130122591; 20130177983 and 20130177960 and U.S. Application No. 14/278,903. For instance, a pair of nucleases (*e.g.*, zinc finger nucleases, TALENs) is typically used to cleave genomic sequences. Each member of the pair generally includes an engineered (non-naturally occurring) DNA-binding protein linked to one or more cleavage domains (or half-domains) of a nuclease. When the DNA-binding proteins bind to their target

Date Reçue/Date Received 2021-09-17

sites, the cleavage domains that are linked to those DNA binding proteins are positioned such that dimerization and subsequent cleavage of the genome can occur, generally between the pair of the zinc finger nucleases or TALENs.

[0005] It has been shown that cleavage activity of the nuclease pair is related to the length of the linker joining the zinc finger and the cleavage domain ("ZC" linker), the amino acid composition, and the distance between the target sites (binding sites). See, for example, U.S. Patent Nos. 8,772,453; 7,888,121 and 8,409,861; Smith *et al.* (2000) *Nucleic Acids Res.* 28:3361-3369; Bibikova *et al.* (2001) *Mol. Cell. Biol.* 21:289-297. When using pairs of nuclease fusion proteins, optimal cleavage with currently available ZC linkers and cleavage half domains has been obtained when the binding sites for the fusion proteins are located 5 or 6 nucleotides apart (as measured from the near edge of each binding site). See, *e.g.*, U.S. Patent No. 7,888,121. U.S. Patent Publication 20090305419 describes linking DNA-binding domains and cleavage domains by using a ZC linker and modifying the N-terminal residues of the *Fok*I cleavage domain.

[0006] Thus, there remains a need for methods and compositions that allow targeted modification where the artificial nucleases can cleave endogenous genomic sequences with binding site separations other than 5 bp or 6 bp. The ability to target sequences with different spacings would increase the number of genomic targets that can be cleaved. Altering the preferences between target sites separated by different numbers of base pairs could also allow the artificial nucleases to act with greater specificity.

SUMMARY

[0006a] Certain exemplary embodiments provide a fusion protein comprising: a DNA-binding domain having an N-terminus and a C-terminus, wherein the DNA-binding domain binds to a nucleotide target site; a truncated *FokI* cleavage domain in which the N-terminal residues QLVKS are deleted; and a linker between the C-terminus of the DNA-binding domain and the N-terminus of the cleavage domain, wherein the linker is a sequence selected from the group consisting of (N)GICPPPRPRTSPP (L8a); (T)GTAPIEIPPEVYP (L8b); (N)GSYAPMPPLALASP (L8c); (P)GIYTAPTSRPTVPP (L8d); (N)GSQTPKRFQPTHPSA (L8e); TGLMPPSHPRQPIHINF (L8g); TGTVHTSPICPQTYP (L8i); TGSGTPPRPHPPLPP (L8j); (H)LPKPANPFPLD (L7-1); (H)RDGPRNLPPTSPP

Date Reçue/Date Received 2021-09-17

(L7-2); (H)RLPDSPTALAPDTL (L7-6); (D)PNSPISRARPLNPHP (L7-3); (Y)GPRPTPRLRCPIDSLIFR (L7-5); (H)CPASRPIHP (L6-2); (G)LQSLIPQQLL (L6-6); (G)LQPTVNHEYNN (L6-7); and (P)ANIHSLSSPPPL (L6-1); and further wherein the amino acid residue shown in round parentheses is optionally present.

[0006b] Other exemplary embodiments provide a fusion protein comprising: a DNA-binding domain having an N-terminus and a C-terminus, wherein the DNA-binding domain binds to a nucleotide target site; a *FokI* cleavage domain having an N-terminus and a C-terminus, wherein the N-terminus residues of the *FokI* cleavage domain are ELEEK; and a linker between the C-terminus of the DNA-binding domain and the N-terminus of the cleavage domain, wherein the linker comprises LRGSPISRARPLNPHP; LRGSISRARPLNPHP; LRGSPSRARPLNPHP; LRGSSRARPLNPHP; LRGSYAPMPPLALASP; LRGSAPMPPLALASP; LRGSPMPPLALASP; or LRGSMPPLALASP.

[0007] Disclosed herein are compositions for linking DNA-binding domains and cleavage domains to form nucleases, for example nucleases with altered target site separation (gap) preferences as compared to conventional linkers. Also described are fusion proteins comprising these linkers. The disclosure also provides methods of using these fusion proteins and compositions thereof for targeted cleavage of cellular

Date Reçue/Date Received 2021-09-17

chromatin in a region of interest and/or homologous recombination at a predetermined region of interest in cells.

[0008] Thus, in one aspect, described herein are amino acid sequences linking a DNA-binding domain (*e.g.*, zinc finger protein or TAL-effector domain) to a cleavage domain (*e.g.*, wild type or engineered *Fok*I cleavage domain). In certain embodiments, the amino acid linker sequences extend between the last residue of the N- or C-terminal of the DNA-binding domain and the N- or C-terminal of the cleavage domain. In other embodiments, the linker may replace one or more residues of the DNA-binding domain and/or cleavage domain. In certain embodiments, the linker is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 or more residues in length. In other embodiments, the linker between the DNA-binding domain and the cleavage domain (or cleavage half-domain) comprises any of the linkers shown in Figures 4, 7, 9, 11, 12, 13, 15 and 16 (SEQ ID NOs:2-21, 45-49, 51, 62-226, 233-240, and 258-312) (with or without 1 or more of the N-terminal amino acid residues shown in these Figures), including but not limited to (N)GICPPPRPRTSPP (SEQ ID NO:2); (T)GTAPIEIPPEVYP (SEQ ID NO:3); (N)GSYAPMPPLALASP (SEQ ID NO:4); (P)GIYTAPTSRPTVPP (SEQ ID NO:5) ; (N)GSQTPKRFQPTHPSA (SEQ ID NO:6); (H)LPKPANPFPLD (SEQ ID NO:7); (H)RDGPRNLPPTSPP (SEQ ID NO:8); (H)RLPDSPTALAPDTL (SEQ ID NO:9); (D)PNSPISRARPLNPHP (SEQ ID NO:10); (Y)GPRPTPRLRCPIDSLIFR (SEQ ID NO:11); (H)CPASRPIHP (SEQ ID NO:12); (G)LQSLIPQQLL (SEQ ID NO:13); (G)LQPTVNHEYNN (SEQ ID NO:14); (P)ANIHSLSSPPPL (SEQ ID NO:15); (P)AGLNTPCSPRSRSN (SEQ ID NO:16); (A)TITDPNP(SEQ ID NO:17); (P)PHKGLLP (SEQ ID NO:18); (S)VSLPDTHH(SEQ ID NO:19); (G)THGATPTHSP (SEQ ID NO:20); (V)APGESSMTSL(SEQ ID NO:21), (LRGS)PISRARPLNPHP (SEQ ID NO:22), (LRGS)ISRARPLNPHP (SEQ ID NO:23), (LRGS)PSRARPLNPHP (SEQ ID NO:24), (LRGS)SRARPLNPHP (SEQ ID NO:25), (LRGS)YAPMPPLALASP (SEQ ID NO:26), (LRGS)APMPPLALASP (SEQ ID NO:27), (LRGS)PMPPLALASP (SEQ ID NO:28), (LRGS)MPPLALASP (SEQ ID NO:29) and HLPKPANPFPLD (SEQ ID NO:30), wherein the amino acid residue(s) shown in round parentheses is(are) optionally present. In certain embodiments, the linker comprises a sequence selected from the group consisting of PKPAN (SEQ ID NO:31), RARPLN (SEQ ID NO:32); PMPPLA (SEQ ID NO:33) or PPPRP (SEQ ID NO:34). In certain embodiments, the linkers as described herein further comprise a ZC linker sequence at

Date Reçue/Date Received 2021-09-17

their N-terminal ends. In other embodiments, the linker is includes modifications at the junction with the cleavage domain (FokI), for example as shown in Figures 4, 7, 9, 11, 12, 13, 15 and 16.

[0009] In still further aspects, described herein is a fusion protein comprising a DNA-binding domain, a modified FokI cleavage domain and a ZC linker between the DNA-binding domain and the FokI cleavage domain. The FokI cleavage domain may be modified in any way, including addition, deletion and/or substitution of one or more amino acids residues. In certain embodiments, the modifications to the FokI cleavage domain comprises one or more additions, deletions and/or substitutions to the N-terminal region of FokI (residues 158-169 of SEQ ID NO:1), including, for example, addition, deletion and/or substitution of 1, 2, 3, 4 amino acid residues. In certain embodiments, the modified FokI cleavage domain comprises deletions of 1, 2, 3, 4 or more amino acids from the N-terminal region of FokI (*e.g.*, deletion of one or more of residues 158, 159, 160 and/or 161 of the wild-type FokI domain of SEQ ID NO:1). In other embodiments, the modified FokI cleavage domain comprises one or more deletions from, and one or more substitutions within, the N-terminal region of FokI (*e.g.*, deletion of one or more of residues 158-161 and substitution of one or more of the remaining residues). In other embodiments, the modified FokI cleavage domain comprises one or more substitutions in the N-terminal region of FokI. In still further embodiments, the modified FokI cleavage domain comprises one or more additional amino acid residues (*e.g.*, 1, 2, 3, 4 or more) N-terminal to the N-terminal-most residue of FokI (residue 158 of SEQ ID NO:1). In other embodiments, the modified FokI cleavage domain comprises one or more additional amino acid residues (*e.g.*, 1, 2, 3, 4 or more) N-terminal to the N-terminal-most residue of FokI (residue 158 of SEQ ID NO:1) and/or one or more substitutions within the N-terminal region of FokI. In certain embodiments, the fusion protein comprises a modified FokI cleavage domain as shown in Figure 15 or Figure 16. In other embodiments, the fusion protein comprises a modified FokI cleavage domain as shown in Figure 15 or Figure 16 and further comprising one or more modifications within the N-terminal as shown in U.S. Patent Publication No. 20090305419.

[0010] In another aspect, described herein is a dimer comprising at least two fusion proteins, each fusion protein comprising a DNA-binding domain, linker and cleavage domain. In certain embodiments, at least one fusion protein comprises a linker as described herein. In other embodiments, both fusion proteins comprise a

Date Reçue/Date Received 2021-09-17

linker as described herein. In still further embodiments, the DNA-binding domains of the dimer target sequences (*e.g.*, in double-stranded DNA) separated by 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 (or more) base pairs, preferably 5 to 8 base pairs. Any of the fusion proteins may comprise wild-type or engineered cleavage domains.

[0011] In another aspect, fusion polypeptides comprising a DNA-binding domain (*e.g.*, a zinc finger or TALE DNA-binding domain), a cleavage half-domain and a linker as described herein are provided.

[0012] In another aspect, polynucleotides encoding any of the linkers or fusion proteins as described herein are provided. In some embodiments, the polynucleotides are RNAs.

[0013] In yet another aspect, cells comprising any of the polypeptides (*e.g.*, fusion polypeptides) and/or polynucleotides as described herein are also provided. In one embodiment, the cells comprise a pair of fusion polypeptides, each comprising a cleavage domain as disclosed herein.

[0014] In yet another aspect, methods for targeted cleavage of cellular chromatin in a region of interest; methods of causing homologous recombination to occur in a cell; methods of treating infection; and/or methods of treating disease are provided. The methods involve cleaving cellular chromatin at a predetermined region of interest in cells by expressing a pair of fusion polypeptides, at least one of which comprises a linker as described herein. In certain embodiments, one fusion polypeptide comprises a linker as described herein and in other embodiments, both fusion polypeptides comprise a linker as described herein. Furthermore, in any of the methods described herein, the pair of fusion polypeptides cleaves the targeted region when the binding sites for the zinc finger nucleases are 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or even more base pairs apart.

[0015] The polypeptides comprising the linkers as described herein can be used in methods for targeted cleavage of cellular chromatin in a region of interest and/or homologous recombination at a predetermined region of interest in cells. Cells include cultured cells, cells in an organism and cells that have been removed from an organism for treatment in cases where the cells and/or their descendants will be returned to the organism after treatment. A region of interest in cellular chromatin can be, for example, a genomic sequence or portion thereof.

[0016] A fusion protein can be expressed in a cell, *e.g.*, by delivering the fusion protein to the cell or by delivering a polynucleotide encoding the fusion protein

Date Reçue/Date Received 2021-09-17

to a cell, wherein the polynucleotide, if DNA, is transcribed, and an RNA molecule delivered to the cell or a transcript of a DNA molecule delivered to the cell is translated, to generate the fusion protein. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0017] Accordingly, in another aspect, a method for cleaving cellular chromatin in a region of interest can comprise (a) selecting a first sequence in the region of interest; (b) engineering a first DNA-binding domain to bind to the first sequence; (c) expressing a first fusion protein in the cell, the first fusion protein comprising the first DNA-binding domain (*e.g.*, zinc finger or TALE), and a cleavage half-domain; and (d) expressing a second fusion protein in the cell, the second fusion protein comprising a second DNA-binding domain, and a second cleavage half-domain, wherein at least one of the fusion proteins comprises a linker as described herein, and further wherein the first fusion protein binds to the first sequence, and the second fusion protein binds to a second sequence located between 2 and 50 nucleotides from the first sequence, such that cellular chromatin is cleaved in the region of interest. In certain embodiments, both fusion proteins comprise a linker as described herein.

[0018] In other embodiments, the disclosure provides methods of cleaving cellular chromatin by introducing one more nucleases comprising a linker as described herein into a cell such that the nucleases target and cleave the cellular chromatin of the cell. The nuclease may comprise a zinc finger nuclease (ZFN), a TALE-nuclease (TALEN), TtAgo or a CRISPR/Cas nuclease system or a combination thereof. The nuclease(s) may be introduced into the cell in any form, for example in protein form, in mRNA form or carried on a viral (AAV, IDLV, etc.) vector or non-viral vector (*e.g.*, plasmid). In certain embodiments, the methods comprise (a) selecting first and second sequences in a region of interest, wherein the first and second sequences are between 2 and 50 nucleotides apart; (b) engineering a first DNA-binding domain to bind to the first sequence; (c) engineering a second DNA-binding domain to bind to the second sequence; (d) expressing a first fusion protein in the cell, the first fusion protein comprising the first engineered DNA-binding domain, a first linker as described herein, and a first cleavage half domain; (e) expressing a second fusion protein in the cell, the second fusion protein comprising the second engineered DNA-binding domain, a second linker and a second cleavage half-domain; wherein the first fusion protein binds to the first sequence and the

Date Reçue/Date Received 2021-09-17

second fusion protein binds to the second sequence, thereby cleaving the cellular chromatin in the region of interest. In certain embodiments, the first and second fusion proteins comprise a linker as described herein.

[0019] Also provided are methods of altering a region of cellular chromatin, for example to introduce targeted mutations. In certain embodiments, methods of altering cellular chromatin comprise introducing into the cell one or more targeted nucleases to create a double-stranded break in cellular chromatin at a predetermined site, and a donor polynucleotide, having homology to the nucleotide sequence of the cellular chromatin in the region of the break. Cellular DNA repair processes are activated by the presence of the double-stranded break and the donor polynucleotide is used as a template for repair of the break, resulting in the introduction of all or part of the nucleotide sequence of the donor into the cellular chromatin. Thus, a sequence in cellular chromatin can be altered and, in certain embodiments, can be converted into a sequence present in a donor polynucleotide.

[0020] Targeted alterations include, but are not limited to, point mutations (*i.e.*, conversion of a single base pair to a different base pair), substitutions (*i.e.*, conversion of a plurality of base pairs to a different sequence of identical length), insertions or one or more base pairs, deletions of one or more base pairs and any combination of the aforementioned sequence alterations.

[0021] The donor polynucleotide can be DNA or RNA, can be linear or circular, and can be single-stranded or double-stranded. It can be delivered to the cell as naked nucleic acid, as a complex with one or more delivery agents (*e.g.*, liposomes, poloxamers) or contained in a viral delivery vehicle, such as, for example, an adenovirus or an adeno-associated Virus (AAV). Donor sequences can range in length from 10 to 1,000 nucleotides (or any integral value of nucleotides therebetween) or longer. In some embodiments, the donor comprises a full length gene flanked by regions of homology with the targeted cleavage site. In some embodiments, the donor lacks homologous regions and is integrated into a target locus through homology independent mechanism (*i.e.* NHEJ). In other embodiments, the donor comprises a smaller piece of nucleic acid flanked by homologous regions for use in the cell (*i.e.* for gene correction). In some embodiments, the donor comprises a gene encoding a functional or structural component such as a shRNA, RNAi, miRNA or the like. In other embodiments the donor comprises sequences encoding a regulatory element that binds to and/or modulates expression of a gene of

Date Reçue/Date Received 2021-09-17

interest. In other embodiments, the donor is a regulatory protein of interest (*e.g.* ZFP TFs, TALE TFs or a CRISPR/Cas TF) that binds to and/or modulates expression of a gene of interest.

[0022] In certain embodiments, the frequency of homologous recombination can be enhanced by arresting the cells in the G2 phase of the cell cycle and/or by activating the expression of one or more molecules (protein, RNA) involved in homologous recombination and/or by inhibiting the expression or activity of proteins involved in non-homologous end-joining.

[0023] In any of the methods described herein in which a pair of nucleases is used, the first and second nucleases of the nuclease pair can bind to target sites 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more base pairs apart. In addition, in any of the methods, the second zinc finger binding domain may be engineered to bind to the second sequence.

[0024] Furthermore, in any of the methods described herein, the fusion proteins may be encoded by a single polynucleotide.

[0025] For any of the aforementioned methods, the cellular chromatin can be in a chromosome, episome or organellar genome. Cellular chromatin can be present in any type of cell including, but not limited to, prokaryotic and eukaryotic cells, fungal cells, plant cells, animal cells, mammalian cells, primate cells and human cells.

[0026] In another aspect, described herein is a kit comprising a linker as described herein or a polynucleotide encoding a linker as described herein; ancillary reagents; and optionally instructions and suitable containers. The kit may also include one or more nucleases or polynucleotides encoding such nucleases.

[0027] In any of the proteins, methods and kits described herein, the cleavage domain (or cleavage half-domain) may comprise a TypeIIS cleavage domain, such as a cleavage half-domain from *Fok*I.

[0028] These and other aspects will be readily apparent to the skilled artisan in light of disclosure as a whole.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] **Figure 1** depicts the sequence of an exemplary zinc finger nuclease that binds to a target site in CCR5 (SEQ ID NO:1). The zinc finger DNA binding domain is doubly underlined. The entire *Fok*I cleavage domain is underlined and the

Date Reçue/Date Received 2021-09-17

N-terminal region is underlined and bolded. A "ZC" linker (LRGS; SEQ ID NO:35) is shown in plain text between the zinc finger and cleavage domains.

[0030] **Figures 2A and B** depict library design and selection. Figure 2A shows host ZFN used in the bacterial selection assay (Example 3, SEQ ID NOs:52-55, respectively, in order of appearance), indicating the location of the randomized linker library. The zinc finger protein domain used for selections included the recognition helices regions of ZFP 8196 as depicted in U.S. Patent No. 7,951,925, which binds to the CCR5 gene. Figure 2B is a schematic showing an overview of the bacterial system used for selection of linkers.

[0031] **Figure 3** shows an overview of the target sequences (including indicated spacings between target sites) used in the bacterial ccdB toxin plasmid of the bacterial selection system (SEQ ID NOs:56-61, respectively, in order of appearance).

[0032] **Figure 4** depicts exemplary linkers obtained from bacterial selections using libraries of 8 to 17 amino acid long linkers for the indicated target spacings (SEQ ID NOs:62-226, respectively, in order of appearance (top to bottom, left to right)). "x2" and "x7" indicate the number of times the sequence was found in the screen.

[0033] **Figure 5** shows the distribution of the linker length (8 to 17 amino acids) relative to target spacing (5 to 16 base pairs) after selection.

[0034] **Figure 6** depicts an overview of the target sites for the modified CCR5 locus in mammalian cells, including target sequences and spacings (SEQ ID NOs:227-232, respectively, in order of appearance).

[0035] **Figure 7** depicts results of ZFN cleavage with an 8 base pair gap between target sites using the indicated linkers (SEQ ID NOs:233, 2-6, 234, 161 and 235, respectively, in order of appearance).

[0036] **Figure 8** is a summary of results obtained with all linkers tested for the indicated target site gaps.

[0037] **Figures 9A to C,** show ZFN activity and dimer gap preference results. Figures 9A and 9B show the results of ZFN-modifications ("indels") obtained with the indicated linkers (SEQ ID NOs:233, 2-6, 234, 161 and 235, respectively, in order of appearance) at the indicated gaps between the paired target sites. Figure 9C shows dimer gap preferences of the indicated linkers.

Date Reçue/Date Received 2021-09-17

[0038] **Figures 10 A and B** show the design, assembly scheme, and results of portability studies conducted with various linkers. Figure 10A shows the design of the vectors used, including nucleotide sequences(SEQ ID NOs:241, 243, 245, 247, 249, 251, 253, 257 and 255, respectively, in order of appearance) and amino acid sequences (SEQ ID NOs:242, 244, 246, 248, 250, 252, 254, 256 and 256, respectively, in order of appearance). Solid blocks labeled 'A', 'B' and 'C'' indicate zinc finger modules. Figure 10B depicts a summary of the results obtained.

[0039] **Figures 11A and B** show the activity of linkers selected to recognize an 8 or 9 bp gap between the target sites. Figure 11A depicts gels showing the Cel-I results for 8 pairs of ZFNs modified by 4 different linkers. Lanes are numbered according to the Group shown at the bottom of the figure. GFP indicates the GFP expression vector control. An 'X' over the lane indicates a faulty run. Figure 11B indicates the percent NHEJ activity from the Cel 1 gels in Figure 11A. Figures 11A and B disclose SEQ ID NOs:45 and 45-49, respectively, in order of appearance.

[0040] **Figure 12** shows activity, as determined by Cel1 assay and sequence analysis, of the indicated linkers in the indicated zinc finger and cleavage domain vectors (SEQ ID NOs:258-261, 258-261 and 262-270, respectively, in order of appearance). As shown the linker sequence extends from the amino acid residue immediately C-terminal to the 2nd histidine residue of the ZFP sequence to the amino acid residue immediately N-terminal to the EL FokI sequence. Arrows indicate the two most active 7 bp gap linker sequences.

[0041] **Figure 13** shows activity of the indicated shows activity (%NHEJ) of the indicated linkers in the indicated zinc finger and cleavage domain vectors (SEQ ID NOs:271-279, respectively, in order of appearance). As shown the linker sequence extends from the amino acid residue immediately C-terminal to the 2nd histidine residue of the ZFP sequence to the amino acid residue immediately N-terminal to the EL FokI sequence.

[0042] **Figure 14** is a schematic illustrating dimerization (*e.g.*, heterodimerization) of FokI cleavage domains upon binding of ZFNs to their target sites with the indicated gaps.

[0043] **Figure 15** shows activity (%NHEJ) of ZFNs including linkers as described herein and including modified junctions as between the linker and cleavage domain (in the left and/or right ZFNs of the pair as indicated) (SEQ ID NOs:280-303 and 51, respectively, in order of appearance).

Date Reçue/Date Received 2021-09-17

[0044] **Figure 16** shows activity (%NHEJ) of different ZFNs including linkers as described herein and including modified junctions as between the linker and cleavage domain (in the left and/or right ZFNs of the pair as indicated). The linker portion of the sequences extends from the amino acid residue immediately C-terminal to the 2nd histidine residue of the ZFP sequence to the amino acid residue immediately N-terminal to the EL FokI sequence. The activities of the ZFN pairs with both conventional junction sequences are boxed. Figure 16 discloses "GSKS," "GSVKS," "GSEVKS," "GSQSVKS," "GSKQLVKS," "GSGKQLVKS," "GSVTKQLVKS," "GSQLVKS" and "GSTKQLVKS" as SEQ ID NOs304-312, respectively.

DETAILED DESCRIPTION

[0045] Described herein are compositions for linking DNA-binding domains and cleavage domains to form artificial nucleases and methods of using these nucleases for targeted alteration of a cellular nucleotide sequence, *e.g.*, by targeted cleavage followed by non-homologous end joining; by targeted cleavage followed by homologous recombination between an exogenous polynucleotide (comprising one or more regions of homology with the cellular nucleotide sequence) and a genomic sequence; by targeted inactivation of one or more endogenous genes.

[0046] Exemplary linkers as shown Figures 4 and 9 increase the ability of a pair of ZFNs to cleave when the ZFN target sites are more than 5 or 6 base pairs apart. Thus, certain linkers described herein significantly increase the ability to perform targeted genomic alteration by increasing the cleavage activity when the zinc finger target sites are not separated by 5 or 6 base pairs.

General

[0047] Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See*, for example, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons,

Date Reçue/Date Received 2021-09-17

New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definitions

[0048] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

[0049] The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

[0050] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (*e.g.*, between a protein and a nucleic acid). Not all components of a binding interaction need be sequence-specific (*e.g.*, contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. Such interactions are generally characterized by a dissociation constant (K_d) of 10^{-6} M^{-1} or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d.

[0051] A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding

Date Reçue/Date Received 2021-09-17

activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0052] A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0053] A "TALE DNA binding domain" or "TALE" is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains are involved in binding of the TALE to its cognate target DNA sequence. A single "repeat unit" (also referred to as a "repeat") is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein.

[0054] Zinc finger and TALE binding domains can be "engineered" to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a naturally occurring zinc finger or TALE protein. Therefore, engineered DNA binding proteins (zinc fingers or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP and/or TALE designs and binding data. See, for example, U.S. Patents 8,586,526; 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

[0055] A "selected" zinc finger protein or TALE is a protein not found in nature whose production results primarily from an empirical process such as phage display, interaction trap or hybrid selection. See e.g., US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197, WO 02/099084 and U.S. Publication No. 20110301073.

[0056] "TtAgo" is a prokaryotic Argonaute protein thought to be involved in gene silencing. TtAgo is derived from the bacteria *Thermus thermophilus*. See, *e.g.*

Date Reçue/Date Received 2021-09-17

Swarts *et al, ibid*, G. Sheng *et al.*, (2013) *Proc. Natl. Acad. Sci. U.S.A.* 111, 652). A "TtAgo system" is all the components required including, for example, guide DNAs for cleavage by a TtAgo enzyme.

[0057] "Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

[0058] A "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains;" "+ and – cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize.

[0059] An "engineered cleavage half-domain" is a cleavage half-domain that has been modified so as to form obligate heterodimers with another cleavage half-domain (e.g., another engineered cleavage half-domain). See, also, U.S. Patent Nos. 8,623,618; 7,888,121; 7,914,796; and 8,034,598 and U.S. Publication No. 20110201055.

[0060] The term "sequence" refers to a nucleotide sequence of any length, which can be DNA or RNA; can be linear, circular or branched and can be either single-stranded or double stranded. The term "donor sequence" refers to a nucleotide sequence that is inserted into a genome. A donor sequence can be of any length, for example between 2 and 10,000 nucleotides in length (or any integer value therebetween or thereabove), preferably between about 100 and 1,000 nucleotides in length (or any integer therebetween), more preferably between about 200 and 500 nucleotides in length.

[0061] A "homologous, non-identical sequence" refers to a first sequence which shares a degree of sequence identity with a second sequence, but whose sequence is not identical to that of the second sequence. For example, a polynucleotide comprising the wild-type sequence of a mutant gene is homologous

Date Reçue/Date Received 2021-09-17

and non-identical to the sequence of the mutant gene. In certain embodiments, the degree of homology between the two sequences is sufficient to allow homologous recombination therebetween, utilizing normal cellular mechanisms. Two homologous non-identical sequences can be any length and their degree of non-homology can be as small as a single nucleotide (*e.g.*, for correction of a genomic point mutation by targeted homologous recombination) or as large as 10 or more kilobases (*e.g.*, for insertion of a gene at a predetermined ectopic site in a chromosome). Two polynucleotides comprising the homologous non-identical sequences need not be the same length. For example, an exogenous polynucleotide (*i.e.*, donor polynucleotide) of between 20 and 10,000 nucleotides or nucleotide pairs can be used.

[0062] Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their percent identity. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. With respect to sequences described herein, the range of desired degrees of sequence identity is approximately 80% to 100% and any integer value therebetween. Typically the percent identities between sequences are at least 70-75%, preferably 80-82%, more preferably 85-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence identity.

[0063] Alternatively, the degree of sequence similarity between polynucleotides can be determined by hybridization of polynucleotides under conditions that allow formation of stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two nucleic acid, or two polypeptide sequences are substantially homologous to each other when the sequences exhibit at least about 70%-75%, preferably 80%-82%, more preferably 85%-90%, even more preferably 92%, still more preferably 95%, and most preferably 98% sequence

Date Reçue/Date Received 2021-09-17

identity over a defined length of the molecules, as determined using the methods above. As used herein, substantially homologous also refers to sequences showing complete identity to a specified DNA or polypeptide sequence. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, *e.g.*, Sambrook et al., *supra*; Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

[0064] Selective hybridization of two nucleic acid fragments can be determined as follows. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit the hybridization of a completely identical sequence to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (*e.g.*, Southern (DNA) blot, Northern (RNA) blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

[0065] When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a reference nucleic acid sequence, and then by selection of appropriate conditions the probe and the reference sequence selectively hybridize, or bind, to each other to form a duplex molecule. A nucleic acid molecule that is capable of hybridizing selectively to a reference sequence under moderately stringent hybridization conditions typically hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14

Date Reçue/Date Received 2021-09-17

nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for probe/reference sequence hybridization, where the probe and reference sequence have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

[0066] Conditions for hybridization are well-known to those of skill in the art. Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of hybrids containing mismatched nucleotides, with higher stringency correlated with a lower tolerance for mismatched hybrids. Factors that affect the stringency of hybridization are well-known to those of skill in the art and include, but are not limited to, temperature, pH, ionic strength, and concentration of organic solvents such as, for example, formamide and dimethylsulfoxide. As is known to those of skill in the art, hybridization stringency is increased by higher temperatures, lower ionic strength and lower solvent concentrations.

[0067] With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of the sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (*e.g.*, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. "Recombination" refers to a process of exchange of genetic information between two polynucleotides. For the purposes of this disclosure, "homologous recombination (HR)" refers to the specialized form of such exchange that takes place, for example, during repair of double-strand breaks in cells. This process requires nucleotide sequence homology, uses a "donor" molecule to template repair of a "target" molecule (*i.e.*, the one that experienced the double-strand break), and is variously known as "non-crossover gene conversion" or "short tract gene conversion," because it leads to the transfer of genetic information from the donor to the target. Without wishing to be bound by any particular theory, such transfer can involve mismatch correction of heteroduplex DNA that forms between the broken target and the donor, and/or "synthesis-dependent strand annealing," in which the donor is used to resynthesize genetic information that will become part of the target,

Date Reçue/Date Received 2021-09-17

and/or related processes. Such specialized HR often results in an alteration of the sequence of the target molecule such that part or all of the sequence of the donor polynucleotide is incorporated into the target polynucleotide.

[0068] "Chromatin" is the nucleoprotein structure comprising the cellular genome. Cellular chromatin comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term "chromatin" is meant to encompass all types of cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

[0069] A "chromosome," is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

[0070] An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

[0071] An "accessible region" is a site in cellular chromatin in which a target site present in the nucleic acid can be bound by an exogenous molecule which recognizes the target site. Without wishing to be bound by any particular theory, it is believed that an accessible region is one that is not packaged into a nucleosomal structure. The distinct structure of an accessible region can often be detected by its sensitivity to chemical and enzymatic probes, for example, nucleases.

[0072] A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist. For example, the sequence 5'-GAATTC-3' is a target site for the Eco RI restriction endonuclease.

[0073] An "exogenous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other

Date Reçue/Date Received 2021-09-17

methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule, a malfunctioning version of a normally-functioning endogenous molecule or an ortholog (functioning version of endogenous molecule from a different species).

[0074] An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors, methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

[0075] An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, an exogenous protein or nucleic acid. For example, an exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

[0076] By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome,

Date Reçue/Date Received 2021-09-17

the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0077] A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules. Examples of the first type of fusion molecule include, but are not limited to, fusion proteins (for example, a fusion between a ZFP DNA-binding domain and a cleavage domain) and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein described *supra*). Examples of the second type of fusion molecule include, but are not limited to, a fusion between a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

[0078] Expression of a fusion protein in a cell can result from delivery of the fusion protein to the cell or by delivery of a polynucleotide encoding the fusion protein to a cell, wherein the polynucleotide is transcribed, and the transcript is translated, to generate the fusion protein. Trans-splicing, polypeptide cleavage and polypeptide ligation can also be involved in expression of a protein in a cell. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0079] A "gene," for the purposes of the present disclosure, includes a DNA region encoding a gene product (see *infra*), as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0080] "Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins

Date Reçue/Date Received 2021-09-17

modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0081] "Modulation" of gene expression refers to a change in the activity of a gene. Modulation of expression can include, but is not limited to, gene activation and gene repression. Gene inactivation refers to any reduction in gene expression as compared to a cell that does not include a ZFP as described herein. Thus, gene inactivation may be partial or complete.

[0082] "Eukaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (*e.g.*, T-cells).

[0083] A "region of interest" is any region of cellular chromatin, such as, for example, a gene or a non-coding sequence within or adjacent to a gene, in which it is desirable to bind an exogenous molecule. Binding can be for the purposes of targeted DNA cleavage and/or targeted recombination. A region of interest can be present in a chromosome, an episome, an organellar genome (*e.g.*, mitochondrial, chloroplast), or an infecting viral genome, for example. A region of interest can be within the coding region of a gene, within transcribed non-coding regions such as, for example, leader sequences, trailer sequences or introns, or within non-transcribed regions, either upstream or downstream of the coding region. A region of interest can be as small as a single nucleotide pair or up to 2,000 nucleotide pairs in length, or any integral value of nucleotide pairs.

[0084] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in *cis* with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

Date Reçue/Date Received 2021-09-17

[0085] With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a DNA-binding domain is fused to a cleavage domain, the DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the DNA-binding domain portion is able to bind its target site and/or its binding site, while the cleavage domain is able to cleave DNA in the vicinity of the target site (*e.g.*, 1 to 500 base pairs or any value therebetween on either side of the target site).

[0086] In the methods of the disclosure, one or more targeted nucleases as described herein create a double-stranded break (DSB) in the target sequence (*e.g.*, cellular chromatin) at a predetermined site. The DSB may result in deletions and/or insertions by homology-directed repair or by non-homology-directed repair mechanisms. Deletions may include any number of base pairs. Similarly, insertions may include any number of base pairs including, for example, integration of a "donor" polynucleotide, optionally having homology to the nucleotide sequence in the region of the break. The donor sequence may be physically integrated or, alternatively, the donor polynucleotide is used as a template for repair of the break via homologous recombination, resulting in the introduction of all or part of the nucleotide sequence as in the donor into the cellular chromatin. Thus, a first sequence in cellular chromatin can be altered and, in certain embodiments, can be converted into a sequence present in a donor polynucleotide. Thus, the use of the terms "replace" or "replacement" can be understood to represent replacement of one nucleotide sequence by another, (*i.e.*, replacement of a sequence in the informational sense), and does not necessarily require physical or chemical replacement of one polynucleotide by another.

[0087] Additional pairs of zinc-finger proteins, TALENs, TtAgo or CRIPSR/Cas systems can be used for additional double-stranded cleavage of additional target sites within the cell.

[0088] In any of the methods described herein, additional pairs of zinc-finger proteins, TALENs, TtAgo or CRIPSR/Cas systems can be used for additional double-stranded cleavage of additional target sites within the cell.

[0089] A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length

Date Reçue/Date Received 2021-09-17

protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (*e.g.*, coding function, ability to hybridize to another nucleic acid) are well-known in the art. Similarly, methods for determining protein function are well-known. For example, the DNA-binding function of a polypeptide can be determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. DNA cleavage can be assayed by gel electrophoresis. See Ausubel *et al., supra.* The ability of a protein to interact with another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical. See, for example, Fields *et al.* (1989) *Nature* **340:**245-246; U.S. Patent No. 5,585,245 and PCT WO 98/44350.

Linkers

[0090] Described herein are amino acid sequences that fuse (link) a DNA binding domain (*e.g.*, zinc finger protein, TALE, etc.) and a nuclease (*e.g.*, a cleavage domain or cleavage half-domain).

[0091] Currently, when a pair of nucleases is used to cleave a genomic sequence, optimal cleavage is obtained when the DNA-binding proteins bind to target sites separated by less than 6 base pairs. In particular, optimal cleavage for zinc finger nuclease pairs is obtained when the target sites are separated by 5-6 base pairs and a flexible "ZC" linker rich in glycine and serine is used to join each zinc finger of the pair to the cleavage domain. In particular, the "ZC" linker used to date consists of the amino acid sequence LRGS (SEQ ID NO:2) between the C-terminal of the zinc finger binding domain and the N-terminal residues of the cleavage domain, which in the case of *Fok*I is a Q residue. See, *e.g.*, U.S. Patent No. 7,888,121. In addition, fusion proteins comprising ZC linkers and additional modifications to the N-terminal region of the cleavage domain have also been described. *See,* U.S. Patent Publication No. 20090305419. Furthermore, U.S. Patent No. 8,772,453 describes linkers useful for obtaining cleavage when the DNA-binding sites target sequences separated by less than 5 base pairs.

[0092] The linkers described herein allow cleavage when the target sites of a pair of zinc finger nucleases are not 0-6 base pairs apart, for example target sites that

Date Reçue/Date Received 2021-09-17

are 7, 8, 9, 10 or more base pairs apart. The linker sequences described herein are typically between about 8 and 17 amino acids in length and may link the N- or C-terminal of the DNA-binding domain to the N- or C-terminal of cleavage domain. In certain embodiments, the linker extends between the C-terminal residue of the DNA-binding domain and the N-terminal residue of the cleavage domain (*e.g.*, (Q) of *Fok*I.

[0093] Non-limiting examples of linkers as described herein are shown in Figures 4, 7, 9, 11, 12, 13, 15 and 16 (SEQ ID NOs:2-21, 45-49, 51, 62-226, 233-240 and 258-312) (with or without 1 or 2 of the C-terminal amino acid residues shown in these Figures), including but not limited to (N)GICPPPRPRTSPP (SEQ ID NO:2); (T)GTAPIEIPPEVYP (SEQ ID NO:3); (N)GSYAPMPPLALASP (SEQ ID NO:4); (P)GIYTAPTSRPTVPP (SEQ ID NO:5) ; (N)GSQTPKRFQPTHPSA (SEQ ID NO:6); (H)LPKPANPFPLD (SEQ ID NO:7); (H)RDGPRNLPPTSPP (SEQ ID NO:8); (H)RLPDSPTALAPDTL (SEQ ID NO:9); (DPNS)PISRARPLNPHP (SEQ ID NO:10); (Y)GPRPTPRLRCPIDSLIFR (SEQ ID NO:11); (H)CPASRPIHP (SEQ ID NO:12); (G)LQSLIPQQLL (SEQ ID NO:13); (G)LQPTVNHEYNN (SEQ ID NO:14); (P)ANIHSLSSPPPL (SEQ ID NO:15); (P)AGLNTPCSPRSRSN (SEQ ID NO:16); (A)TITDPNP(SEQ ID NO:17); (P)PHKGLLP (SEQ ID NO:18); (S)VSLPDTHH(SEQ ID NO:19); (G)THGATPTHSP (SEQ ID NO:20); (V)APGESSMTSL(SEQ ID NO:21), (LRGS)PISRARPLNPHP (SEQ ID NO:22), (LRGS)ISRARPLNPHP (SEQ ID NO:23), (LRGS)PSRARPLNPHP (SEQ ID NO:24), (LRGS)SRARPLNPHP (SEQ ID NO:25), (LRGS)YAPMPPLALASP (SEQ ID NO:26), (LRGS)APMPPLALASP (SEQ ID NO:27), (LRGS)PMPPLALASP (SEQ ID NO:28), (LRGS)MPPLALASP (SEQ ID NO:29) HLPKPANPFPLD (SEQ ID NO:30), wherein the amino acid residue(s) shown in round parentheses is(are) optionally present. In certain embodiments, the linker comprises a sequence selected from the group consisting of PKPAN (SEQ ID NO:31), RARPLN (SEQ ID NO:32); PMPPLA (SEQ ID NO:33) or PPPRP (SEQ ID NO:34). In certain embodiments, the linkers further comprise a ZC linker (LRGS, SEQ ID NO:35) at their N-terminal.

[0094] In certain embodiments, the linker comprises a sequence as shown in Figure 4 (SEQ ID NOs:62-226, respectively, in order of appearance). In other embodiments, the linker comprises a sequence as shown in Figure 7, Figure 9A (top), Figure 12 and Figure 13 (L8 linkers including, for example, L8a (SEQ ID NO:2), L8b (SEQ ID NO:3), L8c (SEQ ID NO:4), L8d (SEQ ID NO:5), L8e (SEQ ID NO:6), L8g (SEQ ID NO:234), L8i (SEQ ID NO:161), L8j (SEQ ID NO:235), L8a (SEQ ID

Date Reçue/Date Received 2021-09-17

NO:2), L8c (SEQ ID NO:4), L8a3 (SEQ ID NO:268) and L8c3 (SEQ ID NO:269)). In certain embodiments, the L8 linkers are used when the DNA-binding domains of the dimerizing nuclease pair used for cleavage bind to target sites separate by 8 base pairs. In other embodiments, the linker comprises a sequence as shown in Figure 9A (bottom), Figure 12 and Figure 13 (L7 linkers including, for example, L7-1 (SEQ ID NO:7), L7-2 (SEQ ID NO:8), L7-6 (SEQ ID NO:9), L7-3 (SEQ ID NO:10), L7-5 (SEQ ID NO:11), L7-4 (SEQ ID NO:236), L7-b (SEQ ID NO:258), L7-c (SEQ ID NO:259), L7-b3 (SEQ ID NO:266) and L7-c3 (SEQ ID NO:267)). In certain embodiments, the L7 linkers are used when the DNA-binding domains of the dimerizing nuclease pair used for cleavage bind to target sites separate by 7 base pairs. In other embodiments, the linker comprises a sequence as shown in Figure 9B (top) (L6 linkers, including, for example, L6-2 (SEQ ID NO:12), L6-6 (SEQ ID NO:13), L6-7 (SEQ ID NO:14), L6-1 (SEQ ID NO:15), L6-5 (SEQ ID NO:16), L6-3 (SEQ ID NO:115), or L6-4 (SEQ ID NO:237)). In certain embodiments, the L6 linkers are used when the DNA-binding domains of the dimerizing nuclease pair used for cleavage bind to target sites separate by 6 base pairs. In still other embodiments, the linker comprises a sequence as shown in Figure 9B (bottom) (L5-6 (SEQ ID NO:17), L5-1 (SEQ ID NO:18), L5-7 (SEQ ID NO:19), L5-8 (SEQ ID NO:20), L5-9 (SEQ ID NO:21), L5-2 (SEQ ID NO:238), L5-3 (SEQ ID NO:239), L5-4 (SEQ ID NO:240), L5-5 (SEQ ID NO:71)). In certain embodiments, the L5 linkers are used when the DNA-binding domains of the dimerizing nuclease pair used for cleavage bind to target sites separate by 5 base pairs.

[0095] The fusion proteins described herein may also include alterations to the N-terminal region of the selected cleavage domain. Alteration may include substitutions, additions and/or deletions of one or more N-terminal residues of the cleavage domain. In certain embodiments, the cleavage domain is derived from *Fok*I and one or more amino acids of the wild-type *Fok*I N-terminal region are replaced and additional amino acids added to this region. For example, as shown in FIG. 2, amino acid residues 4 and 5 of the wild-type *Fok*I cleavage domain (i.e., residues K and S) are replaced with residues E and A, respectively and the residues AAR is added C-terminal to the 2nd replaced residue. Another exemplary embodiment (FIG. 3) includes a seven residue insertion (KSEAAAR; SEQ ID NO:36) in the N-terminal region of the *Fok*I cleavage domain.

Date Reçue/Date Received 2021-09-17

[0096] The sequence joining the DNA-binding domain and the cleavage domain can comprise any amino acid sequence that does not substantially hinder the ability of the DNA-binding domain to bind to its target site or the cleavage domain to dimerize and/or cleave the genomic sequences. In wild-type *Fok*I, the N-terminal region of the cleavage domain includes an alpha helical region extending from residues 389-400 (ELEEKKSELRHK; SEQ ID NO:37). See, *e.g.*, Wah *et al.* (1997) *Nature* 388:97-100). Therefore, in certain embodiments, the linker sequences are designed to extend and/or conserve this structural motif, for example by inserting a 3-5 amino sequence N-terminal to ELEEKKSELRHK (SEQ ID NO:37) of a wild-type *Fok*I cleavage domain.

[0097] Thus, the linker may include a sequence such as EXXXR (SEQ ID NO:38) or EXXXK (SEQ ID NO:39) where the X residues are any residues that form an alpha helix, namely any residue except proline or glycine (*e.g.*, EAAAR (SEQ ID NO:40)) adjacent to the wild-type alpha helical region to form a stable alpha helix linker. See, *e.g.*, Yan *et al.* (2007) *Biochemistry* 46:8517-24 and Merutka and Stellwagen (1991) *Biochemistry* 30:4245-8. Placing an EXXXR (SEQ ID NO:38) or EXXXK (SEQ ID NO:39) peptide adjacent (or near to) to the ELEEKKSELRHK (SEQ ID NO:37) peptide is designed to extend this alpha helix in *Fok*I cleavage domain. This creates a more rigid linker between the ZFP and *Fok*I cleavage domain which allows the resulting ZFN pair to cleave a target with more than 6 bp between the half sites without the loss in activity and specificity that can be observed when a long flexible linker is used between the ZFP and the *Fok*I domain (Bibikova *et al.* (2001) *Molecular and Cellular Biology* 21:289-297). In addition, the linkers described herein show a greater preference for a 6 bp spacing over a 5 bp spacing as compared to current ZFNs.

[0098] Furthermore, in certain embodiments, the junction as between the linker and the cleavage domain and/or DNA-binding domain is modified, for example to substitute, add and/or delete amino acids to the linkers as described herein. In certain embodiments, 1, 2, 3, 4 or more amino acids are deleted or added. In other embodiment, 1, 2 or 3, 4 amino acid substitutions are made. Non-limiting examples of modified linkers as described herein are shown in Figures 15 and 16 (SEQ ID NOs:280-312 and 51).

[0099] Also described herein is a fusion protein comprising a DNA-binding domain, a modified FokI cleavage domain and a ZC linker between the DNA-binding

Date Reçue/Date Received 2021-09-17

domain and the FokI cleavage domain. The FokI cleavage domain may be modified in any way. Non-limiting examples of modifications include additions, deletions and/or substitutions to the N-terminal region of FokI (residues 158-169 of SEQ ID NO:1). In certain embodiments, the modified FokI cleavage domain comprises deletion of 1, 2, 3, 4 or more amino acids from the N-terminal region of FokI (*e.g.*, deletion of one or more of residues 158, 159, 160 and/or 161 of the wild-type FokI domain of SEQ ID NO:1). Non-limiting examples of proteins with deletions of N-terminal FokI amino acid residues include the proteins designated V1, V3 and V7 as shown in Figure 15. In other embodiments, the modified FokI cleavage domain comprises one or more deletions and one or more substitutions from the N-terminal region of FokI (*e.g.*, deletion of one or more of residues 158-161 and substitution of one or more of the remaining residues). Non-limiting examples of proteins with deletions and substitutions in the N-terminal FokI amino acid residues include the proteins designated V2, V4, V5, V6 and V8 as shown in Figure 15. In other embodiments, the modified FokI cleavage domain comprises one or more substitutions in the N-terminal region of FokI. Non-limiting examples of proteins with substitutions in the N-terminal amino acid residues of FokI include the proteins designated V9 through V16 as shown in Figure 15. In still further embodiments, the modified FokI cleavage domain comprises one or more additional amino acid residues (*e.g.*, 1, 2, 3, 4 or more) N-terminal to the N-terminal-most residue of FokI (residue 158 of SEQ ID NO:1). Non-limiting examples of proteins with additions to the N-terminal FokI amino acid residues include the proteins designated V17 through V24 as shown in Figure 15. In other embodiments, the modified FokI cleavage domain comprises one or more additional amino acid residues (*e.g.*, 1, 2, 3, 4 or more) N-terminal to the N-terminal-most residue of FokI (residue 158 of SEQ ID NO:1) and one or more substitutions within the N-terminal region of FokI. Non-limiting examples of proteins with additions include those shown in Figure 15 or Figure 16 and substitutions within the N-terminal FokI amino acid residues as described in U.S. Patent Publication No 20090305419. In certain embodiments, the fusion protein comprises a modified FokI cleavage domain as shown in Figure 15 or Figure 16.

[0100] Typically, the linkers of the invention are made by making recombinant nucleic acids encoding the linker and the DNA-binding domains, which are fused via the linker amino acid sequence. Optionally, the linkers can also be made using peptide synthesis, and then linked to the polypeptide DNA-binding domains.

Date Reçue/Date Received 2021-09-17

Nucleases

[0101] The linker sequences described herein are advantageously used to link DNA-binding domains, for example zinc finger proteins, TALEs, homing endonucleases, CRISPR/Cas and/or Ttago guide RNAs, to nuclease cleavage domains or half domains to form specifically targeted, non-naturally occurring nucleases.

A. DNA-binding domains

[0102] Any DNA-binding domain can be used in the methods disclosed herein. In certain embodiments, the DNA binding domain comprises a zinc finger protein. Preferably, the zinc finger protein is non-naturally occurring in that it is engineered to bind to a target site of choice. *See*, for example, Beerli *et al.* (2002) *Nature Biotechnol.* **20:**135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* **70:**313-340; Isalan *et al.* (2001) *Nature Biotechnol.* **19:**656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* **12:**632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* **10:**411-416. An engineered zinc finger binding domain can have a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, co-owned U.S. Patents 6,453,242 and 6,534,261.

[0103] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0104] Selection of target sites; ZFPs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Application Publication Nos. 20050064474 and 20060188987.

[0105] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any

Date Reçue/Date Received 2021-09-17

suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0106] In certain embodiments, the composition and methods described herein employ a meganuclease (homing endonuclease) DNA-binding domain for binding to the donor molecule and/or binding to the region of interest in the genome of the cell. Naturally-occurring meganucleases recognize 15-40 base-pair cleavage sites and are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cyst box family and the HNH family. Exemplary homing endonucleases include I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III. Their recognition sequences are known. *See* also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.*(1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.*(1994) Nucleic Acids Res. 22, 1125–1127; Jasin (1996) *Trends Genet.***12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue. In addition, the DNA-binding specificity of homing endonucleases and meganucleases can be engineered to bind non-natural target sites. *See*, for example, Chevalier *et al.* (2002) *Molec. Cell* **10**:895-905; Epinat *et al.*(2003) *Nucleic Acids Res.* **31**:2952-2962; Ashworth *et al.* (2006) *Nature* **441**:656-659; Paques *et al.* (2007) *Current Gene Therapy* **7**:49-66; U.S. Patent Publication No. 20070117128. The DNA-binding domains of the homing endonucleases and meganucleases may be altered in the context of the nuclease as a whole (*i.e.*, such that the nuclease includes the cognate cleavage domain) or may be fused to a heterologous cleavage domain.

[0107] In other embodiments, the DNA-binding domain of one or more of the nucleases used in the methods and compositions described herein comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding domain. *See, e.g.*, U.S. Patent No. 8,586,526. The plant pathogenic bacteria of the genus *Xanthomonas* are known to cause many diseases in important crop plants. Pathogenicity of *Xanthomonas* depends on a conserved type III secretion (T3S) system which injects more than 25 different effector proteins into the plant cell. Among these injected proteins are transcription activator-like (TAL) effectors which

Date Reçue/Date Received 2021-09-17

mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay *et al* (2007) *Science* 318:648-651). These proteins contain a DNA binding domain and a transcriptional activation domain. One of the most well characterized TAL-effectors is AvrBs3 from *Xanthomonas campestgris* pv. *Vesicatoria* (see Bonas *et al*(1989) *Mol Gen Genet* 218: 127-136 and WO2010079430). TAL-effectors contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, they contain a nuclear localization sequence and an acidic transcriptional activation domain (for a review see Schornack S, *et al* (2006) *J Plant Physiol* 163(3): 256-272). In addition, in the phytopathogenic bacteria *Ralstonia solanacearum* two genes, designated brg11 and hpx17 have been found that are homologous to the AvrBs3 family of *Xanthomonas* in the *R. solanacearum* biovar 1 strain GMI1000 and in the biovar 4 strain RS1000 (See Heuer *et al* (2007) *Appl and Envir Micro* 73(13): 4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ by a deletion of 1,575 bp in the repeat domain of hpx17. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of *Xanthomonas*. *See, e.g.,* U.S. Patent No. 8,586,526.

[0108] Specificity of these TAL effectors depends on the sequences found in the tandem repeats. The repeated sequence comprises approximately 102 bp and the repeats are typically 91-100% homologous with each other (Bonas *et al, ibid*). Polymorphism of the repeats is usually located at positions 12 and 13 and there appears to be a one-to-one correspondence between the identity of the hypervariable diresidues (RVD) at positions 12 and 13 with the identity of the contiguous nucleotides in the TAL-effector's target sequence (see Moscou and Bogdanove, (2009) *Science* 326:1501 and Boch *et al* (2009) *Science* 326:1509-1512). Experimentally, the natural code for DNA recognition of these TAL-effectors has been determined such that an HD sequence at positions 12 and 13 leads to a binding to cytosine (C), NG binds to T, NI to A, C, G or T, NN binds to A or G, and ING binds to T. These DNA binding repeats have been assembled into proteins with new combinations and numbers of repeats, to make artificial transcription factors that are able to interact with new sequences and activate the expression of a non-endogenous reporter gene in plant cells (Boch *et al, ibid*). Engineered TAL proteins have been linked to a *Fok*I cleavage half domain to yield a TAL effector domain nuclease fusion (TALEN). *See, e.g.,* U.S. Patent No. 8,586,526; Christian *et al* ((2010)<*Genetics*

epub 10.1534/genetics.110.120717). In certain embodiments, TALE domain comprises an N-cap and/or C-cap as described in U.S. Patent No. 8,586,526. In still further embodiments, the nuclease comprises a compact TALEN (cTALEN). These are single chain fusion proteins linking a TALE DNA binding domain to a TevI nuclease domain. The fusion protein can act as either a nickase localized by the TALE region, or can create a double strand break, depending upon where the TALE DNA binding domain is located with respect to the TevI nuclease domain (see Beurdeley *et al* (2013) *Nat Comm*: 1-8 DOI: 10.1038/ncomms2782). Any TALENs may be used in combination with additional TALENs (*e.g.*, one or more TALENs (cTALENs or FokI-TALENs) with one or more mega-TALs).

[0109] In certain embodiments, the DNA-binding domain is part of a CRISPR/Cas nuclease system. *See, e.g.*, U.S. Patent No. 8,697,359 and U.S. Patent Application No. 14/278,903. The CRISPR (clustered regularly interspaced short palindromic repeats) locus, which encodes RNA components of the system, and the cas (CRISPR-associated) locus, which encodes proteins (Jansen *et al.*, 2002. *Mol. Microbiol.* 43: 1565-1575; Makarova *et al.*, 2002. *Nucleic Acids Res.* 30: 482-496; Makarova *et al.*, 2006. *Biol. Direct* 1: 7; Haft *et al.*, 2005. *PLoSComput. Biol.* 1: e60) make up the gene sequences of the CRISPR/Cas nuclease system. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

[0110] The Type II CRISPR is one of the most well characterized systems and carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNA, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of alien DNA sequences into the CRISPR array to prevent future attacks, in a process called 'adaptation', (ii) expression of the relevant proteins, as well as expression and processing of the array,

Date Reçue/Date Received 2021-09-17

followed by (iii) RNA-mediated interference with the alien nucleic acid. Thus, in the bacterial cell, several of the so-called 'Cas' proteins are involved with the natural function of the CRISPR/Cas system and serve roles in functions such as insertion of the alien DNA etc.

[0111] In certain embodiments, Cas protein may be a "functional derivative" of a naturally occurring Cas protein. A "functional derivative" of a native sequence polypeptide is a compound having a qualitative biological property in common with a native sequence polypeptide. "Functional derivatives" include, but are not limited to, fragments of a native sequence and derivatives of a native sequence polypeptide and its fragments, provided that they have a biological activity in common with a corresponding native sequence polypeptide. A biological activity contemplated herein is the ability of the functional derivative to hydrolyze a DNA substrate into fragments. The term "derivative" encompasses both amino acid sequence variants of polypeptide, covalent modifications, and fusions thereof. Suitable derivatives of a Cas polypeptide or a fragment thereof include but are not limited to mutants, fusions, covalent modifications of Cas protein or a fragment thereof. Cas protein, which includes Cas protein or a fragment thereof, as well as derivatives of Cas protein or a fragment thereof, may be obtainable from a cell or synthesized chemically or by a combination of these two procedures. The cell may be a cell that naturally produces Cas protein, or a cell that naturally produces Cas protein and is genetically engineered to produce the endogenous Cas protein at a higher expression level or to produce a Cas protein from an exogenously introduced nucleic acid, which nucleic acid encodes a Cas that is same or different from the endogenous Cas. In some case, the cell does not naturally produce Cas protein and is genetically engineered to produce a Cas protein.

[0112] In some embodiments, the DNA binding domain is part of a TtAgo system (see Swarts *et al, ibid*; Sheng *et al, ibid*). In eukaryotes, gene silencing is mediated by the Argonaute (Ago) family of proteins. In this paradigm, Ago is bound to small (19-31 nt) RNAs. This protein-RNA silencing complex recognizes target RNAs via Watson-Crick base pairing between the small RNA and the target and endonucleolytically cleaves the target RNA (Vogel (2014) *Science* 344:972-973). In contrast, prokaryotic Ago proteins bind to small single-stranded DNA fragments and likely function to detect and remove foreign (often viral) DNA (Yuan *et al.*, (2005) *Mol. Cell* 19, 405; Olovnikov, *et al.* (2013) *Mol. Cell* 51, 594; Swarts *et al., ibid*).

Date Reçue/Date Received 2021-09-17

Exemplary prokaryotic Ago proteins include those from *Aquifex aeolicus*, *Rhodobacter sphaeroides*, and *Thermus thermophilus*.

[0113] One of the most well-characterized prokaryotic Ago protein is the one from *T. thermophilus* (TtAgo; Swarts *et al. ibid*). TtAgo associates with either 15 nt or 13-25 nt single-stranded DNA fragments with 5' phosphate groups. This "guide DNA" bound by TtAgo serves to direct the protein-DNA complex to bind a Watson-Crick complementary DNA sequence in a third-party molecule of DNA. Once the sequence information in these guide DNAs has allowed identification of the target DNA, the TtAgo-guide DNA complex cleaves the target DNA. Such a mechanism is also supported by the structure of the TtAgo-guide DNA complex while bound to its target DNA (G. Sheng *et al., ibid*). Ago from *Rhodobacter sphaeroides* (RsAgo) has similar properties (Olivnikov *et al. ibid*).

[0114] Exogenous guide DNAs of arbitrary DNA sequence can be loaded onto the TtAgo protein (Swarts *et al. ibid.*). Since the specificity of TtAgo cleavage is directed by the guide DNA, a TtAgo-DNA complex formed with an exogenous, investigator-specified guide DNA will therefore direct TtAgo target DNA cleavage to a complementary investigator-specified target DNA. In this way, one may create a targeted double-strand break in DNA. Use of the TtAgo-guide DNA system (or orthologous Ago-guide DNA systems from other organisms) allows for targeted cleavage of genomic DNA within cells. Such cleavage can be either single- or double-stranded. For cleavage of mammalian genomic DNA, it would be preferable to use of a version of TtAgo codon optimized for expression in mammalian cells. Further, it might be preferable to treat cells with a TtAgo-DNA complex formed *in vitro* where the TtAgo protein is fused to a cell-penetrating peptide. Further, it might be preferable to use a version of the TtAgo protein that has been altered via mutagenesis to have improved activity at 37°C. Ago-RNA-mediated DNA cleavage could be used to effect a panopoly of outcomes including gene knock-out, targeted gene addition, gene correction, targeted gene deletion using techniques standard in the art for exploitation of DNA breaks.

B. Cleavage Domains

[0115] The nucleases described herein (*e.g.*, ZFNs, TALENs, CRISPR/Cas nuclease) also comprise a nuclease (cleavage domain, cleavage half-domain). The cleavage domain portion of the fusion proteins disclosed herein can be obtained from

Date Reçue/Date Received 2021-09-17

any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. *See*, for example, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; and Belfort *et al.* (1997) *Nucleic Acids Res.* **25:**3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; *see* also Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press,1993). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0116] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof).

[0117] In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However any integral number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0118] As noted above, the cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a TALEN DNA-binding domain and a cleavage domain, or meganuclease DNA-binding domain and cleavage domain from a different nuclease, or a DNA binding domain from a CRISPR/Cas system and a cleavage domain from a difference nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. Additional enzymes which cleave DNA

Date Reçue/Date Received 2021-09-17

are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease. One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0119] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However any integral number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0120] Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding. Certain restriction enzymes (*e.g.*, Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme *Fok* I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. See, for example, US Patents 5,356,802; 5,436,150 and 5,487,994; as well as Li *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89:**4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90:**2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* **91:**883-887; Kim *et al.* (1994b) *J. Biol. Chem.* **269:**31,978-31,982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

Date Reçue/Date Received 2021-09-17

[0121] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok* I. This particular enzyme is active as a dimer. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95:** 10,570-10,575. Accordingly, for the purposes of the present disclosure, the portion of the Fok I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok* I fusions, two fusion proteins, each comprising a *Fok*I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a zinc finger binding domain and two *Fok* I cleavage half-domains can also be used. Parameters for targeted cleavage and targeted sequence alteration using zinc finger-*Fok* I fusions are provided elsewhere in this disclosure.

[0122] A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (*e.g.*, dimerize) to form a functional cleavage domain.

[0123] Exemplary Type IIS restriction enzymes are described in International Publication WO 07/014275. Additional restriction enzymes also contain separable binding and cleavage domains, and these are contemplated by the present disclosure. See, for example, Roberts *et al.* (2003) *Nucleic Acids Res.* **31:**418-420.

[0124] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Nos. 8,623,618; 8,409,861; 8,034,598; 7,914,796; and 7,888,121. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok* I are all targets for influencing dimerization of the *Fok* I cleavage half-domains.

[0125] Exemplary engineered cleavage half-domains of *Fok* I that form obligate heterodimers include a pair in which a first cleavage half-domain includes mutations at amino acid residues at positions 490 and 538 of *Fok* I and a second cleavage half-domain includes mutations at amino acid residues 486 and 499.

[0126] Thus, in one embodiment, a mutation at 490 replaces Glu (E) with Lys (K); the mutation at 538 replaces Iso (I) with Lys (K); the mutation at 486 replaced Gln (Q) with Glu (E); and the mutation at position 499 replaces Iso (I) with Lys (K). Specifically, the engineered cleavage half-domains described herein were prepared by

Date Reçue/Date Received 2021-09-17

mutating positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated "E490K:I538K" and by mutating positions 486 (Q→E) and 499 (I→L) in another cleavage half-domain to produce an engineered cleavage half-domain designated "Q486E:I499L". The engineered cleavage half-domains described herein are obligate heterodimer mutants in which aberrant cleavage is minimized or abolished. *See, e.g.,* U.S. Patent No. 7,888,121.

[0127] Cleavage domains with more than one mutation may be used, for example mutations at positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated "E490K:I538K" and by mutating positions 486 (Q→E) and 499 (I→L) in another cleavage half-domain to produce an engineered cleavage half-domain designated "Q486E:I499L;" mutations that replace the wild type Gln (Q) residue at position 486 with a Glu (E) residue, the wild type Iso (I) residue at position 499 with a Leu (L) residue and the wild-type Asn (N) residue at position 496 with an Asp (D) or Glu (E) residue (also referred to as a "ELD" and "ELE" domains, respectively); engineered cleavage half-domain comprising mutations at positions 490, 538 and 537 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue, the wild type Iso (I) residue at position 538 with a Lys (K) residue, and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as "KKK" and "KKR" domains, respectively); and/or engineered cleavage half-domain comprises mutations at positions 490 and 537 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as "KIK" and "KIR" domains, respectively). *See, e.g.,* U.S. Patent Nos. 7,914,796; 8,034,598 and 8,623,618. In other embodiments, the engineered cleavage half domain comprises the "Sharkey" and/or "Sharkey' " mutations (see Guo *et al,* (2010) *J. Mol. Biol.* 400(1):96-107).

[0128] Alternatively, nucleases may be assembled *in vivo* at the nucleic acid target site using so-called "split-enzyme" technology (see *e.g.* U.S. Patent Publication No. 20090068164). Components of such split enzymes may be expressed either on separate expression constructs, or can be linked in one open reading frame where the individual components are separated, for example, by a self-cleaving 2A peptide or

Date Reçue/Date Received 2021-09-17

IRES sequence. Components may be individual zinc finger binding domains or domains of a meganuclease nucleic acid binding domain.

[0129] Nucleases can be screened for activity prior to use, for example in a yeast-based chromosomal system as described in U.S. Patent No. 8,563,314.

[0130] The Cas9 related CRISPR/Cas system comprises two RNA non-coding components: tracrRNA and a pre-crRNA array containing nuclease guide sequences (spacers) interspaced by identical direct repeats (DRs). To use a CRISPR/Cas system to accomplish genome engineering, both functions of these RNAs must be present (see Cong *et al*, (2013) *Sciencexpress* 1/10.1126/science 1231143). In some embodiments, the tracrRNA and pre-crRNAs are supplied via separate expression constructs or as separate RNAs. In other embodiments, a chimeric RNA is constructed where an engineered mature crRNA (conferring target specificity) is fused to a tracrRNA (supplying interaction with the Cas9) to create a chimeric cr-RNA-tracrRNA hybrid (also termed a single guide RNA). (see Jinek *ibid* and Cong, *ibid*).

Target Sites

[0131] As described in detail above, DNA-binding domains of the fusion proteins comprising the linkers as described herein can be engineered to bind to any sequence of choice. An engineered DNA-binding domain can have a novel binding specificity, compared to a naturally-occurring DNA-binding domain.

[0132] Non-limiting examples of suitable target genes a beta (β) globin gene (HBB), a gamma (δ) globin gene (HBG1), a B-cell lymphoma/leukemia 11A (BCL11A) gene, a Kruppel-like factor 1 (KLF1) gene, a CCR5 gene, a CXCR4 gene, a PPP1R12C (AAVS1) gene, an hypoxanthine phosphoribosyltransferase (HPRT) gene, an albumin gene, a Factor VIII gene, a Factor IX gene, a Leucine-rich repeat kinase 2 (LRRK2) gene, a Hungtingin (Htt) gene, a rhodopsin (RHO) gene, a Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene, a surfactant protein B gene (SFTPB), a T-cell receptor alpha (TRAC) gene, a T-cell receptor beta (TRBC) gene, a programmed cell death 1 (PD1) gene, a Cytotoxic T-Lymphocyte Antigen 4 (CTLA-4) gene, an human leukocyte antigen (HLA) A gene, an HLA B gene, an HLA C gene, an HLA-DPA gene, an HLA-DQ gene, an HLA-DRA gene, a LMP7 gene, a Transporter associated with Antigen Processing (TAP) 1 gene, a TAP2 gene, a tapasin gene (TAPBP), a class II major histocompatibility complex transactivator

Date Reçue/Date Received 2021-09-17

(CIITA) gene, a dystrophin gene (DMD), a glucocorticoid receptor gene (GR), an IL2RG gene, a Rag-1 gene, an RFX5 gene, a FAD2 gene, a FAD3 gene, a ZP15 gene, a KASII gene, a MDH gene, and/or an EPSPS gene.

[0133] In certain embodiments, the nuclease targets a "safe harbor" loci such as the AAVS1, HPRT, albumin and CCR5 genes in human cells, and Rosa26 in murine cells (*see, e.g.*, U.S. Patent Nos. 7,888,121; 7,972,854; 7,914,796; 7,951,925; 8,110,379; 8,409,861; 8,586,526; U.S. Patent Publications 20030232410; 20050208489; 20050026157; 20060063231; 20080159996; 201000218264; 20120017290; 20110265198; 20130137104; 20130122591; 20130177983 and 20130177960) and the Zp15 locus in plants (*see* United Stated Patent U.S. 8,329,986).

Donors

[0134] In certain embodiments, the present disclosure relates to nuclease-mediated modification of the genome of a stem cell. As noted above, insertion of an exogenous sequence (also called a "donor sequence" or "donor" or "transgene"), for example for deletion of a specified region and/or correction of a mutant gene or for increased expression of a wild-type gene. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence where it is placed. A donor sequence can contain a non-homologous sequence flanked by two regions of homology to allow for efficient HDR at the location of interest or can be integrated via non-homology directed repair mechanisms. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous to the region of interest in cellular chromatin. A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. Further, for targeted insertion of sequences not normally present in a region of interest, said sequences can be present in a donor nucleic acid molecule and flanked by regions of homology to sequence in the region of interest.

[0135] As with nucleases, the donors can be introduced in any form. In certain embodiments, the donors are introduced in mRNA form to eliminate residual virus in the modified cells. In other embodiments, the donors may be introduced using DNA and/or viral vectors by methods known in the art. *See, e.g.*, U.S. Patent Publication Nos. 20100047805 and 20110207221. The donor may be introduced into the cell in circular or linear form. If introduced in linear form, the ends of the donor sequence can be protected (*e.g.*, from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are

Date Reçue/Date Received 2021-09-17

added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. *See*, for example, Chang *et al.* (1987) *Proc. Natl. Acad. Sci. USA*84:4959-4963; Nehls *et al.* (1996) *Science*272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues.

[0136] In certain embodiments, the donor includes sequences (*e.g.*, coding sequences, also referred to as transgenes) greater than 1 kb in length, for example between 2 and 200 kb, between 2and 10kb (or any value therebetween). The donor may also include at least one nuclease target site. In certain embodiments, the donor includes at least 2 target sites, for example for a pair of ZFNs, TALENs, TtAgo or CRISPR/Cas nucleases. Typically, the nuclease target sites are outside the transgene sequences, for example, 5' and/or 3' to the transgene sequences, for cleavage of the transgene. The nuclease cleavage site(s) may be for any nuclease(s). In certain embodiments, the nuclease target site(s) contained in the double-stranded donor are for the same nuclease(s) used to cleave the endogenous target into which the cleaved donor is integrated via homology-independent methods.

[0137] The donor can be inserted so that its expression is driven by the endogenous promoter at the integration site, namely the promoter that drives expression of the endogenous gene into which the donor is inserted. However, it will be apparent that the donor may comprise a promoter and/or enhancer, for example a constitutive promoter or an inducible or tissue specific promoter. The donor molecule may be inserted into an endogenous gene such that all, some or none of the endogenous gene is expressed. Furthermore, although not required for expression, exogenous sequences may also include transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals.

[0138] The transgenes carried on the donor sequences described herein may be isolated from plasmids, cells or other sources using standard techniques known in the art such as PCR. Donors for use can include varying types of topology, including circular supercoiled, circular relaxed, linear and the like. Alternatively, they may be chemically synthesized using standard oligonucleotide synthesis techniques. In

Date Reçue/Date Received 2021-09-17

addition, donors may be methylated or lack methylation. Donors may be in the form of bacterial or yeast artificial chromosomes (BACs or YACs).

[0139] The donor polynucleotides described herein may include one or more non-natural bases and/or backbones. In particular, insertion of a donor molecule with methylated cytosines may be carried out using the methods described herein to achieve a state of transcriptional quiescence in a region of interest.

[0140] The exogenous (donor) polynucleotide may comprise any sequence of interest (exogenous sequence). Exemplary exogenous sequences include, but are not limited to any polypeptide coding sequence (*e.g.*, cDNAs), promoter sequences, enhancer sequences, epitope tags, marker genes, cleavage enzyme recognition sites and various types of expression constructs. Marker genes include, but are not limited to, sequences encoding proteins that mediate antibiotic resistance (*e.g.*, ampicillin resistance, neomycin resistance, G418 resistance, puromycin resistance), sequences encoding colored or fluorescent or luminescent proteins (*e.g.*, green fluorescent protein, enhanced green fluorescent protein, red fluorescent protein, luciferase), and proteins which mediate enhanced cell growth and/or gene amplification (*e.g.*, dihydrofolate reductase). Epitope tags include, for example, one or more copies of FLAG, His, myc, Tap, HA or any detectable amino acid sequence.

[0141] In some embodiments, the donor further comprises a polynucleotide encoding any polypeptide of which expression in the cell is desired, including, but not limited to antibodies, antigens, enzymes, receptors (cell surface or nuclear), hormones, lymphokines, cytokines, reporter polypeptides, growth factors, and functional fragments of any of the above. The coding sequences may be, for example, cDNAs.

[0142] In certain embodiments, the exogenous sequences can comprise a marker gene (described above), allowing selection of cells that have undergone targeted integration, and a linked sequence encoding an additional functionality. Non-limiting examples of marker genes include GFP, drug selection marker(s) and the like.

[0143] In certain embodiments, the transgene may include, for example, wild-type genes to replace mutated endogenous sequences. For example, a wild-type (or other functional) gene sequence may be inserted into the genome of a stem cell in which the endogenous copy of the gene is mutated. The transgene may be inserted at the endogenous locus, or may alternatively be targeted to a safe harbor locus.

Date Reçue/Date Received 2021-09-17

[0144] Construction of such expression cassettes, following the teachings of the present specification, utilizes methodologies well known in the art of molecular biology (see, for example, Ausubel or Maniatis). Before use of the expression cassette to generate a transgenic animal, the responsiveness of the expression cassette to the stress-inducer associated with selected control elements can be tested by introducing the expression cassette into a suitable cell line (*e.g.*, primary cells, transformed cells, or immortalized cell lines).

[0145] Furthermore, although not required for expression, exogenous sequences may also transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals. Further, the control elements of the genes of interest can be operably linked to reporter genes to create chimeric genes (*e.g.*, reporter expression cassettes). Exemplary splice acceptor site sequences are known to those of skill in the art and include, by way of example only, CTGACCTCTTCTCTTCCTCCCACAG, (SEQ ID NO:302) (from the human *HBB* gene) and TTTCTCTCCACAG (SEQ ID NO:303) (from the human Immunoglobulin-gamma gene).

[0146] Targeted insertion of non-coding nucleic acid sequence may also be achieved. Sequences encoding antisense RNAs, RNAi, shRNAs and micro RNAs (miRNAs) may also be used for targeted insertions.

[0147] In additional embodiments, the donor nucleic acid may comprise non-coding sequences that are specific target sites for additional nuclease designs. Subsequently, additional nucleases may be expressed in cells such that the original donor molecule is cleaved and modified by insertion of another donor molecule of interest. In this way, reiterative integrations of donor molecules may be generated allowing for trait stacking at a particular locus of interest or at a safe harbor locus.

Delivery

[0148] The nucleases, polynucleotides encoding these nucleases, donor polynucleotides and compositions comprising the proteins and/or polynucleotides described herein may be delivered by any suitable means. In certain embodiments, the nucleases and/or donors are delivered *in vivo*. In other embodiments, the nucleases and/or donors are delivered to isolated cells (*e.g.*, autologous or

Date Reçue/Date Received 2021-09-17

heterologous stem cells) for the provision of modified cells useful in *ex vivo* delivery to patients.

[0149] Methods of delivering nucleases as described herein are described, for example, in U.S. Patent Nos. 7,888,121; 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824.

[0150] Nucleases and/or donor constructs as described herein may also be delivered using any nucleic acid delivery mechanism, including naked DNA and/or RNA (*e.g.*, mRNA) and vectors containing sequences encoding one or more of the components. Any vector systems may be used including, but not limited to, plasmid vectors, DNA minicircles, retroviral vectors, lentiviral vectors, adenovirus vectors, poxvirus vectors; herpesvirus vectors and adeno-associated virus vectors, etc., and combinations thereof. *See, also,* U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, and U.S. Patent Application No. 14/271,008. Furthermore, it will be apparent that any of these systems may comprise one or more of the sequences needed for treatment. Thus, when one or more nucleases and a donor construct are introduced into the cell, the nucleases and/or donor polynucleotide may be carried on the same delivery system or on different delivery mechanisms. When multiple systems are used, each delivery mechanism may comprise a sequence encoding one or multiple nucleases and/or donor constructs (*e.g.*, mRNA encoding one or more nucleases and/or mRNA or AAV carrying one or more donor constructs).

[0151] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding nucleases and donor constructs in cells (*e.g.*, mammalian cells) and target tissues. Non-viral vector delivery systems include DNA plasmids, DNA minicircles, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell.

[0152] Methods of non-viral delivery of nucleic acids include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, nanoparticles, polycation or lipid:nucleic acid conjugates, naked DNA, naked RNA, capped RNA, artificial virions, and agent-enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

Date Reçue/Date Received 2021-09-17

[0153] Additional exemplary nucleic acid delivery systems include those provided by Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) and Copernicus Therapeutics Inc, (*see* for example U.S. Patent No. 6,008,336). Lipofection is described in *e.g.*, U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (*e.g.*, Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024.

[0154] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered CRISPR/Cas systems take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to subjects (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to subjects (*ex vivo*). Conventional viral based systems for the delivery of CRISPR/Cas systems include, but are not limited to, retroviral, lentivirus, adenoviral, adeno-associated, vaccinia and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0155] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system depends on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof.

Date Reçue/Date Received 2021-09-17

[0156] In applications in which transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and high levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*see, e.g.*, West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989). Any AAV serotype can be used, including AAV1, AAV3, AAV4, AAV5, AAV6 and AAV8, AAV 8.2, AAV9, and AAV rh10 and pseudotyped AAV such as AAV2/8, AAV2/5 and AAV2/6.

[0157] At least six viral vector approaches are currently available for gene transfer in clinical trials, which utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

[0158] pLASN and MFG-S are examples of retroviral vectors that have been used in clinical trials (Dunbar *et al.*, *Blood* 85:3048-305 (1995); Kohn *et al.*, *Nat. Med.* 1:1017-102 (1995); Malech *et al.*, *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al.*, *Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al.*, *Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al.*, *Hum. Gene Ther.* 1:111-2 (1997).

[0159] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 base pair (bp) inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al.*, *Lancet* 351:9117 1702-3 (1998), Kearns *et al.*, *Gene*

Date Reçue/Date Received 2021-09-17

Ther. 9:748-55 (1996)). Other AAV serotypes, including AAV1, AAV3, AAV4, AAV5, AAV6, AAV8, AAV9 and AAVrh10, and all variants thereof, can also be used in accordance with the present invention.

[0160] Replication-deficient recombinant adenoviral vectors (Ad) can be produced at high titer and readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; subsequently the replication defective vector is propagated in human 293 cells that supply deleted gene function in *trans*. Ad vectors can transduce multiple types of tissues *in vivo*, including non-dividing, differentiated cells such as those found in liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for anti-tumor immunization with intramuscular injection (Sterman *et al., Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al., Infection* 24:1 5-10 (1996); Sterman *et al., Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al., Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al., Hum. Gene Ther.* 5:597-613 (1997); Topf *et al., Gene Ther.* 5:507-513 (1998); Sterman *et al., Hum. Gene Ther.* 7:1083-1089 (1998).

[0161] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host (if applicable), other viral sequences being replaced by an expression cassette encoding the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts

Date Reçue/Date Received 2021-09-17

due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

[0162] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Accordingly, a viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *Proc. Natl. Acad. Sci. USA* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other virus-target cell pairs, in which the target cell expresses a receptor and the virus expresses a fusion protein comprising a ligand for the cell-surface receptor. For example, filamentous phage can be engineered to display antibody fragments (*e.g.*, FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences which favor uptake by specific target cells.

[0163] Gene therapy vectors can be delivered *in vivo* by administration to an individual subject, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0164] Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, etc.) containing nucleases and/or donor constructs can also be administered directly to an organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and,

Date Reçue/Date Received 2021-09-17

although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0165] Vectors suitable for introduction of polynucleotides described herein include non-integrating lentivirus vectors (IDLV). *See*, for example, Ory *et al.* (1996) *Proc. Natl. Acad. Sci. USA***93:**11382-11388; Dull *et al.* (1998) *J. Virol.***72:**8463-8471; Zuffery *et al.* (1998) *J. Virol.***72:**9873-9880; Follenzi *et al.* (2000) *Nature Genetics* **25:**217-222; U.S. Patent Publication No 2009/054985.

[0166] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (*see, e.g.*, *Remington's Pharmaceutical Sciences*, 17th ed., 1989).

[0167] It will be apparent that the nuclease-encoding sequences and donor constructs can be delivered using the same or different systems. For example, a donor polynucleotide can be carried by an AAV, while the one or more nucleases can be carried by mRNA. Furthermore, the different systems can be administered by the same or different routes (intramuscular injection, tail vein injection, other intravenous injection, intraperitoneal administration and/or intramuscular injection. The vectors can be delivered simultaneously or in any sequential order.

[0168] Formulations for both *ex vivo* and *in vivo* administrations include suspensions in liquid or emulsified liquids. The active ingredients often are mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients include, for example, water, saline, dextrose, glycerol, ethanol or the like, and combinations thereof. In addition, the composition may contain minor amounts of auxiliary substances, such as, wetting or emulsifying agents, pH buffering agents, stabilizing agents or other reagents that enhance the effectiveness of the pharmaceutical composition.

Kits

[0169] Also provided are kits comprising any of the linkers described herein and/or for performing any of the above methods. The kits typically contain a linker sequence as described herein (or a polynucleotide encoding a linker as described herein). The kit may supply the linker alone or may provide vectors into which a

Date Reçue/Date Received 2021-09-17

DNA-binding domain and/or nuclease of choice can be readily inserted into. The kits can also contain cells, buffers for transformation of cells, culture media for cells, and/or buffers for performing assays. Typically, the kits also contain a label which includes any material such as instructions, packaging or advertising leaflet that is attached to or otherwise accompanies the other components of the kit.

Applications

[0170] The disclosed linkers are advantageously used to link with engineered DNA-binding domains with cleavage domains to form nucleases for cleaving DNA. The linkers as described herein allow for the cleavage of DNA when the target sites of a pair of nucleases used for cleavage are of variable spacings, for example target sites that are not 5 or 6 base pairs apart (*e.g.*, 7, 8, 9 or more base pairs apart). Cleavage can be at a region of interest in cellular chromatin (*e.g.*, at a desired or predetermined site in a genome, for example, in a gene, either mutant or wild-type); to replace a genomic sequence (*e.g.*, a region of interest in cellular chromatin) with a homologous non-identical sequence (*i.e.*, targeted recombination); to delete a genomic sequence by cleaving DNA at one or more sites in the genome, which cleavage sites are then joined by non-homologous end joining (NHEJ); to screen for cellular factors that facilitate homologous recombination; and/or to replace a wild-type sequence with a mutant sequence, or to convert one allele to a different allele. Such methods are described in detail, for example, in U.S. Patent No. 7,888,121.

[0171] Accordingly, the disclosed linkers can be used in any nuclease for any method in which specifically targeted cleavage is desirable and/or to replace any genomic sequence with a homologous, non-identical sequence. For example, a mutant genomic sequence can be replaced by its wild-type counterpart, thereby providing methods for treatment of *e.g.*, genetic disease, inherited disorders, cancer, and autoimmune disease. In like fashion, one allele of a gene can be replaced by a different allele using the methods of targeted recombination disclosed herein. Indeed, any pathology dependent upon a particular genomic sequence, in any fashion, can be corrected or alleviated using the methods and compositions disclosed herein.

[0172] Exemplary genetic diseases include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, Alsheimer's disease, androgen insensitivity

Date Reçue/Date Received 2021-09-17

syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangictasia, barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercum's disease, ectodermal dysplasia, fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis, Gaucher's disease, generalized gangliosidoses (*e.g.*, GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefleter syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920), leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipdius, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, Parkinson's disease, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, X-linked lymphoproliferative syndrome (XLP, OMIM No. 308240) and X-linked SCID.

[0173] Additional exemplary diseases that can be treated by targeted DNA cleavage and/or homologous recombination include acquired immunodeficiencies, lysosomal storage diseases (*e.g.*, Gaucher's disease, GM1, Fabry disease and Tay-Sachs disease), mucopolysaccahidosis (*e.g.* Hunter's disease, Hurler's disease), hemoglobinopathies (*e.g.*, sickle cell diseases, HbC, α-thalassemia, β-thalassemia) and hemophilias.

[0174] Targeted cleavage of infecting or integrated viral genomes can be used to treat viral infections in a host. Additionally, targeted cleavage of genes encoding receptors for viruses can be used to block expression of such receptors, thereby preventing viral infection and/or viral spread in a host organism. Targeted mutagenesis of genes encoding viral receptors (*e.g.*, the CCR5 and CXCR4 receptors for HIV) can be used to render the receptors unable to bind to virus, thereby preventing new infection and blocking the spread of existing infections. *See*,

Date Reçue/Date Received 2021-09-17

International Patent Publication WO 2007/139982. Non-limiting examples of viruses or viral receptors that may be targeted include herpes simplex virus (HSV), such as HSV-1 and HSV-2, varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV), HHV6 and HHV7. The hepatitis family of viruses includes hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV) and hepatitis G virus (HGV). Other viruses or their receptors may be targeted, including, but not limited to, Picornaviridae (*e.g.*, polioviruses, etc.); Caliciviridae; Togaviridae (*e.g.*, rubella virus, dengue virus, *etc.*); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (*e.g.*, rabies virus, *etc.*); Filoviridae; Paramyxoviridae (*e.g.*, mumps virus, measles virus, respiratory syncytial virus, *etc.*); Orthomyxoviridae (*e.g.*, influenza virus types A, B and C, *etc.*); Bunyaviridae; Arenaviridae; Retroviradae; lentiviruses (*e.g.*, HTLV-I; HTLV-II; HIV-1 (also known as HTLV-III, LAV, ARV, hTLR, *etc.*) HIV-II); simian immunodeficiency virus (SIV), human papillomavirus (HPV), influenza virus and the tick-borne encephalitis viruses. *See, e.g.* Virology, 3rd Edition (W. K. Joklik ed. 1988); Fundamental Virology, 2nd Edition (B. N. Fields and D. M. Knipe, eds. 1991), for a description of these and other viruses. Receptors for HIV, for example, include CCR-5 and CXCR-4.

[0175] Nucleases containing the disclosed linkers can also be used for inactivation (partial or complete) of one or more genomic sequences. Inactivation can be achieved, for example, by a single cleavage event, by cleavage followed by non-homologous end joining, by cleavage at two sites followed by joining so as to delete the sequence between the two cleavage sites, by targeted recombination of a missense or nonsense codon into the coding region, by targeted recombination of an irrelevant sequence (*i.e.*, a "stuffer" sequence) into the gene or its regulatory region, so as to disrupt the gene or regulatory region, or by targeting recombination of a splice acceptor sequence into an intron to cause mis-splicing of the transcript.

[0176] Nuclease-mediated inactivation (*e.g.*, knockout) of endogenous genes can be used, for example, to generate cell lines deficient in genes involved in apoptosis or protein production (e.g., post-translational modifications such as fucosylation). ZFN-mediated inactivation can also be used to generate transgenic organisms (*e.g.*, plants, rodents and rabbits).

[0177] In addition, because nucleases don't appear to have specificity for the DNA sequence between the two paired half sites, nucleases with linkers as described

Date Reçue/Date Received 2021-09-17

herein can be designed to cleave DNA such that the resulting single-stranded overhangs have any desired sequence. In particular, linkers as described herein can be designed to influence both the size and position of these single-stranded overhangs with respect to the starting sequence. Thus, when incorporated into one or more nucleases of a nuclease pair, linkers as described herein can result in more uniform ends following cleavage. Accordingly, the linkers described herein can also be used to more efficiently clone DNA cut with nucleases, which is broadly applicable in many areas of biotechnology and basic science.

[0178] Thus, the linkers described herein provide broad utility for improving nuclease-mediated cleavage in gene modification applications. Linkers as described herein may be readily incorporated into any existing nuclease by either site directed mutagenesis or subcloning to be used in many applications in standard cloning, constructing large genomes for synthetic biology, new types of RFLP analysis of large sequences or even allow new types of cloning involving extremely large DNA sequences. The potential properties of nucleases with rigid linkers could also be ideal in applications such as DNA computing.

[0179] The following Examples relate to exemplary embodiments of the present disclosure in which the nuclease comprises one or more ZFNs. It will be appreciated that this is for purposes of exemplification only and that other nucleases can be used, for instance TALENs, homing endonucleases (meganucleases) with engineered DNA-binding domains and/or fusions of naturally occurring of engineered homing endonucleases (meganucleases) DNA-binding domains and heterologous cleavage domains, and nuclease systems such as TtAgo and CRISPR/Cas using engineered single guide RNAs.

EXAMPLES

Example 1: Design and construction of ZFNs with rigid linkers

[0180] Zinc finger nuclease constructs targeted to the human CCR5 locus were prepared as disclosed in U.S. Patent No. 7,951,925. "Wild-type constructs" included the "ZC" linker.

[0181] In addition, pairs of ZFNs targeted to sequences in the human mitochondria containing the mutation that causes MELAS (mitochondrial myopathy,

Date Reçue/Date Received 2021-09-17

encephalopathy, lactic acidosis, and stroke) were also prepared to include the L7a or L6a linker as described in U.S. Patent Publication No. 20090305419.

Example 2: ZFN activity

A. CCR5-targeted ZFNs

[0182] Constructs encoding CCR5-targeted ZFN SBS #8266 were initially tested in a yeast Mel-I reporter system as described in U.S. Patent No. 8,563,314. In particular, yeast strains having an inverted repeat of the SBS #8266 target site separated by 3, 4, 5, 6, 7, or 8 bp were used to characterize the constructs.

[0183] The wild-type ZFN (with the standard LRGSQLVKSELEEKKS (SEQ ID NO:301) linker showed strong activity with 5 bp and 6 bp half site spacings. In addition, the constructs with the L6a linker sequence (FIG. 2) showed activity at 6p spacings and the L7a linker sequence showed significant activity with 7 bp and 8 bp spacings.

[0184] *In vitro* DNA binding and cleavage activity of the MELAS-targeted ZFNs was also assayed and pairs of ZFNs including the rigid L7a linker cleaved their target.

[0185] Finally, the CCR5 ZFNs including the L7a linker were tested for NHEJ activity at the endogenous human CCR5 locus in cell lines that contain various numbers of base pairs between the half sites. Results are shown in Table 1.

Table 1

ZFN	target sites separated by	% NHEJ (exp't #1)	% NHEJ (exp't #2)	% NHEJ (average)
Wt ZFNs	4 bp	1.2	1.1	1.2
Wt ZFNs	5 bp	36.0	34.0	35.0
Wt ZFNs	6 bp	13.4	8.8	11.1
Wt ZFNs	7 bp	0.0	0.0	0.0
Wt ZFNs	8 bp	0.0	0.0	0.0
L6a ZFNs	4 bp	0.0	0.0	0.0
L6a ZFNs	5 bp	44.4	34.1	39.3
L6a ZFNs	6 bp	26.2	24.6	25.4
L6a ZFNs	7 bp	6.5	3.7	5.1

Date Reçue/Date Received 2021-09-17

L6a ZFNs	8 bp	0.0	0.0	0.0
L7a ZFNs	4 bp	0.0	0.0	0.0
L7a ZFNs	5 bp	0.0	0.0	0.0
L7a ZFNs	6 bp	33.1	30.5	31.8
L7a ZFNs	7 bp	41.1	38.1	39.6
L7a ZFNs	8 bp	7.9	4.6	6.1

[0186] As expected, the wild-type ZFNs only showed high activity at half-sites separated by 5 or 6 bp. However, CCR5-targeted ZFNs including the rigid L7a linker showed high activity with a 7 bp spacing and noticeable activity with the 8 bp spacing. It should be noted that the efficiency of the L7a constructs with the 7 bp spacing is very similar to the efficiency of the wild type ZFNs with a 5 bp spacing (either in the wild-type cell line or a cell line with a different sequence of the 5 bp in between the half sites).

[0187] In addition, combinations of linkers were also tested in CCR5-targeted ZFN pairs. Briefly, K562 cells were engineered to have gaps of 4 to 8 base pairs (bp) between the CCR5 ZFN binding sites. Two CCR5 ZFNs with different linkers combinations (Wt/L7a) were transfected into these K562 cells by Amaxa Shuttle. Samples were harvested 3 days after transfection and subjected to CELI-I assay analysis. CEL-I mismatch assays were performed essentially as per the manufacturer's instructions (Trangenomic SURVEYOR™).

[0188] The results indicate that the Wt/Wt linker ZFN has the highest activity with 5bp gap target sequence; the L7a/L7a linker ZFN had the highest activity with a 7bp gap sequence, and the ZFNs with Wt/L7a or L7a/Wt linker combinations had the highest activity with a 6bp gap sequence.

B. ROSA-targeted ZFNs

[0189] Neuro2A cells were transfected with combinations of mROSA-targeted ZFNs (*see, e.g.*, U.S. Patent Publication No. 2007/0134796) by Amaxa Shuttle using a target site with a 6bp gap. One ZFN of the pairs included a wild-type linker ("ZC") and the other included either wild-type or L7a linker as described

Date Reçue/Date Received 2021-09-17

herein. Samples were harvested 3 days after transfection and subjected to CEL-I analysis, as described above.

[0190] As shown in Table 2 below, the wild type (WT)/L7a linker in a pair of ZFNs is active with a 6 bp gap.

Table 2

Sample	Linker #1	Linker #2	%NHEJ
mock transfection (no ZFN)	NA	NA	0.4
Rosa-ZFN pairs	Wt	Wt	22.6
Rosa-ZFN pairs	Wt	L7a	7.5
Rosa-ZFN pairs	Wt	Wt	23.2
Rosa-ZFN pairs	Wt	Wt	18.7
GFP-ZFN pairs	Wt	L7a	5.3
GFP-ZFN pairs	Wt	Wt	21.5

C. Rat IgM

[0191] Rat C6 cells were transfected with combinations of rat IgM-targeted ZFNs (see, *e.g.*, U.S. Publication No. 20100218264) by Amaxa Shuttle using a target site with a 6 bp gap. One ZFN of the pairs included a wild-type linker ("ZC") and the other included either wild-type or L7a linker. Samples were harvested 9 days after transfection and subjected to CEL-I analysis, as described above and in U.S. Patent Publication No. 2007/0134796.

[0192] Cells containing the pair of ZFNs that included the L7a linkers showed 2.43% NHEJ as compared to cells containing a pair of ZFNs that included the ZC linker, which showed 1.93% NHEJ. Furthermore, the L7a-containing linker ZFN pair was used to inject into rat ES cells (as described in U.S. Publication No. 20100218264) and these ES cells successfully produced homozygous IgM gene knockout rat offspring.

Example 3: Additional Linker Designs

[0193] Additional linkers were generated from a bacterial selection system that was modified from Barbas *et al.* (2010) *J. Mol. Biol.* 400:96-107. Briefly, bacteria were transformed with a ZFN-encoding plasmid (expression of the ZFN driven by the arabinose inducible promoter) and a plasmid that expressed the bacterial

Date Reçue/Date Received 2021-09-17

ccdB toxin from the T7 promoter and included the ZFN target sites. *See,* Figure 2. This system allowed the ability to query very large libraries with complexities of ~10^8 and a high stringency option (> 100 cleavage events required for survival).

[0194] The expressed ZFN cleaved pTox leading to degradation of pTox; the ccdB toxin was then switched on for cell killing; the survivors (ZFN-cleaved pTox) were amplified and the genes encoding the linkers re-cloned into new plasmids for the next cycle. *See,* Figures 2A and 2B. The ZFNs included fully randomized linkers of between 8 and 17 amino acids in length between the zinc finger protein domain and the cleavage (*Fok*I) domain. *See,* Figure 2A.

[0195] Engineered cleavage domains and canonical (C2H2) and non-canonical finger structures (*see,* U.S. Patent Publication No. 20080182332) may be used; for this study wild-type *Fok*I domains (which form homodimers) and CCHC finger structures were employed. Linkers were selected on gaps between ZFN target sequences of 5-16 base pairs.

[0196] To select for active pZFNs from the vast excess of inactive ones, the following steps were performed: (i) High-complexity DNA libraries were constructed, *see* Figure 2A, (ii) plasmid DNA library was transformed into cells bearing pTox_8196-bs; (iii) expressed ZFNs for 2 hours; (iv) induced ccdB overnight; (v) minipreped plasmids and subclone linker DNA into new pZFN; repeated (ii) – (v) for 7 more cycles.

[0197] Exemplary functional linkers obtained from the selections are shown in Figure 4. A summary of the length distribution of the linkers which cleaved at the indicated gaps is shown in Figure 5.

Example 4: Verifying Activity and Portability Studies

[0198] Selected candidates (ZFNs with linkers) were screened for modification of an endogenous target in six variant K562 cell lines, which were engineered at the CCR5 locus to replace the gap between heterodimeric target sites (8166 and 8266 ZFNs as described in U.S. Patent No. 7,951,925) with new gaps of 5-16 base pairs (*e.g.*, 5, 6, 7, 8, 15 or 16 bp). Linker cassettes (~80/selection = 480 total) were substituted for L0 linkers in CCR5 ZFN expression vectors with the CCHC structure and constructs were transfected into appropriate K562 cells. *See,* Figure 6. Activity was assessed using a Cel1 assay, as described above and in U.S. Patent Publication No. 2007/0134796.

Date Reçue/Date Received 2021-09-17

[0199] Figure 8 shows a summary of the results obtained with the selected linkers with the indicated target gaps. Figures 9A and 9B show the most active linkers for a 5-8 base pair gap between target sites. "Indels" refers to insertions or deletions, usually small, within the target sequence following ZFN-mediated modification (*e.g.*, following cleavage and NHEJ repair). Linkers were also tested for activity with different gap spacings. Figure 9C shows the gap preference for the indicated linkers.

[0200] For portability studies, four L8 linkers were chosen for cloning and testing with a large target gene to ensure consistent high design scores for all test sets. A schematic of the vector designs is shown in Figure 10A. Sets of the 32 ZFNs were generated for each of L0 (5, 6 bp gap), L7a (7 bp gap), and L8 (8 bp gap) linkers for a total of 92 ZFNs. See, Figure 7 for L8 linkers. Activity was verified by Cel-1 assay, as described above and in U.S. Patent Publication No. 2007/0134796.

[0201] Figure 10B shows a summary of the results of the portability studies and demonstrates many ZFNs tested were active. The number active and level of modification of the selected linkers was comparable to previously used linkers at a 5 or 6 (L0) or 7 bp spacing (L7a).

[0202] The L8c linkers were also additionally modified. A different target was chosen and eight pairs of C2H2 ZFNs were used to select for activity for 8 bp (termed A, B, C, D) and 9 bp (termed E, F, G, and H) gaps against their specific targets. The linkers in the ZFN pairs were altered with 4 different linker types (termed L8n, L8o, L8m and L8p) and the pairs were selected for activity against target sites where the gaps between the ZFN binding sites were separated by 8 or 9 bp. Additionally, the original L8c linker was used in the pairs to attach either a wt *Fok*I nuclease domain or enhanced obligated heterodimeric *Fok*I nuclease domains (eHiFi) for comparison. A GFP expression vector was used as a transfection control. The peptide sequences of the original L8c linker and the 4 modified linkers are shown below (linker sequences underlined):

Group 1: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[wt FokI] (SEQ ID NO:229)
Group 2: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI] (SEQ ID NO:229)
Group 3: L8n [ZFP]HTKIH-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI] (SEQ ID NO:230)
Group 4: L8o [ZFP]HTKIH-GGSYAPMPPLALASPELEEKKSEL[eHiFi FokI] (SEQ ID NO:231)
Group 5: L8m [ZFP]HTKIH--GSYAPMPPLALASPELEEKKSEL[eHiFi FokI] (SEQ ID NO:232)
Group 6: L8p [ZFP]HTKIHLRGSYAPMPPLALASPELEEKKSEL[eHiFi FokI] (SEQ ID NO:233)

Date Reçue/Date Received 2021-09-17

[0203] The results (see Figures 11A and 11B) demonstrated that the L8c type linker worked well in both the CCHC and C2H2 ZFP backgrounds. Additionally, the L8p linker worked well with both 8 bp and 9 bp gaps.

Example 5: Further linker studies

[0204] The "L7a" linker is currently used in as the ZFP-FokI linker to target the 7-bp gap sites in both C2H2 and CCHC ZFP architectures. Linkers shown in Figure 9A and 9B were identified within a CCHC ZFP architecture so further studies were conducted in C2H2 architecture of ZFPs that target 7-bp gap sites.

[0205] Four linker peptide sequences as shown in Figure 9A (L7-1, L7-3, L8a, and L8c) were initially selected for these studies, including these linkers with and without one or more optional residues as shown in parentheses as follows: (H)LPKPANPFPLD (SEQ ID NO:199); (D)PNSPISRARPLNPHP (SEQ ID NO:202); (N)GICPPPRPRTSPP (SEQ ID NO:194); and (N)GSYAPMPPLALASP (SEQ ID NO:196).

[0206] A pair of ZFPs that targets a 7-bp gap site was used to test the linkers. This pair of ZFNs was engineered in the context of CCHC or C2H2 architectures, with various linker sequences, and fused to either wild-type ("wt") or enhanced obligated heterodimeric FokI ("eHiFi") domains. *See, e.g.,* U.S. Patent Nos. 8,623,618; 7,888,121; 7,914,796; and 8,034,598 and U.S. Publication No. 20110201055 for exemplary eHiFi domains. These ZFN pairs were tested in K562 cells and their nuclease cleavage activities were measured by Cel1 assay and MiSeq sequencing.

[0207] The results, as shown in Figure 12, demonstrated that (1) these linkers are effective in both wild type and eHiFi FokI architectures (comparing Group 1 and Group 2); (2) these linkers are effective in the context of both C3H and C2H2 architecture (comparing Group 2 and Group 3); and (3) removing one amino acid residue (from the N-terminal) in each linker (shown in Group 3) resulted in significantly higher activity as compared to linkers with the C-terminal residue and, in most instance, as compared to the currently-used L7a linker (sample #17 in Figure 12).

[0208] Subsequently, the two most active 7-bp linker sequences ("L7c3" and "L8c3", indicated by arrows) were modified to make the C-terminus of a ZFP ends

Date Reçue/Date Received 2021-09-17

with the conventional "HTKIHLRGS" peptide sequence, which allows ZFN assembly using the same process as all other C2H2 ZFNs.

[0209] To ensure the modified linkers possessed the same or higher activities than the conventional "L7a" linker on a 7-bp gap site, four linker variants were designed for both "L7c3" and "L8c3" and engineered into two ZFN pairs (see Figure 13). These ZFNs were tested in human CD34 cells and their NHEJ activities were measured with MiSeq analyses.

[0210] As shown in Figure 13, while all variants tested were active, many of them showed higher activities than the conventional "L7a" bearing ZFNs, including the highest activity as seen with "L7c5" and "L7e4" variants.

Example 6: Optimizing ZFN activity by changing the ZFP-FokI junction sequences

[0211] The conventional junction sequence between ZFP and FokI nuclease domain is "---HTKIHLRGSQLVKSELEEK---" (SEQ ID NO:259). Since ZFN activity can be affected by many factors including the gap length between the two ZFP binding sites and the ZFP affinity to its binding sites, we tested whether the cleavage efficiency of a pair of ZFNs could be modulated by changing the length and/or compositions at the junction sites to optimize the heterodimer interaction between the two FokI cleavage domains. *See,* Figure 14. ZFNs of C2H2 or C3H structure can be used.

[0212] In particular, a series of 24 junction variants were designed and engineered into a pair of ZFNs target to a 6-bp gap site. *See,* Figure 15. Some junction variants have the same length as the conventional one but with different (substituted) amino acid residue(s). In addition, these variants also contain additional amino acids or have some amino acids removed (as shown in Figure 15).

[0213] These ZFNs were tested in K562 cells in the following ways: (1) the 24 left ZFN variants were paired with the right ZFN containing the conventional linker, (b) the 24 right ZFN variants were paired with the left ZFN containing the conventional linker, and (c) the 24 left ZFN variants were paired with each of the right ZFN variants. The ZFN NHEJ activities were measured by MiSeq.

[0214] As shown in Figure 15, ZFN activities were affected by pairing ZFNs containing the various junction linkers. Notably, many of the modified junction

Date Reçue/Date Received 2021-09-17

regions showed significantly higher NHEJ activities when comparing to the pair of ZFNs containing both of the conventional junction sequences.

[0215] Selected junction sequences were also tested using ZFN pairs that target a different site. Two ZFNs pairs with the junction variants were transfected into human HepG2 cells and their NHEJ activities were measured by MiSeq. The data as shown in Figure 16 demonstrated again that the ZFN activities can be further improved by modulating the ZFP-FokI junction sequences.

[0216] Although disclosure has been provided in some detail by way of illustration and example for the purposes of clarity of understanding, it will be apparent to those skilled in the art that various changes and modifications can be practiced without departing from the scope of the disclosure. Accordingly, the foregoing descriptions and examples should not be construed as limiting.

Date Reçue/Date Received 2021-09-17

EMBODIMENTS

Embodiment 1. A fusion protein comprising a DNA-binding domain having an N-terminus and a C-terminus, wherein the DNA-binding domain binds to a nucleotide target site; a *FokI* cleavage domain having an N-terminus and a C-terminus; and a linker between the C-terminus of the DNA-binding domain and the N-terminus of the cleavage domain, wherein the linker comprises a sequence selected from the group consisting of PKPAN (SEQ ID NO:289), RARPLN (SEQ ID NO:290); PMPPLA (SEQ ID NO:291) or PPPRP (SEQ ID NO:292).

Embodiment 2. The fusion protein of Embodiment 1, wherein the linker further comprises a ZC sequence (SEQ ID NO:2).

Embodiment 3. The fusion protein of Embodiment 2, wherein the ZC sequence is at the N-terminus of the linker.

Embodiment 4. The fusion protein of any of Embodiments 1 to 3, wherein the DNA-binding domain is a zinc finger protein.

Embodiment 5. A dimer comprising two fusion proteins according to any of Embodiments 1 to 4.

Embodiment 6. The dimer of Embodiment 5, wherein the dimer is a homodimer or heterodimer.

Embodiment 7. A polynucleotide encoding at least one fusion protein according to any of Embodiments 1 to 4.

Embodiment 8. A cell comprising a fusion protein according to any of Embodiments 1 to 4, a dimer according to Embodiment 5 or 6 or a polynucleotide according to Embodiment 7.

Date Reçue/Date Received 2021-09-17

Embodiment 9. A method for targeted cleavage of cellular chromatin in a region of interest in a cell, the method comprising: expressing a pair of nucleases in the cell under conditions such that cellular chromatin is cleaved at the region of interest, wherein the nucleases bind to target sites in the region of interest and further wherein at least one nuclease comprises a fusion protein according to any of Embodiments 1 to 4.

Embodiment 10. The method of Embodiment 9, wherein both nucleases comprise fusions proteins according to any of Embodiments 1 to 4.

Embodiment 11. The method of Embodiment 9 or 10, wherein the target sites for the zinc finger nucleases are 3 to 20 base pairs apart.

Embodiment 12. The method of any of Embodiments 9 to 11, further comprising the step of introducing a donor polynucleotide into the cell, wherein all or part of the donor polynucleotide is incorporated into the region of interest following cleavage.

Embodiment 13. A kit for producing a nuclease, the kit comprising a fusion protein according to any of Embodiments 1 to 4 or a polynucleotide according to Embodiment 7 contained in one or more containers, optional hardware, and instructions for use of the kit.

Embodiment 14. The kit of Embodiment 13, further comprising a donor polynucleotide.

Embodiment 15. A fusion protein comprising a DNA-binding domain having an N-terminus and a C-terminus, wherein the DNA-binding domain binds to a nucleotide target site; a *FokI* cleavage domain having an N-terminus and a C-terminus, wherein the N-terminus includes one or more amino acid modifications selected from the group consisting of addition of one or more amino acid residues to the N-terminus; deletion of one or more amino acid residues from the N-terminus; and substitution of one or more amino acid residues in the N-terminus; and a ZC linker between the C-terminus of the DNA-binding domain and the N-terminus of the cleavage domain.

Date Reçue/Date Received 2021-09-17

Embodiment 16. The fusion protein of Embodiment 1, wherein the DNA-binding domain is a zinc finger protein.

Embodiment 17. A dimer comprising two fusion proteins according to Embodiment 15 or Embodiment 16.

Embodiment 18. The dimer of Embodiment 17, wherein the dimer is a homodimer or heterodimer.

Embodiment 19. A polynucleotide encoding at least one fusion protein according to Embodiment 15 or Embodiment 16.

Embodiment 20. A cell comprising a fusion protein according to Embodiment 15 or Embodiment 16, a dimer according to Embodiment 17 or 18 or a polynucleotide according to Embodiment 19.

Embodiment 21. A method for targeted cleavage of cellular chromatin in a region of interest in a cell, the method comprising: expressing a pair of nucleases in the cell under conditions such that cellular chromatin is cleaved at the region of interest, wherein the nucleases bind to target sites in the region of interest and further wherein at least one nuclease comprises a fusion protein according to Embodiment 15 or Embodiment 16.

Embodiment 22. The method of Embodiment 21, wherein both nucleases comprise fusions proteins according to Embodiment 15 or Embodiment 16.

Embodiment 23. The method of Embodiment 21 or 22, wherein the target sites for the zinc finger nucleases are 3 to 20 base pairs apart.

Embodiment 24. The method of any of Embodiments 21 to 23, further comprising the step of introducing a donor polynucleotide into the cell, wherein all or part of the donor polynucleotide is incorporated into the region of interest following cleavage.

Date Reçue/Date Received 2021-09-17

Embodiment 25. A kit for producing a nuclease, the kit comprising a fusion protein according to Embodiment 15 or Embodiment 16 or a polynucleotide according to Embodiment 19 contained in one or more containers, optional hardware, and instructions for use of the kit.

Embodiment 26. The kit of Embodiment 25, further comprising a donor polynucleotide.

Date Reçue/Date Received 2021-09-17

CLAIMS

1. A fusion protein comprising

a DNA-binding domain having an N-terminus and a C-terminus, wherein the DNA-binding domain binds to a nucleotide target site;

a *FokI* cleavage domain having an N-terminus and a C-terminus, the N-terminus residues of the *FokI* cleavage domain are ELEEK; and

a linker between the C-terminus of the DNA-binding domain and the N-terminus of the cleavage domain, wherein the linker comprises LRGSPISRARPLNPHP; LRGSISRARPLNPHP; LRGSPSRARPLNPHP; LRGSSRARPLNPHP; LRGSYAPMPPLALASP; LRGSAPMPPLALASP; LRGSPMPPLALASP; or LRGSMPPLALASP.

2. The fusion protein of claim 1, wherein the DNA-binding domain is a zinc finger protein, a TAL-effector domain or a Cas domain.

3. A dimer comprising two fusion proteins according to claim 1 or claim 2, wherein the dimer is a heterodimer or a homodimer.

4. One or more polynucleotides encoding at least one fusion protein according to claim 1 or claim 2.

5. A cell comprising the fusion protein according to claim 1 or claim 2, the dimer according to claim 3 and/or the one or more polynucleotides according to claim 4.

6. An *in vitro* method for targeted cleavage of cellular chromatin in a region of interest in a cell, the method comprising:

expressing a pair of nucleases in the cell under conditions such that cellular chromatin is cleaved at the region of interest, wherein the nucleases bind to target sites separated by 7 base pairs in the region of interest and further wherein at least one nuclease comprises a fusion protein according to claim 1 or claim 2.

Date Reçue/Date Received 2023-01-23

7. The *in vitro* method of claim 6, wherein both nucleases comprise fusions proteins according to claim 1 or claim 2.

8. The *in vitro* method of claim 6 or claim 7, further comprising the step of introducing a donor polynucleotide into the cell, wherein all or part of the donor polynucleotide is incorporated into the region of interest following cleavage.

9. A kit for producing a nuclease that cleaves a target site, the kit comprising the fusion protein according to claim 1 or claim 2, the one or more polynucleotides according to claim 4 and/or the cell of claim 5, contained in one or more containers, and instructions for making the nuclease using one or more of the fusion proteins or polynucleotides.

10. The kit of claim 9, further comprising a donor polynucleotide that is integrated into the target site following cleavage by the nuclease.

11. A pair of nucleases comprising at least one fusion protein according to claim 1 or claim 2 for use to cleave cellular chromatin, wherein the pair of nucleases bind to target sites separated by 7 base pairs.

12. The pair of nucleases for use according to claim 11, wherein the pair of nucleases is encoded by one or more polynucleotides.

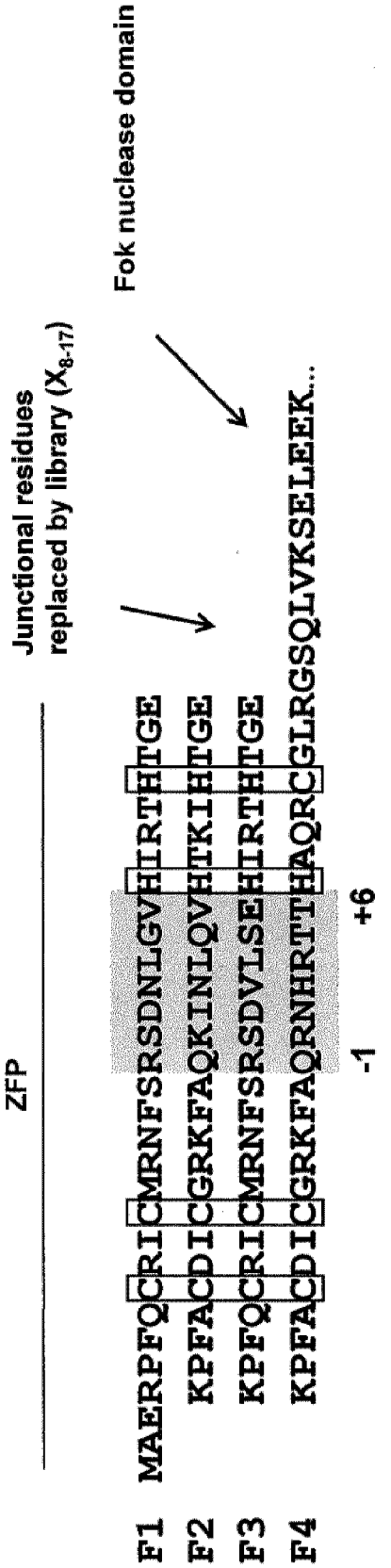
13. The pair of nucleases for use according to claim 11 or claim 12, wherein the pair of nucleases comprises two fusion proteins according to claim 1 or claim 2.

Date Reçue/Date Received 2023-01-23

FIGURE 1 (SEQ ID NO:1):

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAAMAERPFQCRICMR
NFSDRSNLSRHIRTHTGEKPFACDICGRKFAISSNLNSHTKIHTGSQKPFQCRICMRNF
SRSDNLARHIRTHTGEKPFACDICGRKFATSGNLTRHTKIHLRGS**QLVKSELEEKKS**EL
RHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYT
VGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWWKVYPS
SVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVR
RKFNNGEINF

Figure 2A



Date Reçue/Date Received 2021-09-17

Figure 2B

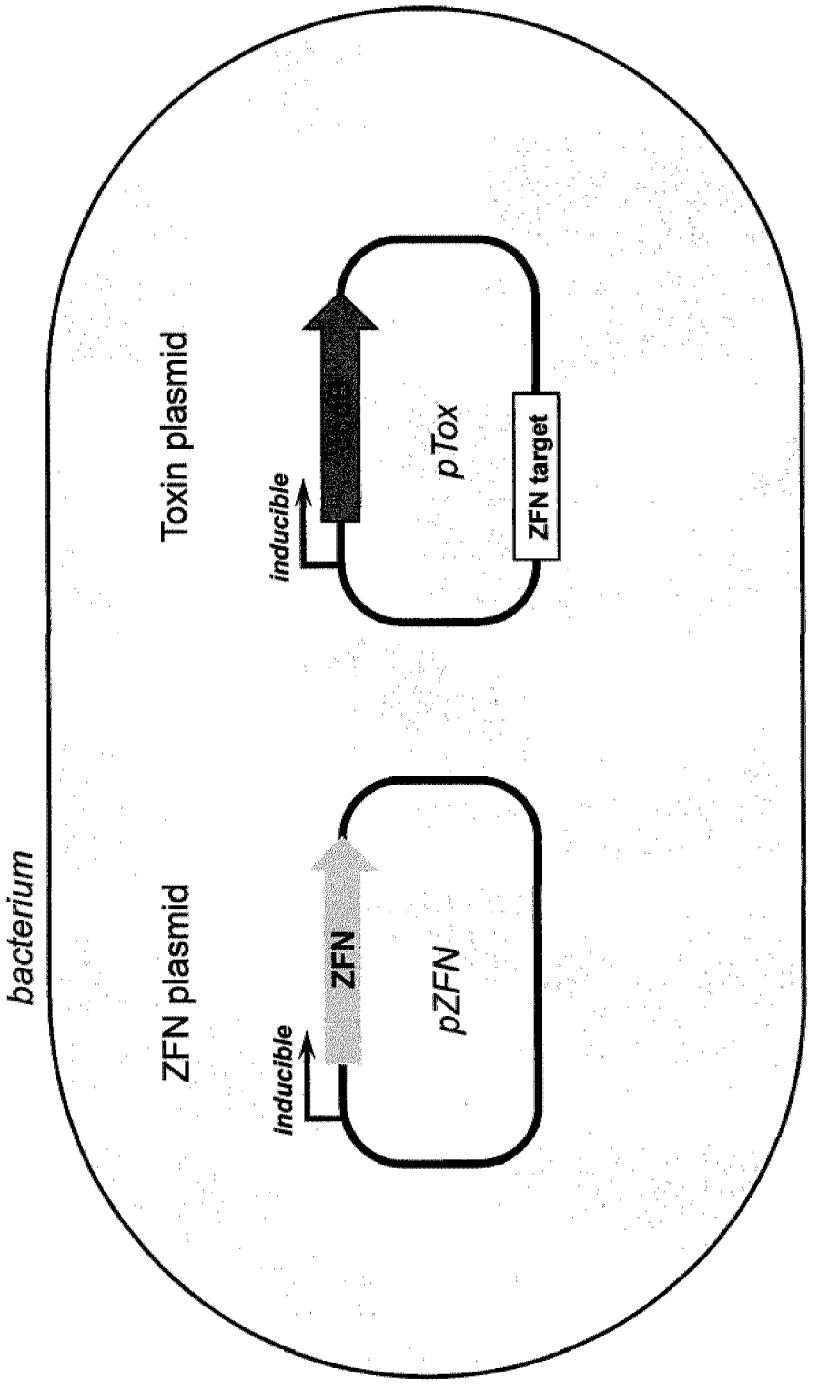

bacterium
ZFN plasmid
inducible
ZFN
pZFN
Toxin plasmid
inducible
pTox
ZFN target

Figure 3

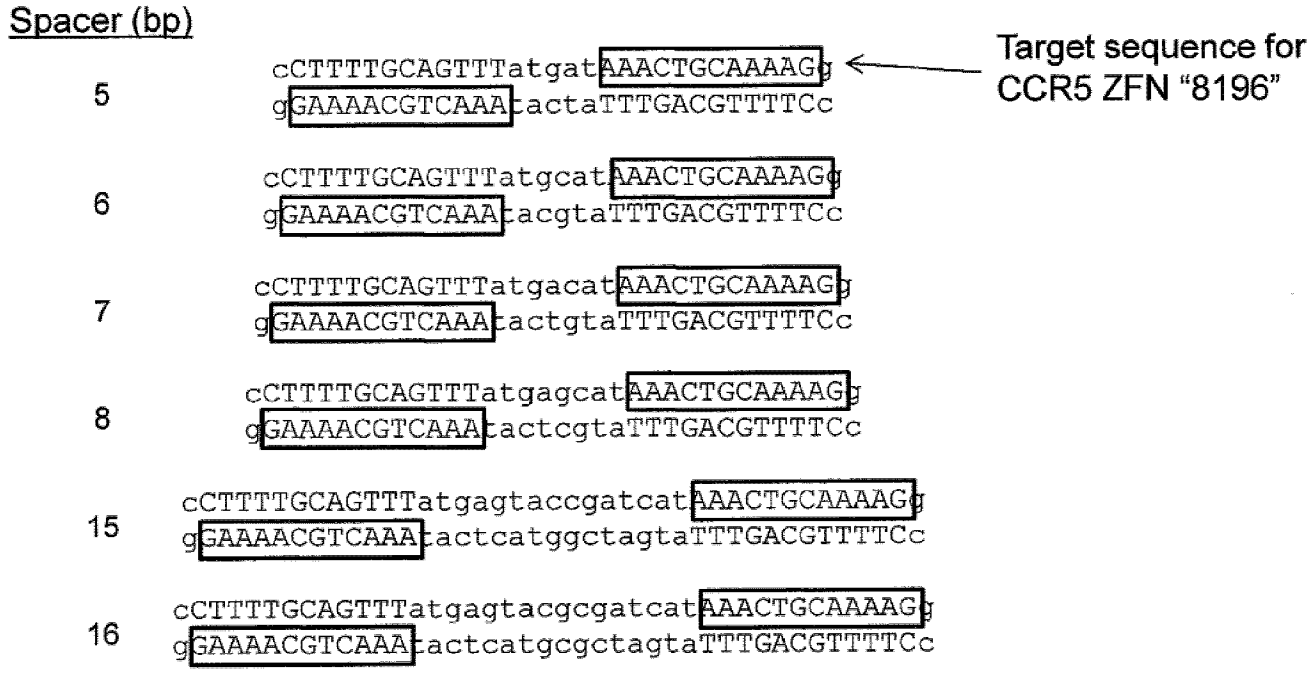


Figure 4

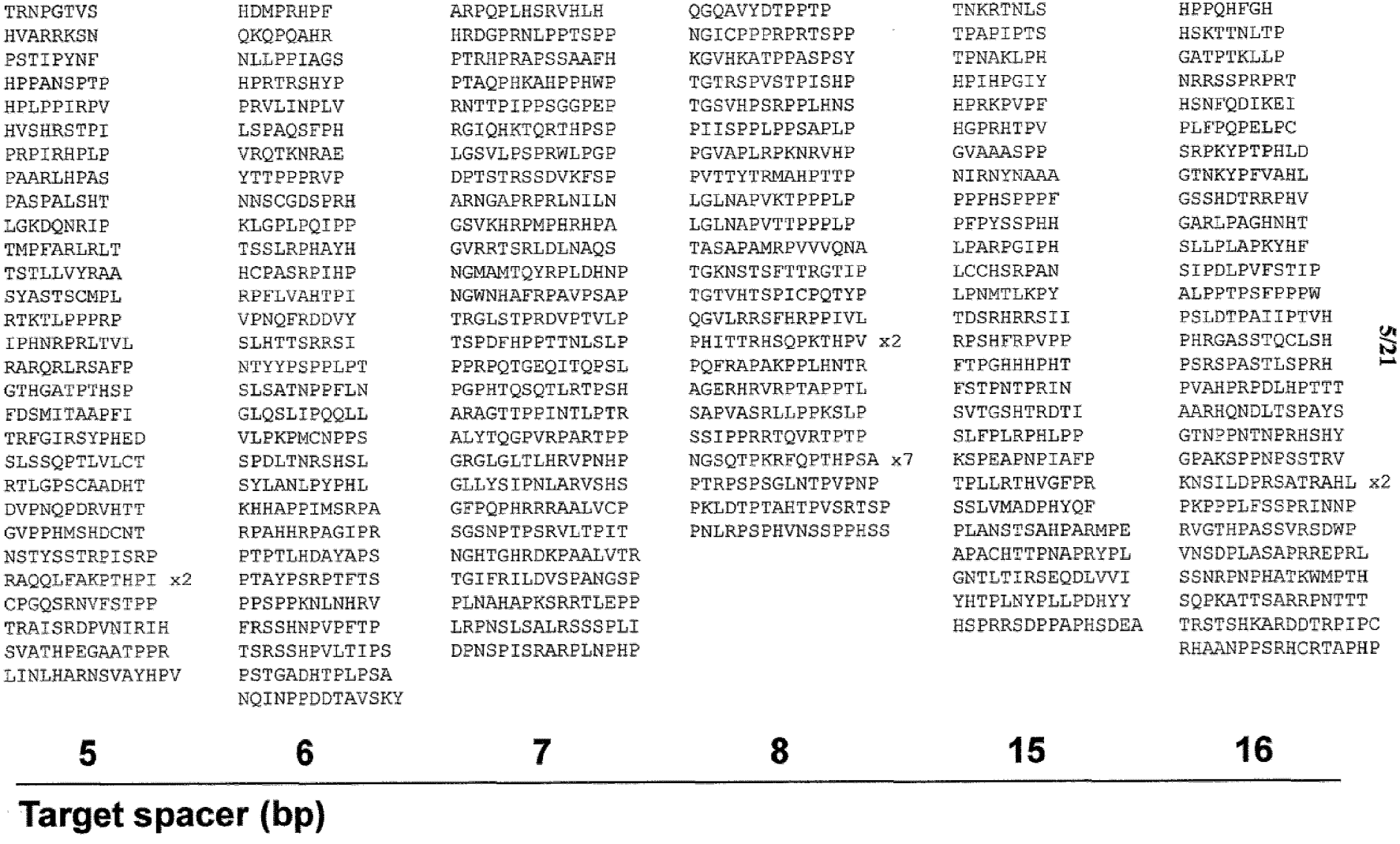


Figure 5

Distribution of selected linker lengths

Target spacer (bp)	length distribution of selected linkers (aa)									
	8	9	10	11	12	13	14	15	16	17
5	3	7	4	4	5	4	2	1		
6	2	6	7	6	6	2	1			
7						1	10	12	5	
8					1		9	10	8*	2
15	7	6	4	2	3	4	1			
16	1	2	3	5	2	3	4	4	3	2

* includes 7 identical sequences

Figure 6

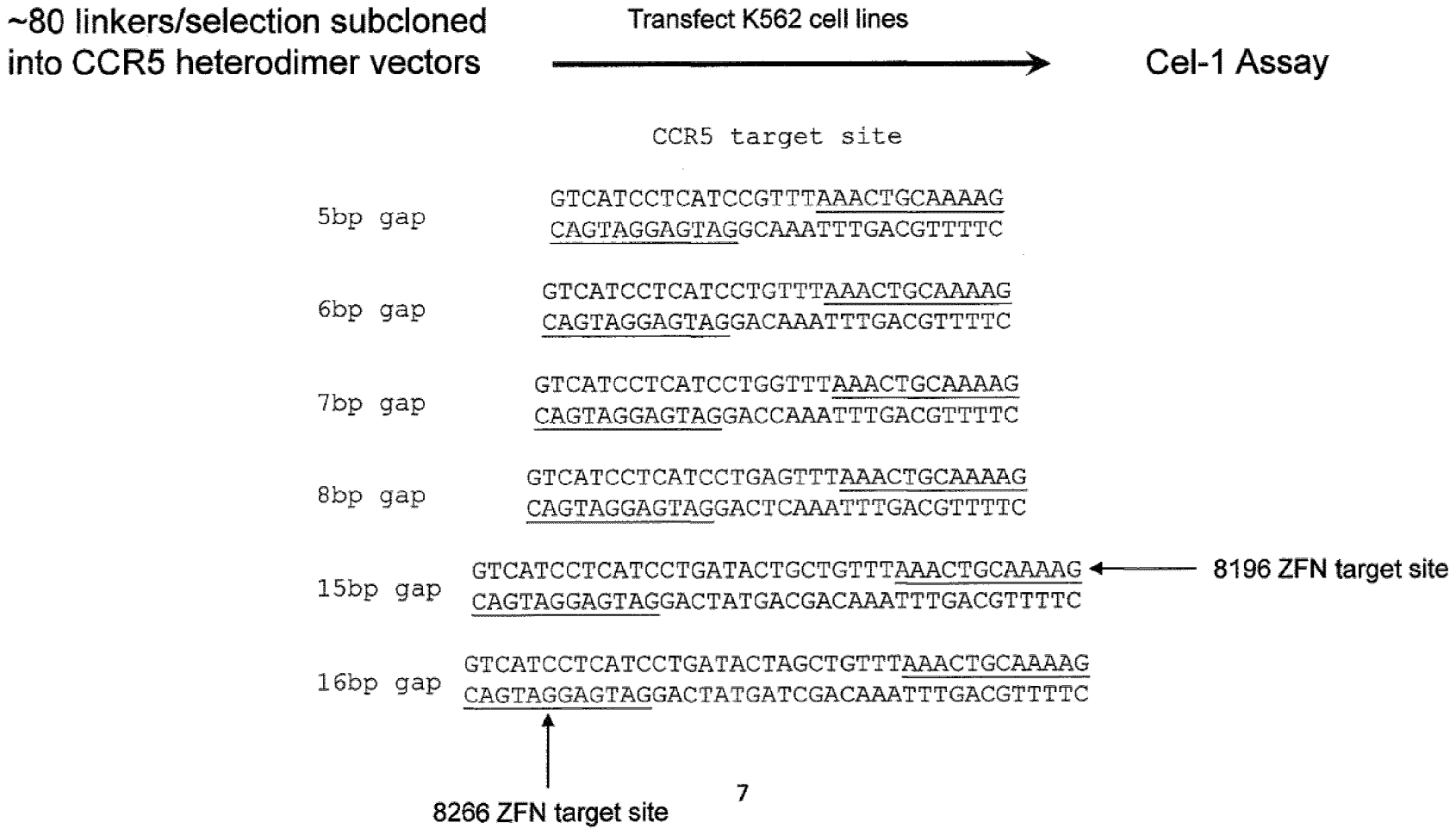
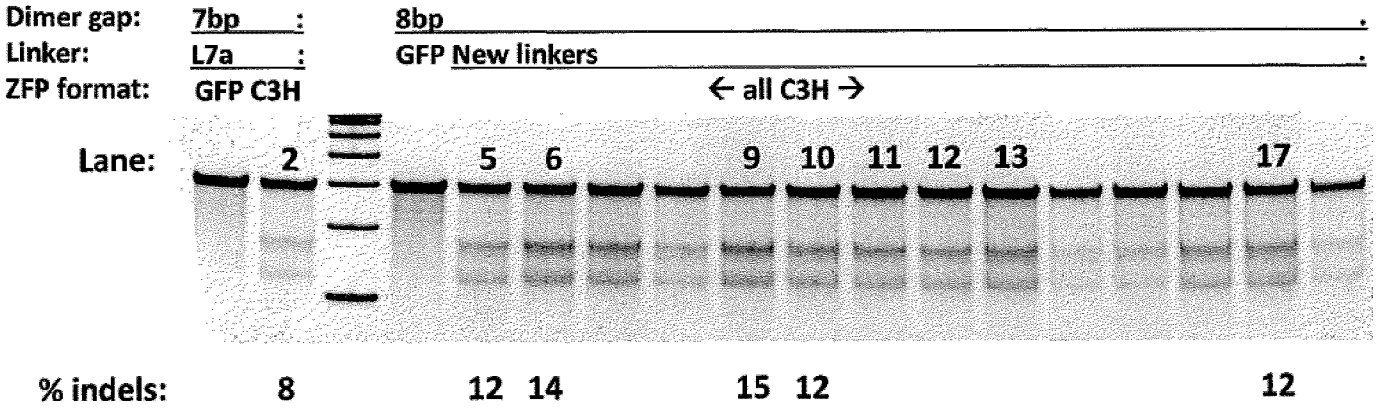


Figure 7: Linkers Selected for 8bp Dimer Gap



Selected linkers:

																% indels	lane
L7a:	G	L	R	G	S	Q	L	V	K	S	K	S	E	A	A	A	R	8	2
L8a	N	G	I	C	P	P	P	R	P	R	T	S	P	P				15	9
L8b	T	G	T	A	P	I	E	I	P	P	E	V	Y	P				14	6
L8c	N	G	S	Y	A	P	M	P	P	L	A	L	A	S	P			12	10
L8d	P	G	I	Y	T	A	P	T	S	R	P	T	V	P	P			12	5
L8e	N	G	S	Q	T	P	K	R	F	Q	P	T	H	P	S	A		12	17
L8g:	T	G	L	M	P	P	S	H	P	R	Q	P	I	H	I	N	F	12	11
L8i	T	G	T	V	H	T	S	P	I	C	P	Q	T	Y	P			11	13
L8j:	T	G	S	G	T	P	T	R	P	H	P	P	L	P	P			11	12

Figure 8

Dimer Gap Target (bp)	Linker performance in cellular studies	
	# screened	# active*
5	80	33
6	80	15
7	80	5
8	57	20
15	75	0
16	78	0

* Defined as activity ≥ that of L7a (~1/3 of L0)

Figure 9A

	8bp dimer gap:																		% indels
L7a	G	L	R	G	S	Q	L	V	K	S	K	S	E	A	A	A	R	(7bp gap)	8
L8a	N	G	I	C	P	P	P	R	P	R	T	S	P	P					15
L8b	T	G	T	A	P	I	E	I	P	P	E	V	Y	P					14
L8c	N	G	S	Y	A	P	M	P	P	L	A	L	A	S	P				12
L8d	P	G	I	Y	T	A	P	T	S	R	P	T	V	P	P				12
L8e	N	G	S	Q	T	P	K	R	F	Q	P	T	H	P	S	A			12
L8g	T	G	L	M	P	P	S	H	P	R	Q	P	I	H	I	N	F		12
L8i	T	G	T	V	H	T	S	P	I	C	P	Q	T	Y	P				11
L8j	T	G	S	G	T	P	T	R	P	H	P	P	L	P	P				11

	7bp dimer gap:																		% indels
L7a	G	L	R	G	S	Q	L	V	K	S	K	S	E	A	A	A	R		12
L7-1	H	L	P	K	P	A	N	P	F	P	L	D							15
L7-2	H	R	D	G	P	R	N	L	P	P	T	S	P	P					12
L7-6	H	R	L	P	D	S	P	T	A	L	A	P	D	T	L				11
L7-3	D	P	N	S	P	I	S	R	A	R	P	L	N	P	H	P			15
L7-5	Y	G	P	R	P	T	P	R	L	R	C	P	I	D	S	L	I	F	R 15
L7-4	G	R	P	R	R	P	T	D	T	P	L	G	S	P	P				10

Figure 9B

	6bp dimer gap:														% indels
L6-2	H	C	P	A	S	R	P	I	H	P					14
L6-6	G	L	Q	S	L	I	P	Q	Q	L	L				16
L6-7	G	L	Q	P	T	V	N	H	E	Y	N	N			15
L6-1	P	A	N	I	H	S	L	S	S	P	P	P	L		14
L6-5	P	A	G	L	N	T	P	C	S	P	R	S	R	S N	13
L6-3	P	T	A	Y	P	S	R	P	T	F	T	S			15
L6-4	H	T	T	P	N	P	H	A	N	T					13

	5bp dimer gap:											% indels
L5-6	A	T	I	T	D	P	N	P				16
L5-1	P	P	H	K	G	L	L	P				27
L5-7	S	V	S	L	P	D	T	H	H			21
L5-8	G	T	H	G	A	T	P	T	H	S	P	15
L5-9	V	A	P	G	E	S	S	M	T	S	L	11
L5-2	I	L	T	P	V	T	H	V	V			12
L5-3	H	R	I	L	S	P	S	R	G			13
L5-4	G	T	P	R	A	S	S	D	S	S	F	10
L5-5	L	G	K	D	Q	N	R	I	P			13

Figure 9C

Linker	Dimer gap target (bp)	% Target modification in indicated cell line			
		5bp	6bp	7bp	8bp
L0	-	16.9	4.8	0.0	0.0
L7a	-	7.8	12.7	4.9	3.0
L5-1	5	18.5	3.3	0.0	0.0
L5-2	5	13.5	2.0	0.0	0.0
L5-3	5	9.7	3.6	0.0	0.0
L5-4	5	10.1	8.2	0.0	0.0
L5-5	5	9.9	0.0	0.0	0.0
L6-1	6	15.5	11.7	0.0	0.0
L6-2	6	12.1	10.1	0.0	0.0
L6-3	6	18.3	15.4	0.0	0.0
L6-4	6	12.5	3.5	0.0	0.0
L6-5	6	10.9	7.2	0.0	0.0

Linker	Dimer gap target (bp)	% Target modification in indicated cell line			
		5bp	6bp	7bp	8bp
L0	-	16.9	4.8	0.0	0.0
L7a	-	7.8	12.7	4.9	3.0
L7-1	7	0.0	10.6	9.6	0.0
L7-2	7	4.3	8.0	4.5	1.9
L7-3	7	2.7	8.2	8.3	6.1
L7-4	7	4.4	8.8	3.3	1.4
L7-5	7	1.3	4.8	3.7	8.8
L8a	8	1.2	12.4	10.2	10.4
L8b	8	0.0	3.9	4.7	10.3
L8d	8	6.6	5.1	6.1	9.0
L8i	8	4.7	6.8	4.3	6.8
L8j	8	8.2	8.0	5.9	8.6

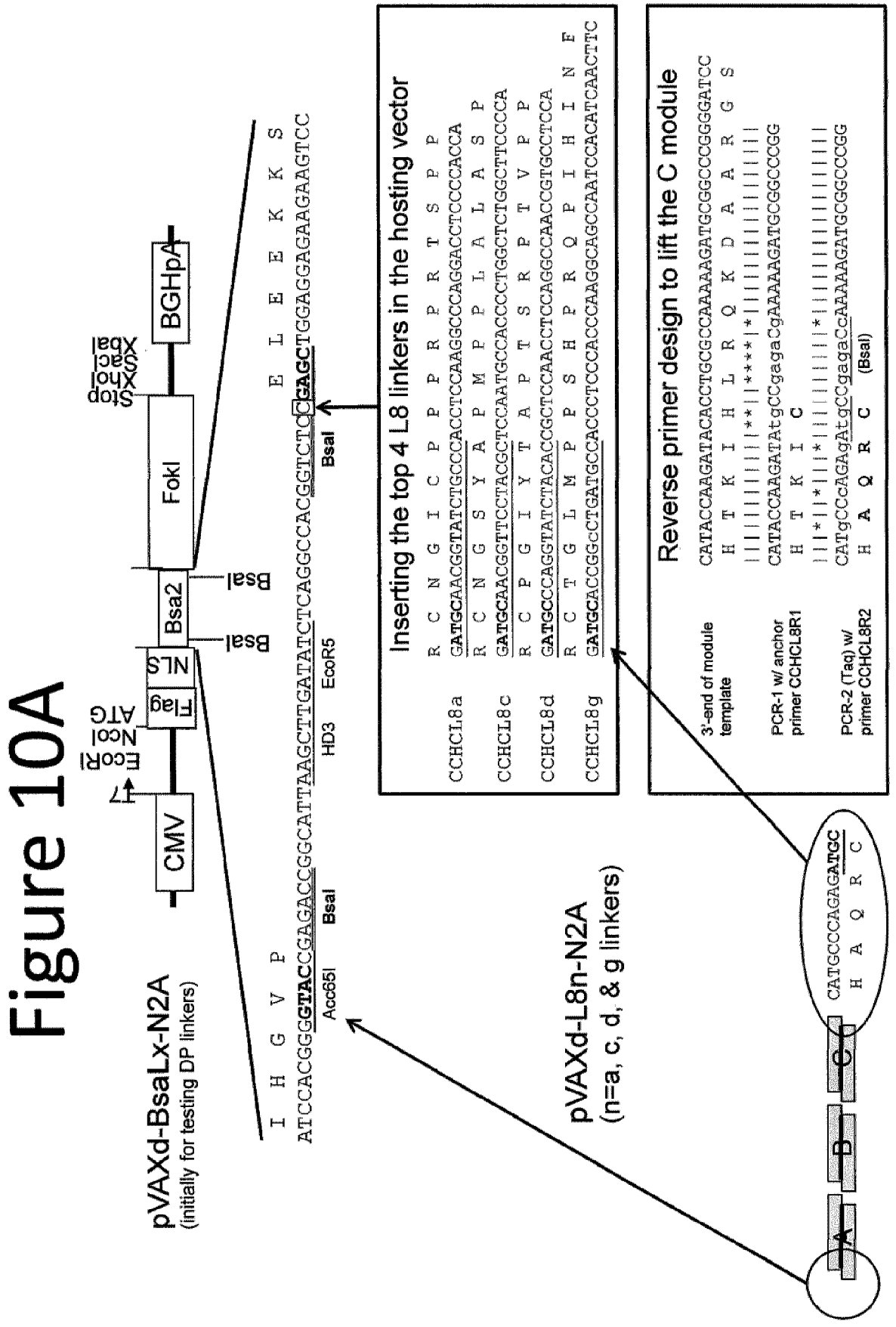
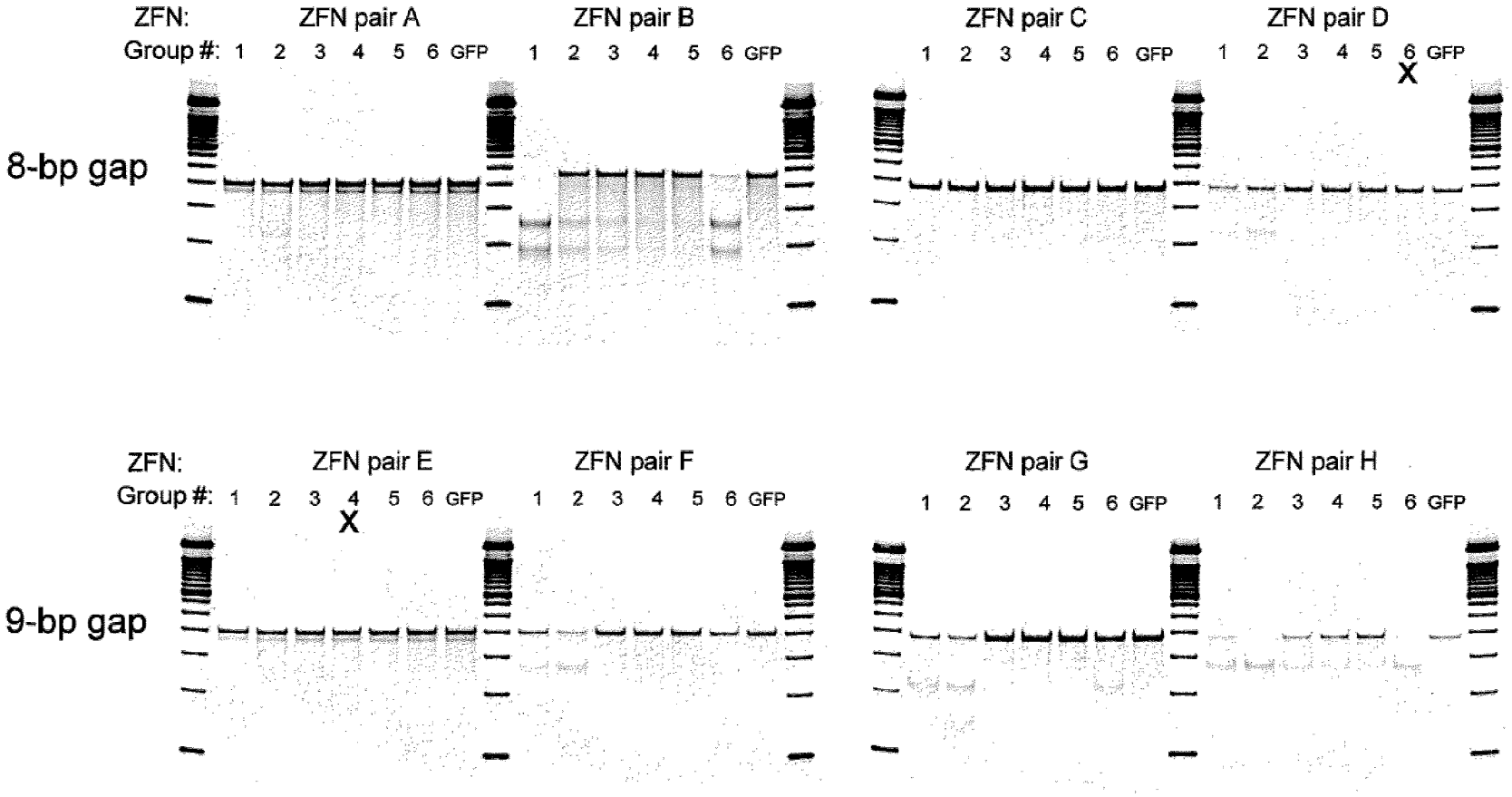


Figure 10A
pVAXd-BsaLx-N2A
(initially for testing DP linkers)
CMV
T7
EcoRI
NcoI
ATG
Flag
NLS
Bsa2
FokI
Stop
XhoI
SacI
XbaI
BGHpA
BsaI
BsaI
I H G V P
ATCCACGGGGTACCGAGACCGGCATTAAGCTTGATATCTCAGGCCACGGTCTCCGAGCTGGAGGAGAAGAAGTCC
E L E E K K S
Acc65I
BsaI
HD3
EcoR5
BsaI
pVAXd-L8n-N2A
(n=a, c, d, & g linkers)
A
B
C
CATGCCCAGAGATGC
H A Q R C
Inserting the top 4 L8 linkers in the hosting vector
R C N G I C P P P P R R P R T S P P
CCHCL8a GATGCAACGGTATCTGCCCACCTCCAAGGCCCAGGACCTCCCCACCA
R C N G S Y A P M P P L A L A S P
CCHCL8c GATGCAACGGTTCCTACGCTCCAATGCCACCCCTGGCTCTGGCTTCCCCA
R C P G G I Y Y T A P T S R R P T V P P
CCHCL8d GATGCCCAGGTATCTACACCGCTCCAACCTCCAGGCCAACCGTGCCTCCA
R C T G G L M P P S H P P R Q P P I H I N F
CCHCL8g GATGCACCGGCCTGATGCCACCCTCCCACCCAAGGCAGCCAATCCACATCCAACTTC
Reverse primer design to lift the C module
3'-end of module template
CATACCAAGATACACCTGCGCCAAAAAGATGCGGCCCGGGGATCC
H T K I H L R Q K D A A R G S
PCR-1 w/ anchor primer CCHCL8R1
CATACCAAGATATgCCgagaCgAAAAAGATGCGGCCCGG
H T K I C
PCR-2 (Taq) w/ primer CCHCL8R2
CATgCCcAGAgAtgCCgagaCcAAAAAGATGCGGCCCGG
H A Q R C (BsaI)

Figure 10B

	Linker					
	L0	L7a	L8a	L8c	L8d	L8g
# ZFNs Tested	16	15	13	12	13	13
# > 1% (fraction)	9 (0.56)	11 (0.73)	9 (0.69)	10 (0.83)	10 (0.77)	9 (0.69)
# > 10% (fraction)	5 (0.31)	7 (0.47)	5 (0.38)	5 (0.42)	6 (0.46)	6 (0.46)
Average Activity	6.5	10.2	9.5	12.1	8.8	11.1

Figure 11A



Group 1: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[wt FokI]
Group 2: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 3: L8n [ZFP]HTKIH-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 4: L8o [ZFP]HTKIH-GGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 5: L8m [ZFP]HTKIH--GSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 6: L8p [ZFP]HTKIHLRGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]

Figure 11B

8 bp gap

ZFN Pair A	
group	%NHEJ
1	5.2
2	6.2
3	0.0
4	0.0
5	0.0
6	0.0
GFP	0.0

ZFN Pair B	
group	%NHEJ
1	78.6
2	24.0
3	15.6
4	8.6
5	7.5
6	59.5
GFP	0.0

ZFN Pair C	
group	%NHEJ
1	2.5
2	5.6
3	2.0
4	1.9
5	2.1
6	3.6
GFP	0.0

ZFN Pair D	
group	%NHEJ
1	13.1
2	20.8
3	0.0
4	0.0
5	0.0
6	nd
GFP	0.0

9 bp gap

ZFN Pair E	
group	%NHEJ
1	6.2
2	4.9
3	0.0
4	nd
5	0.0
6	0.0
GFP	0.0

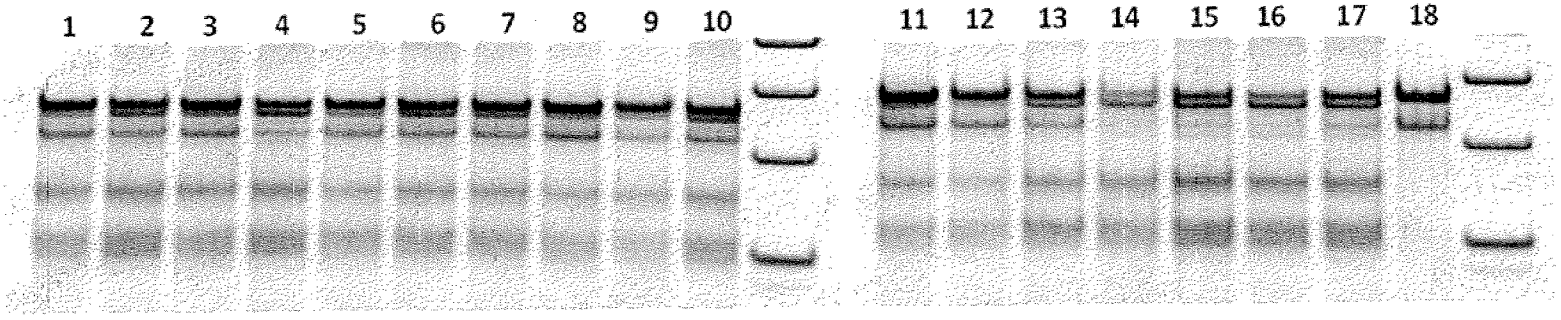
ZFN Pair F	
group	%NHEJ
1	16.6
2	33.6
3	0.0
4	0.0
5	0.0
6	11.1
GFP	0.0

ZFN Pair G	
group	%NHEJ
1	26.9
2	37.6
3	2.2
4	0.0
5	0.0
6	12.4
GFP	0.0

ZFN Pair H	
group	%NHEJ
1	39.8
2	72.6
3	24.8
4	15.0
5	7.7
6	63.3
GFP	0.0

Group 1: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[wt FokI]
Group 2: L8c [ZFP]HAQRC-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 3: L8n [ZFP]HTKIH-NGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 4: L8o [ZFP]HTKIH-GGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 5: L8m [ZFP]HTKIH--GSYAPMPPLALASPELEEKKSEL[eHiFi FokI]
Group 6: L8p [ZFP]HTKIHLRGSYAPMPPLALASPELEEKKSEL[eHiFi FokI]

Figure 12



	Sample #	Vector	Linker	Cel1 %NHEJ	MiSeq %NHEJ	
Group 1	1	CCHC-L7b-wt	HAQRC HLPKPANPFPLD ELEEK	13.9	26.8	
	2	CCHC-L7c-wt	HAQRC DPNSPISRARPLNPHP ELEEK	18.8	41.2	
	3	CCHC-L8a-wt	HAQRC NGICPPPRPRTSPP ELEEK	13.7	31.2	
	4	CCHC-L8c-wt	HAQRC NGSYAPMPPLALASP ELEEK	21.0	51.5	
Group 2	5	CCHC-L7b-eHiFi	HAQRC HLPKPANPFPLD ELEEK	15.3	33.5	
	6	CCHC-L7c-eHiFi	HAQRC DPNSPISRARPLNPHP ELEEK	15.6	38.5	
	7	CCHC-L8a-eHiFi	HAQRC NGICPPPRPRTSPP ELEEK	15.9	36.9	
	8	CCHC-L8c-eHiFi	HAQRC NGSYAPMPPLALASP ELEEK	10.1	22.7	
Group 3	9	C2H2-L7b2-eHiFi	HTKIH HLPKPANPFPLD ELEEK	14.6	31.7	
	10	C2H2-L7c2-eHiFi	HTKIH DPNSPISRARPLNPHP ELEEK	16.9	38.2	
	11	C2H2-L8a2-eHiFi	HTKIH NGICPPPRPRTSPP ELEEK	10.6	21.3	
	12	C2H2-L8c2-eHiFi	HTKIH NGSYAPMPPLALASP ELEEK	12.4	23.9	
Group 4	13	C2H2-L7b3-eHiFi	HTKIH LPKPANPFPLD ELEEK	20.9	43.7	
	14	C2H2-L7c3-eHiFi	HTKIH PNSPISRARPLNPHP ELEEK	31.4	66.0	←
	15	C2H2-L8a3-eHiFi	HTKIH GICPPPRPRTSPP ELEEK	28.2	63.2	
	16	C2H2-L8c3-eHiFi	HTKIH GSYAPMPPLALASP ELEEK	27.8	69.1	←
	17	L7a-eHiFi	HTKIH LRGSQLVKSKSEAAAR ELEEK	25.0	51.7	
	18	eGFP control		0.0	0.5	

Figure 13

Name	Linker Sequence	NHEJ% ZFN pair #1	NHEJ% ZFN pair #1	
L7a	HTKIH LRGSQLVKSKSEAAAR ELEEK	12.6%	11.9%	
L7c4	HTKIH LRGSPISRARPLNPHP ELEEK	8.1%	18.7%	
L7c5	HTKIH LRGSISRARPLNPHP ELEEK	19.8%	18.0%	←
L7c6	HTKIH LRGSPSRARPLNPHP ELEEK	17.3%	14.5%	
L7c7	HTKIH LRGSSRARPLNPHP ELEEK	12.9%	13.4%	
L7e4	HTKIH LRGSYAPMPPLALASP ELEEK	19.6%	18.2%	←
L7e5	HTKIH LRGSAPMPPLALASP ELEEK	14.4%	15.0%	
L7e6	HTKIH LRGSPMPPLALASP ELEEK	15.8%	13.2%	
L7e7	HTKIH LRGSMPPLALASP ELEEK	5.3%	8.8%	

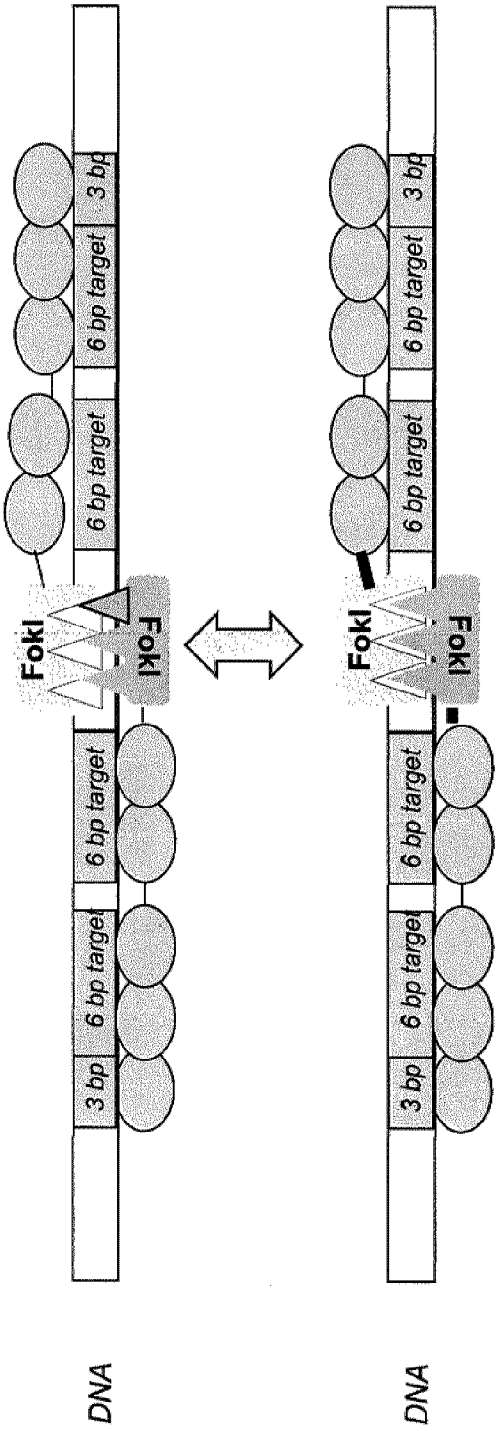

FokI
FokI
3 bp
6 bp target
6 bp target
6 bp target
3 bp
DNA
FokI
FokI
3 bp
6 bp target
6 bp target
6 bp target
3 bp
DNA

Figure 14

Figure 15

Variants	aa	Junction	Left ZFN Variants	Right ZFN variants	Left and right ZFN variants
V1	-3	HTKIHLRGS------KSELEEK	47.1	56.3	68.7
V2	-2	HTKIHLRGS-----GKSELEEK	61.3	58.5	40.6
V3	-2	HTKIHLRGS-----VKSELEEK	50.1	59.0	41.2
V4	-2	HTKIHLRGS-----EKSELEEK	38.2	39.5	52.3
V5	-1	HTKIHLRGS----GVKSELEEK	23.6	39.0	32.6
V6	-1	HTKIHLRGS----SVKSELEEK	32.2	50.8	45.8
V7	-1	HTKIHLRGS----LVKSELEEK	35.5	63.5	50.1
V8	-1	HTKIHLRGS----EVKSELEEK	24.3	64.3	45.3
V9	0	HTKIHLRGS---GLVKSELEEK	44.6	28.8	30.1
V10	0	HTKIHLRGS---SLVKSELEEK	44.6	56.8	16.6
V11	0	HTKIHLRGS---RLVKSELEEK	24.7	47.0	41.1
V12	0	HTKIHLRGS---WLVKSELEEK	16.2	22.3	36.5
V13	0	HTKIHLRGS---QGVKSELEEK	41.3	62.5	42.1
V14	0	HTKIHLRGS---QSVKSELEEK	27.1	56.2	31.8
V15	0	HTKIHLRGS---QLSKSELEEK	40.6	47.5	57.1
V16	+1	HTKIHLRGS--GQLVKSELEEK	20.3	42.4	51.0
V17	+1	HTKIHLRGS--SQLVKSELEEK	62.7	36.7	33.5
V18	+1	HTKIHLRGS--KQLVKSELEEK	50.1	57.7	22.5
V19	+1	HTKIHLRGS--RQLVKSELEEK	27.0	57.9	32.2
V20	+2	HTKIHLRGS-GSQLVKSELEEK	26.8	38.8	41.6
V21	+2	HTKIHLRGS-GKQLVKSELEEK	50.5	49.4	30.6
V22	+2	HTKIHLRGS-TKQLVKSELEEK	46.1	49.8	41.1
V23	+3	HTKIHLRGSGTKQLVKSELEEK	53.8	41.4	16.2
V24	+3	HTKIHLRGSVTKQLVKSELEEK	66.0	52.2	41.1
Conventional	0	HTKIHLRGS---QLVKSELEEK	45.4		
EGFP			0.1		

Figure 16

Pair 1 variants

LEFT\\RIGHT	GS------KS	GS-----VKS	GS----EVKS	GS---QSVKS	GS--KQLVKS	GS-TKQLVKS	GSVTKQLVKS	GS---QLVKS	Average
GS------KS	0.7	3.8	4.6	4.6	3.3	5.7	3.7	7.1	4.2
GS-----VKS	7.1	19.5	21.7	17.6	16.6	21.3	18.9	21.5	18.0
GS----EVKS	17.3	26.5	24.8	23.0	22.5	27.1	24.4	22.3	23.5
GS---QSVKS	16.6	23.9	24.1	22.6	24.2	25.3	25.7	23.5	23.3
GS--KQLVKS	16.0	23.3	29.2	25.0	28.6	27.1	32.3	27.3	26.1
GS-GKQLVKS	15.1	22.4	28.8	26.0	27.2	29.3	31.7	25.8	25.8
GSVTKQLVKS	18.1	26.7	32.4	30.2	31.7	33.4	**36.0**	27.0	29.4
GS---QLVKS	18.1	28.0	31.5	29.0	28.6	25.0	30.0	17.2	25.9
Average	13.6	21.8	24.7	22.2	22.8	24.3	25.3	21.5	

Pair 2 variants

LEFT\\RIGHT	GS------KS	GS-----VKS	GS----EVKS	GS---QSVKS	GS--KQLVKS	GS-TKQLVKS	GSVTKQLVKS	GS---QLVKS	Average
GS------KS	0.4	1.9	1.7	1.9	0.9	1.7	1.6	1.5	1.4
GS-----VKS	3.5	10.9	7.7	8.9	4.1	4.9	6.1	7.1	6.6
GS----EVKS	9.6	16.6	13.8	13.8	12.3	12.1	11.7	15.0	13.1
GS---QSVKS	10.1	16.4	14.2	13.9	13.7	14.0	12.4	14.7	13.7
GS--KQLVKS	13.7	19.2	15.9	16.6	13.9	15.3	15.2	16.4	15.8
GS-GKQLVKS	12.1	18.4	15.5	18.6	16.0	19.7	16.1	15.9	16.5
GSVTKQLVKS	15.1	20.6	15.5	19.3	18.7	19.9	18.0	19.3	18.3
GS---QLVKS	8.3	7.9	5.5	8.0	8.7	6.7	7.1	11.1	7.9
Average	9.1	14.0	11.2	12.6	11.1	11.8	11.0	12.6	