Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2615532 C 2016/06/28

(11)(21) **2 615 532**

(12) **BREVET CANADIEN
CANADIAN PATENT**

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date:** 2006/07/26

(87) **Date publication PCT/PCT Publication Date:** 2007/02/01

(45) **Date de délivrance/Issue Date:** 2016/06/28

(85) **Entrée phase nationale/National Entry:** 2008/01/16

(86) **N° demande PCT/PCT Application No.:** US 2006/029027

(87) **N° publication PCT/PCT Publication No.:** 2007/014275

(30) **Priorités/Priorities:** 2005/07/26 (US60/702,394); 2005/09/26 (US60/721,054)

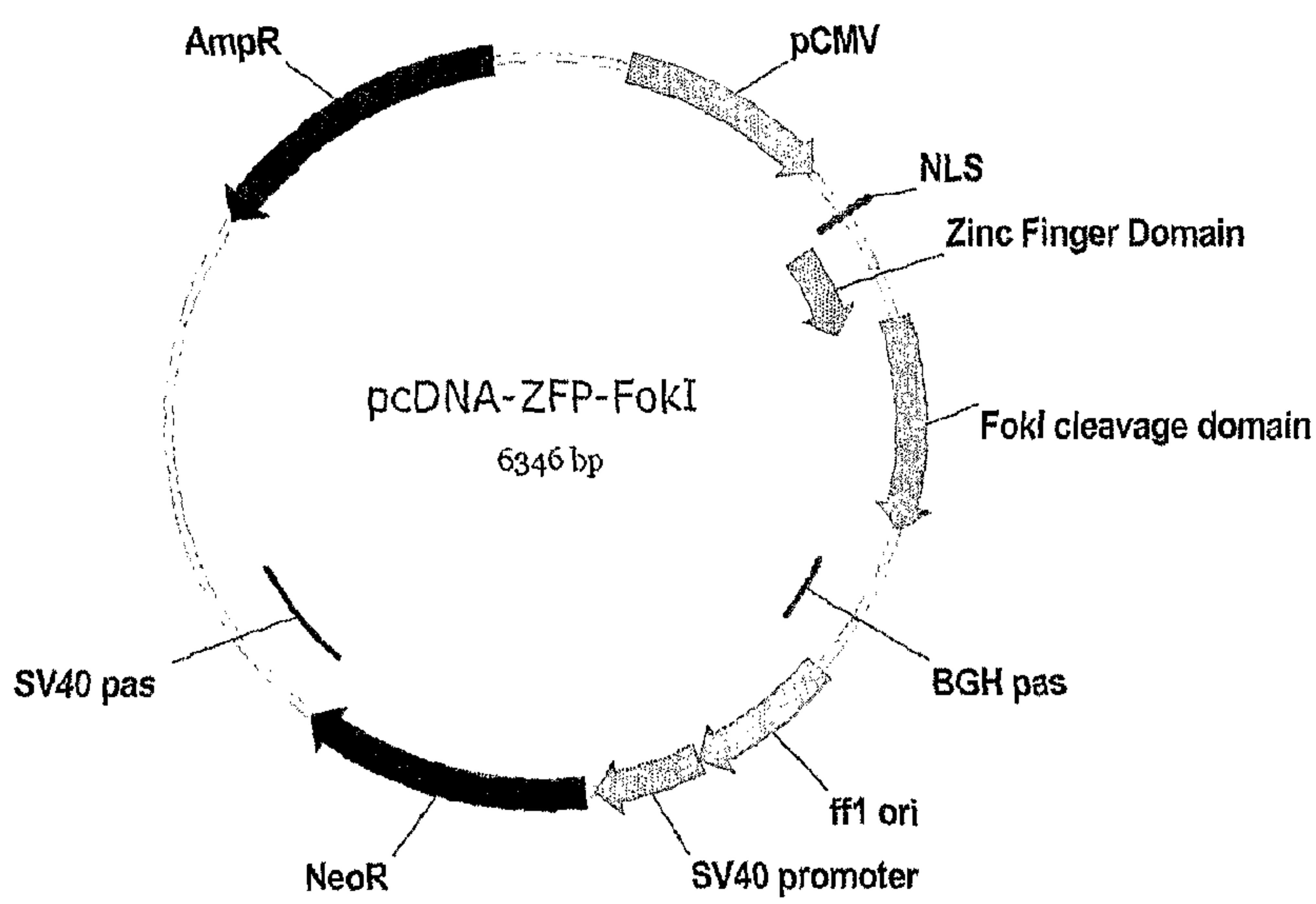
(51) **Cl.Int./Int.Cl.** ***C12P 21/06*** (2006.01), ***C07H 21/04*** (2006.01), ***C12N 15/00*** (2006.01)

(72) **Inventeurs/Inventors:**
HOLMES, MICHAEL C., US;
URNOV, FYODOR, US;
GREGORY, PHILIP D., US;
REBAR, EDWARD J., US;
BRENNAN, SEAN M., US

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

(74) **Agent:** KIRBY EADES GALE BAKER

(54) **Titre : INTEGRATION ET EXPRESSION CIBLEES DE SEQUENCES D'ACIDES NUCLEIQUES EXOGENES**
(54) **Title: TARGETED INTEGRATION AND EXPRESSION OF EXOGENOUS NUCLEIC ACID SEQUENCES**



(57) **Abrégé/Abstract:**
Disclosed herein are methods and compositions for targeted integration of a exogenous sequence into a predetermined target site in a genome for use, for example, in protein expression and gene inactivation.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 February 2007 (01.02.2007)

PCT

(10) International Publication Number
WO 2007/014275 A3

(51) International Patent Classification:
C12P 21/06 (2006.01) *C07H 21/04* (2006.01)
C12N 15/00 (2006.01)

(21) International Application Number:
PCT/US2006/029027

(22) International Filing Date: 26 July 2006 (26.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/702,394 26 July 2005 (26.07.2005) US
60/721,054 26 September 2005 (26.09.2005) US

(71) Applicant *(for all designated States except US)*: **SANGAMO BIOSCIENCES, INC.** [US/US]; Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, California 94804 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HOLMES, Michael C.** [US/US]; 3934 Glen Park Rd., Oakland, California 94602 (US). **URNOV, Fyodor** [RU/US]; 1400 Pinnacle Court #313, Point Richmond, California 94801 (US). **GREGORY, Philip D.** [GB/US]; 155 Alice Lane, Orinda, California 94563 (US). **REBAR, Edward J.** [US/US]; 1609 Roger Ct., El Cerrito, California 94530 (US). **BRENNAN, Sean M.** [US/US]; 3020 Cabrillo Avenue, San Ramon, California 94583 (US).

(74) Agent: **PASTERNAK, Dahna S.**; ROBINS & PASTERNAK LLP, 1731 Embarcadero Road, Suite 230, Palo Alto, California 94303 (US).

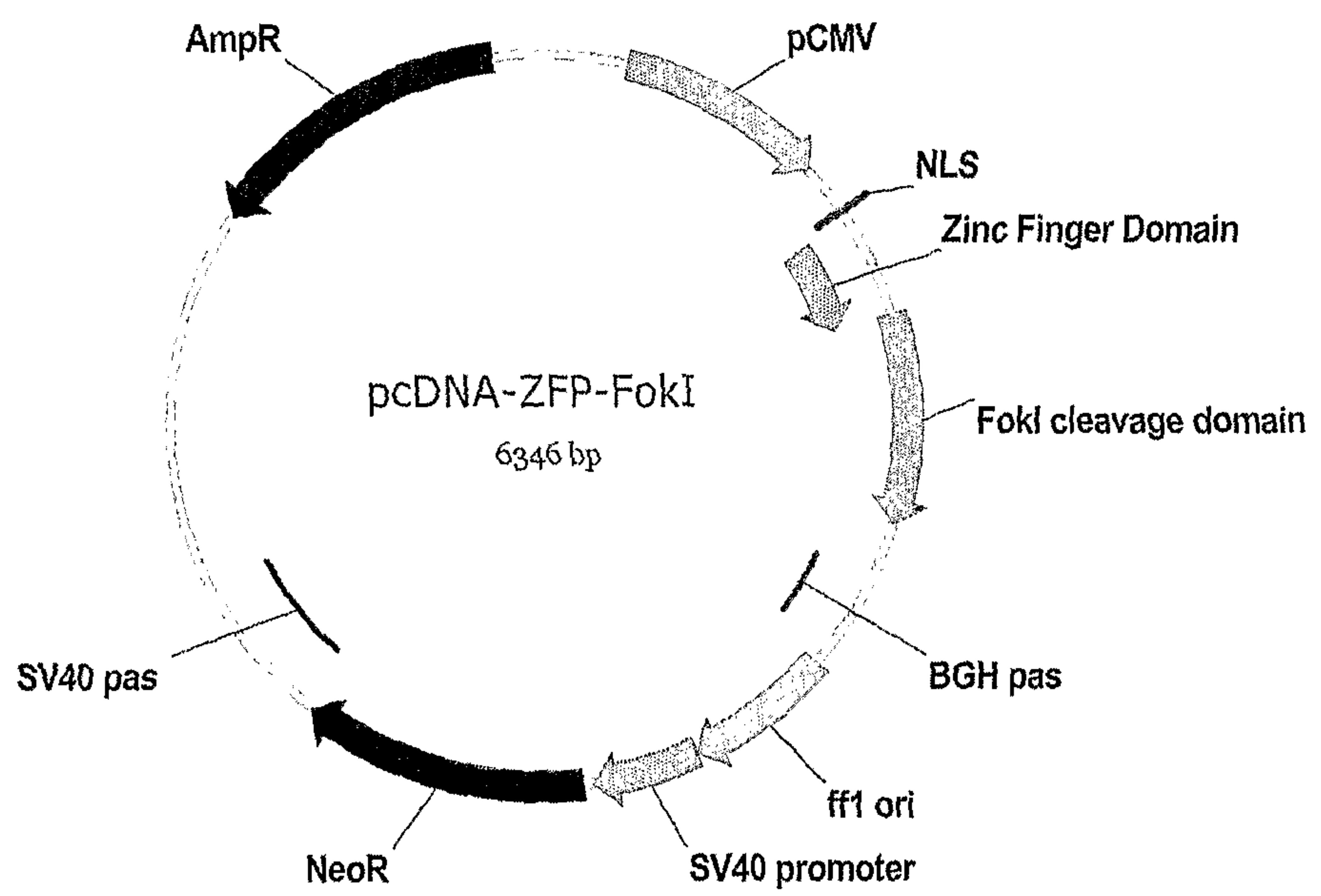
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ,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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) **Title:** TARGETED INTEGRATION AND EXPRESSION OF EXOGENOUS NUCLEIC ACID SEQUENCES



(57) **Abstract:** Disclosed herein are methods and compositions for targeted integration of a exogenous sequence into a predetermined target site in a genome for use, for example, in protein expression and gene inactivation.

WO 2007/014275 A3

WO 2007/014275 A3

(88) Date of publication of the international search report:
24 May 2007

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

TARGETED INTEGRATION AND EXPRESSION OF EXOGENOUS NUCLEIC ACID SEQUENCES

TECHNICAL FIELD

[0002] The present disclosure is in the fields of genome engineering, gene targeting, targeted chromosomal integration and protein expression.

BACKGROUND

[0003] A major area of interest in genome biology, especially in light of the determination of the complete nucleotide sequences of a number of genomes, is the targeted alteration of genome sequences. To provide but one example, sickle cell anemia is caused by mutation of a single nucleotide pair in the human β-globin gene. Thus, the ability to convert the endogenous genomic copy of this mutant nucleotide pair to the wild-type sequence in a stable fashion and produce normal β-globin would provide a cure for sickle cell anemia, as would introduction of a functional β-globin gene into a genome containing a mutant β-globin gene.

[0004] Attempts have been made to alter genomic sequences in cultured cells by taking advantage of the natural phenomenon of homologous recombination. *See,* for example, Capecchi (1989) *Science* **244:**1288-1292; U.S. Patent Nos. 6,528,313 and 6,528,314. If a polynucleotide has sufficient homology to the genomic region containing the sequence to be altered, it is possible for part or all of the sequence of the polynucleotide to replace the genomic sequence by homologous recombination. However, the frequency of homologous recombination under these circumstances is extremely low. Moreover, the frequency of insertion of the exogenous polynucleotide at genomic locations that lack sequence homology exceeds the frequency of homologous recombination by several orders of magnitude.

[0005] The introduction of a double-stranded break into genomic DNA, in the region of the genome bearing homology to an exogenous polynucleotide, has been shown to stimulate homologous recombination at this site by several thousand-fold in cultured cells. Rouet *et al.* (1994) *Mol. Cell. Biol.* **14:**8096-8106; Choulika *et al.* (1995) *Mol. Cell. Biol.* **15:**1968-1973; Donoho *et al.* (1998) *Mol. Cell. Biol.* **18:**4070-4078. *See* also Johnson *et al.* (2001) *Biochem. Soc. Trans.* **29:**196-201; and Yanez *et al.* (1998) *Gene Therapy* **5:**149-159. In these methods, DNA cleavage in the desired genomic region was accomplished by inserting a recognition site for a meganuclease (*i.e.*, an endonuclease whose recognition sequence is so large that it does not occur, or occurs only rarely, in the genome of interest) into the desired genomic region.

[0006] However, meganuclease cleavage-stimulated homologous recombination relies on either the fortuitous presence of, or the directed insertion of, a suitable meganuclease recognition site in the vicinity of the genomic region to be altered. Since meganuclease recognition sites are rare (or nonexistent) in a typical mammalian genome, and insertion of a suitable meganuclease recognition site is plagued with the same difficulties as associated with other genomic alterations, these methods are not broadly applicable.

[0007] Thus, there remain needs for compositions and methods for targeted alteration of sequences in any genome and for compositions and methods for targeted introduction of exogenous sequences into a genome.

SUMMARY

[0008] The present disclosure provides method and compositions for expressing the product of an exogenous nucleic acid sequence (*i.e.* a protein or a RNA molecule) in a cell. The exogenous nucleic acid sequence can comprise, for example, one or more genes or cDNA molecules, or any type of coding or noncoding sequence, and is introduced into the cell such that it is integrated into the genome of the cell in a predetermined region of interest. Integration of the exogenous nucleic acid sequence is facilitated by targeted double-strand cleavage of the genome in the region of interest. Cleavage is targeted to a particular site through the use of fusion proteins comprising a zinc finger binding domain, which can be engineered to bind any sequence of choice in the region of interest, and a cleavage domain or a cleavage half-domain. Such cleavage stimulates integration of exogenous polynucleotide sequences

at or near the cleavage site. Said integration of exogenous sequences can proceed through both homology-dependent and homology-independent mechanisms.

[0009] Also provided are methods and compositions for modulating the expression of an endogenous cellular gene by targeted integration (either homology-dependent or homology-independent) of one or more exogenous sequences. Such exogenous sequences can include, for example, transcriptional control sequences such as promoters and enhancers. Modulation can include transcriptional activation (*e.g.*, enhancement of transcription by, for example, insertion of a promoter and/or enhancer sequence) and transcriptional repression (*e.g.*, functional "knock-out" by, for example, inserting an exogenous sequence into an endogenous transcriptional regulatory sequence, inserting a sequence facilitating transcriptional repression, or inserting a sequence that interrupts a coding region).

[0010] Also provided are methods and compositions for targeted insertion of an exogenous sequence into a genome, by either homology-dependent or homology-independent mechanisms, wherein the exogenous sequence does not express a product or modulate expression of an endogenous gene. For example, a recognition sequence for a sequence-specific DNA-cleaving enzyme can be introduced at a predetermined location in a genome so that targeted cleavage by the cleaving enzyme, at the predetermined location in the genome, can be accomplished. Exemplary DNA-cleaving enzymes include, but are not limited to, restriction enzymes, meganucleases and homing endonucleases.

[0010a] Certain exemplary embodiments provide a protein comprising an engineered zinc finger protein DNA-binding domain, wherein the DNA-binding domain comprises four zinc finger recognition regions ordered F1 to F4 from N-terminus to C-terminus, and (i) wherein F1, F2, F3, and F4 comprise the following amino acid sequences:

F1: QSGALAR (SEQ ID NO:229)
F2: RSDNLRE (SEQ ID NO:230)
F3: QSSDLSR (SEQ ID NO:231)
F4: TSSNRKT (SEQ ID NO:232) or

(ii) wherein F1, F2, F3, and F4 comprise the following amino acid sequences:

F1: RSDTLSE (SEQ ID NO:224)
F2: NNRDRTK (SEQ ID NO:225)

F3: RSDHLSA (SEQ ID NO:226)

F4: QSGHLSR (SEQ ID NO:227).

[0011] In one aspect, disclosed herein is a method for expression the product of an exogenous nucleic acid sequence in a cell, the method comprising: (a) expressing a first fusion protein in the cell, the first fusion protein comprising a first zinc finger binding domain and a first cleavage half-domain, wherein the first zinc finger binding domain has been engineered to bind to a first target site in a region of interest in the genome of the cell; (b) expressing a second fusion protein in the cell, the second fusion protein comprising a second zinc finger binding domain and a second cleavage half domain, wherein the second zinc finger binding domain binds to a second target site in the region of interest in the genome of the cell, wherein the second target site is different from the first target site; and (c) contacting the cell with a polynucleotide comprising an exogenous nucleic acid sequence and a first nucleotide sequence that is homologous to a first sequence in the region of interest; wherein binding of the first fusion protein to the first target site, and binding of the second fusion protein to the second target site, positions the cleavage half-domains such that the genome of the cell is cleaved in the region of interest, thereby resulting in integration of the exogenous sequence into the genome of the cell in the region of interest and expression of the product of the exogenous sequence.

[0012] The exogenous nucleic acid sequence may comprise a cDNA and/or a promoter. In other embodiments, the exogenous nucleic acid sequence encodes a siRNA. The first nucleotide sequence may be identical to the first sequence in the region of interest.

[0013] In certain embodiments, the polynucleotide further comprises a second nucleotide sequence that is homologous to a second sequence in the region of interest. The second nucleotide sequence may be identical to the second sequence in the region of interest. Furthermore, in embodiments comprising first and second nucleotide sequences, the first nucleotide sequence may identical to the first sequence in the region of interest and the second nucleotide sequence may be homologous but non-identical to a second sequence in the region of interest. In any of the methods described herein, the first and second nucleotide sequences flank the exogenous sequence.

[0014] In certain embodiments, the polynucleotide is a plasmid. In other embodiments, the polynucleotide is a linear DNA molecule.

[0015] In any of the methods described herein, the region of interest is in an accessible region of cellular chromatin, a chromosome, and/or a gene (*e.g.*, a gene comprising a mutation such as a point mutation, a substitution, a deletion, an insertion, a duplication, an inversion and/or a translocation). In certain embodiments, the exogenous nucleic acid sequence comprises the wild-type sequence of the gene. In other embodiments, the exogenous nucleic acid sequence comprises a portion of the wild-type sequence of the gene. In still other embodiments, the exogenous nucleic acid sequence comprises a cDNA copy of a transcription product of the gene.

[0016] In any of the methods described herein, the region of interest is in a region of the genome that is not essential for viability. In other embodiments, the region of interest is in a region of the genome that is transcriptionally active. The region of interest is in a region of the genome that is transcriptionally active and not essential for viability (*e.g.*, the human Rosa26 genome, the human homologue of the murine Rosa26 gene, or a CCR5 gene).

[0017] In another aspect, provided herein is a method for integrating an exogenous sequence into a region of interest in the genome of a cell, the method comprising: (a) expressing a first fusion protein in the cell, the first fusion protein comprising a first zinc finger binding domain and a first cleavage half-domain, wherein the first zinc finger binding domain has been engineered to bind to a first target site in a region of interest in the genome of the cell; (b) expressing a second fusion protein in the cell, the second fusion protein comprising a second zinc finger binding domain and a second cleavage half domain, wherein the second zinc finger binding domain binds to a second target site in the region of interest in the genome of the cell, wherein the second target site is different from the first target site; and (c) contacting the cell with a polynucleotide comprising an exogenous nucleic acid sequence; wherein binding of the first fusion protein to the first target site, and binding of the second fusion protein to the second target site, positions the cleavage half-domains such that the genome of the cell is cleaved in the region of interest, thereby resulting in integration of the exogenous sequence into the genome of the cell in the region of interest. In certain embodiments, the integration inactivates gene expression in the region of interest. The exogenous nucleic acid sequence may comprise, for example, a sequence of between 1 and 50 nucleotides in length. Furthermore, the exogenous nucleic acid sequence may encode a detectable amino acid sequence. The region of interest may be in an accessible region of cellular chromatin.

[0018] In any of the methods described herein, the first and second cleavage half-domains are from a Type IIS restriction endonuclease, for example, *Fok*I or *Sts*I. Furthermore, in any of the methods described herein, at least one of the fusion proteins may comprise an alteration in the amino acid sequence of the dimerization interface of the cleavage half-domain.

[0019] In any of the methods described herein, the cell can be a mammalian cell, for example, a human cell. Furthermore, the cell may be arrested in the G2 phase of the cell cycle.

[0020] The present subject matter thus includes, but is not limited to, the following embodiments:

1. A method for expressing the product of an exogenous nucleic acid sequence in a cell, the method comprising:

(a) expressing a first fusion protein in the cell, the first fusion protein comprising a first zinc finger binding domain and a first cleavage half-domain, wherein the first zinc finger binding domain has been engineered to bind to a first target site in a region of interest in the genome of the cell;

(b) expressing a second fusion protein in the cell, the second fusion protein comprising a second zinc finger binding domain and a second cleavage half domain, wherein the second zinc finger binding domain binds to a second target site in the region of interest in the genome of the cell, wherein the second target site is different from the first target site; and

(c) contacting the cell with a polynucleotide comprising an exogenous nucleic acid sequence;

wherein binding of the first fusion protein to the first target site, and binding of the second fusion protein to the second target site, positions the cleavage half-domains such that the genome of the cell is cleaved in the region of interest, thereby resulting in integration of the exogenous sequence into the genome of the cell in the region of interest and expression of the product of the exogenous sequence.

2. The method according to 1, wherein the exogenous nucleic acid sequence comprises a cDNA.

3. The method according to 1, wherein the exogenous sequence comprises a promoter.

4. The method according to 1, wherein the polynucleotide further comprises a first nucleotide sequence that is identical to a first sequence in the region of interest.

5. The method according to 4, wherein the polynucleotide further comprises a second nucleotide sequence that is identical to a second sequence in the region of interest.

6. The method according to 1, wherein the polynucleotide further comprises a first nucleotide sequence that is homologous but non-identical to a first sequence in the region of interest.

7. The method according to 6, wherein the polynucleotide further comprises a second nucleotide sequence that is homologous but non-identical to a second sequence in the region of interest.

8. The method according to 1, wherein the polynucleotide further comprises a first nucleotide sequence that is identical to a first sequence in the region of interest and a second nucleotide sequence that is homologous but non-identical to a second sequence in the region of interest.

9. The method according to 5, wherein the first and second nucleotide sequences flank the exogenous sequence.

10. The method according to 7, wherein the first and second nucleotide sequences flank the exogenous sequence.

11. The method according to 8, wherein the first and second nucleotide sequences flank the exogenous sequence.

12. The method of 1, wherein the polynucleotide is a plasmid.

13. The method of 1, wherein the polynucleotide is a linear DNA molecule.

14. The method according to 1, wherein the region of interest is in an accessible region of cellular chromatin.

15. The method of 1, wherein the region of interest is in a region of the genome that is not essential for viability.

16. The method of 1, wherein the region of interest is in a region of the genome that is transcriptionally active.

17. The method of 1, wherein the region of interest is in a region of the genome that is transcriptionally active and not essential for viability.

18. The method according to 17, wherein the region of interest is the human Rosa26 gene.

19. The method according to 17, wherein the region of interest is the human homologue of the murine Rosa26 gene.

20. The method according to 1, wherein the first and second cleavage half-domains are from a Type IIS restriction endonuclease.

21. The method according to 20, wherein the Type IIS restriction endonuclease is selected from the group consisting of *Fok*I and *Sts*I.

22. The method according to 1, wherein the region of interest is in a chromosome.

23. The method according to 1,wherein the region of interest comprises a gene.

24. The method according to 13, wherein the gene comprises a mutation.

25. The method according to 14, wherein the mutation is selected from the group consisting of a point mutation, a substitution, a deletion, an insertion, a duplication, an inversion and a translocation.

26. The method according to 24, wherein the exogenous nucleic acid sequence comprises the wild-type sequence of the gene.

27. The method according to 24, wherein the exogenous nucleic acid sequence comprises a portion of the wild-type sequence of the gene.

28. The method according to 24, wherein the exogenous nucleic acid sequence comprises a cDNA copy of a transcription product of the gene.

29. The method according to 1, wherein the exogenous nucleic acid sequence encodes a siRNA.

30. The method according to 1, wherein the cell is arrested in the G2 phase of the cell cycle.

31. The method according to 1, wherein at least one of the fusion proteins comprises an alteration in the amino acid sequence of the dimerization interface of the cleavage half-domain.

32. The method according to 1, wherein the cell is a mammalian cell.

33. The method according to 32, wherein the cell is a human cell.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] **Figure 1** shows the nucleotide sequence, in double-stranded form, of a portion of the human hSMC1L1 gene encoding the amino-terminal portion of the protein (SEQ ID NO:1) and the encoded amino acid sequence (SEQ ID NO:2). Target sequences for the hSMC1-specific ZFPs are underlined (one on each DNA strand).

[0022] **Figure 2** shows a schematic diagram of a plasmid encoding a ZFP-*Fok*I fusion for targeted cleavage of the hSMC1 gene.

[0023] **Figure 3 A-D** show a schematic diagram of the hSMC1 gene. Figure 3A shows a schematic of a portion of the human X chromosome which includes the hSMC1 gene. Figure 3B shows a schematic of a portion of the hSMC1 gene including the upstream region (left of +1), the first exon (between +1 and the right end of the arrow labeled "SMC1 coding sequence") and a portion of the first intron. Locations of sequences homologous to the initial amplification primers and to the chromosome-specific primer (see Table 3) are also provided. Figure 3C shows the nucleotide sequence of the human X chromosome in the region of the SMC1 initiation codon (SEQ ID NO: 3), the encoded amino acid sequence (SEQ ID NO: 4), and the target sites for the SMC1-specific zinc finger proteins. Figure 3D shows the sequence of the corresponding region of the donor molecule (SEQ ID NO: 5), with differences between donor and chromosomal sequences underlined. Sequences contained in the donor-specific amplification primer (Table 3) are indicated by double underlining.

[0024] Figure 4 shows a schematic diagram of the hSMC1 donor construct.

[0025] Figure 5 shows PCR analysis of DNA from transfected HEK293 cells. From left, the lanes show results from cells transfected with a plasmid encoding GFP (control plasmid), cells transfected with two plasmids, each of which encodes one of the two hSMC1-specific ZFP-*Fok*I fusion proteins (ZFPs only), cells transfected with two concentrations of the hSMC1 donor plasmid (donor only), and cells transfected with the two ZFP-encoding plasmids and the donor plasmid (ZFPs + donor). See Example 1 for details.

[0026] Figure 6 shows the nucleotide sequence of an amplification product derived from a mutated hSMC1 gene (SEQ ID NO:6) generated by targeted homologous recombination. Sequences derived from the vector into which the amplification product was cloned are single-underlined, chromosomal sequences not present in the donor molecule are indicated by dashed underlining (nucleotides 32-97), sequences common to the donor and the chromosome are not underlined (nucleotides 98-394 and 402-417), and sequences unique to the donor are double-underlined (nucleotides 395-401). Lower-case letters represent sequences that differ between the chromosome and the donor.

[0027] Figure 7 shows the nucleotide sequence of a portion of the human IL2Rγ gene comprising the 3' end of the second intron and the 5' end of third exon (SEQ ID NO:7) and the amino acid sequence encoded by the displayed portion of the third exon (SEQ ID NO:8). Target sequences for the second pair of IL2Rγ-specific ZFPs are underlined. *See* Example 2 for details.

[0028] Figure 8 shows a schematic diagram of a plasmid encoding a ZFP-FokI fusion for targeted cleavage of IL2Rγ gene.

[0029] Figure 9 A-D show a schematic diagram of the IL2Rγ gene. Figure 9A shows a schematic of a portion of the human X chromosome which includes the IL2Rγ gene. Figure 9B shows a schematic of a portion of the IL2Rγ gene including a portion of the second intron, the third exon and a portion of the third intron. Locations of sequences homologous to the initial amplification primers and to the chromosome-specific primer (see Table 5) are also provided. Figure 9C shows the nucleotide sequence of the human X chromosome in the region of the third exon of the IL2Rγ gene (SEQ ID NO: 9), the encoded amino acid sequence (SEQ ID NO: 10), and the target sites for the first pair of IL2Rγ-specific zinc finger proteins. Figure 9D shows the sequence of the corresponding region of the donor molecule (SEQ ID NO: 11), with differences between donor and chromosomal sequences underlined. Sequences contained in the donor-specific amplification primer (Table 5) are indicated by double overlining.

[0030] **Figure 10** shows a schematic diagram of the IL2Rγ donor construct.

[0031] **Figure 11** shows PCR analysis of DNA from transfected K652 cells. From left, the lanes show results from cells transfected with two plasmids, each of which encodes one of a pair of IL2Rγ -specific ZFP-FokI fusion proteins (ZFPs only, lane 1), cells transfected with two concentrations of the IL2Rγ donor plasmid (donor only, lanes 2 and 3), and cells transfected with the two ZFP-encoding plasmids and the donor plasmid (ZFPs + donor, lanes 4-7). Each of the two pairs of IL2Rγ-specific ZFP-*Fok*I fusions were used (identified as "pair 1" and "pair 2") and use of both pairs resulted in production of the diagnostic amplification product (labeled "expected chimeric product" in the Figure). See Example 2 for details.

[0032] **Figure 12** shows the nucleotide sequence of an amplification product derived from a mutated IL2Rγ gene (SEQ ID NO:12) generated by targeted homologous recombination. Sequences derived from the vector into which the amplification product was cloned are single-underlined, chromosomal sequences not present in the donor molecule are indicated by dashed underlining (nucleotides 460-552), sequences common to the donor and the chromosome are not underlined (nucleotides 32-42 and 59-459), and a stretch of sequence containing nucleotides which distinguish donor sequences from chromosomal sequences is double-underlined (nucleotides 44-58). Lower-case letters represent nucleotides whose sequence differs between the chromosome and the donor.

[0033] **Figure 13** shows the nucleotide sequence of a portion of the human beta-globin gene encoding segments of the core promoter, the first two exons and the first intron (SEQ ID NO:13). A missense mutation changing an A (in boldface and underlined) at position 5212541 on Chromosome 11 (BLAT, UCSC Genome Bioinformatics site) to a T results in sickle cell anemia. A first zinc finger/*Fok*I fusion protein was designed such that the primary contacts were with the underlined 12-nucleotide sequence AAGGTGAACGTG (nucleotides 305-316 of SEQ ID NO:13), and a second zinc finger/*Fok*I fusion protein was designed such that the primary contacts were with the complement of the underlined 12-nucleotide sequence CCGTTACTGCCC (nucleotides 325-336 of SEQ ID NO:13).

[0034] **Figure 14** is a schematic diagram of a plasmid encoding ZFP-FokI fusion for targeted cleavage of the human beta globin gene.

[0035] **Figure 15** is a schematic diagram of the cloned human beta globin gene showing the upstream region, first and second exons, first intron and primer binding sites.

[0036] **Figure 16** is a schematic diagram of the beta globin donor construct, pCR4-TOPO-HBBdonor.

[0037] **Figure 17** shows PCR analysis of DNA from cells transfected with two pairs of β-globin-specific ZFP nucleases and a beta globin donor plasmid. The panel on the left is a loading control in which the initial amp 1 and initial amp 2 primers (Table 7) were used for amplification. In the experiment shown in the right panel, the "chromosome-specific and "donor-specific" primers (Table 7) were used for amplification. The leftmost lane in each panel contains molecular weight markers and the next lane shows amplification products obtained from mock-transfected cells. Remaining lanes, from left to right, show amplification product from cells transfected with: a GFP-encoding plasmid, 100ng of each ZFP/*Fok*I-encoding plasmid, 200ng of each ZFP/*Fok*I-encoding plasmid, 200 ng donor plasmid, 600 ng donor plasmid, 200 ng donor plasmid + 100 ng of each ZFP/*Fok*I-encoding plasmid, and 600 ng donor plasmid + 200 ng of each ZFP/*Fok*I-encoding plasmid.

[0038] **Figure 18** shows the nucleotide sequence of an amplification product derived from a mutated beta-globin gene (SEQ ID NO:14) generated by targeted homologous recombination. Chromosomal sequences not present in the donor molecule are indicated by dashed underlining (nucleotides 1-72), sequences common to the donor and the chromosome are not underlined (nucleotides 73-376), and a stretch of sequence containing nucleotides which distinguish donor sequences from chromosomal sequences is double-underlined (nucleotides 377-408). Lower-case letters represent nucleotides whose sequence differs between the chromosome and the donor.

[0039] **Figure 19** shows the nucleotide sequence of a portion of the fifth exon of the Interleukin-2 receptor gamma chain (IL-2Rγ) gene (SEQ ID NO:15). Also shown (underlined) are the target sequences for the 5-8 and 5-10 ZFP/*Fok*I fusion proteins. See Example 5 for details.

[0040] **Figure 20** shows the amino acid sequence of the 5-8 ZFP/*Fok*I fusion targeted to exon 5 of the human IL-2Rγ gene (SEQ ID NO:16). Amino acid residues

1-17 contain a nuclear localization sequence (NLS, underlined); residues 18-130 contain the ZFP portion, with the recognition regions of the component zinc fingers shown in boldface; the ZFP-*Fok*I linker (ZC linker, underlined) extends from residues 131 to 140 and the *Fok*I cleavage half-domain begins at residue 141 and extends to the end of the protein at residue 336. The residue that was altered to generate the Q486E mutation is shown underlined and in boldface.

[0041] **Figure 21** shows the amino acid sequence of the 5-10 ZFP/*Fok*I fusion targeted to exon 5 of the human IL-2Rγ gene (SEQ ID NO:17). Amino acid residues 1-17 contain a nuclear localization sequence (NLS, underlined); residues 18-133 contain the ZFP portion, with the recognition regions of the component zinc fingers shown in boldface; the ZFP-*Fok*I linker (ZC linker, underlined) extends from residues 134 to 143 and the *Fok*I cleavage half-domain begins at residue 144 and extends to the end of the protein at residue 339. The residue that was altered to generate the E490K mutation is shown underlined and in boldface.

[0042] **Figure 22** shows the nucleotide sequence of the enhanced Green Fluorescent Protein gene (SEQ ID NO:18) derived from the *Aequorea victoria* GFP gene (Tsien (1998) *Ann. Rev. Biochem.* **67**:509-544). The ATG initiation codon, as well as the region which was mutagenized, are underlined.

[0043] **Figure 23** shows the nucleotide sequence of a mutant defective eGFP gene (SEQ ID NO:19). Binding sites for ZFP-nucleases are underlined and the region between the binding sites corresponds to the region that was modified.

[0044] **Figure 24** shows the structures of plasmids encoding Zinc Finger Nucleases targeted to the eGFP gene.

[0045] **Figure 25** shows an autoradiogram of a 10% acrylamide gel used to analyze targeted DNA cleavage of a mutant eGFP gene by zinc finger endonucleases. See Example 8 for details.

[0046] **Figure 26** shows the structure of plasmid pcDNA4/TO/GFPmut (see Example 9).

[0047] **Figure 27** shows levels of eGFPmut mRNA, normalized to GAPDH mRNA, in various cell lines obtained from transfection of human HEK293 cells. Light bars show levels in untreated cells; dark bars show levels in cell that had been treated with 2 ng/ml doxycycline. See Example 9 for details.

[0048] **Figure 28** shows the structure of plasmid pCR(R)4-TOPO-GFPdonor5. See Example 10 for details.

[0049] **Figure 29** shows the nucleotide sequence of the eGFP insert in pCR(R)4-TOPO-GFPdonor5 (SEQ ID NO:20). The insert contains sequences encoding a portion of a non-modified enhanced Green Fluorescent Protein, lacking an initiation codon. See Example 10 for details.

[0050] **Figure 30** shows a FACS trace of T18 cells transfected with plasmids encoding two ZFP nucleases and a plasmid encoding a donor sequence, that were arrested in the G2 phase of the cell cycle 24 hours post-transfection with 100 ng/ml nocodazole for 48 hours. The medium was replaced and the cells were allowed to recover for an additional 48 hours, and gene correction was measured by FACS analysis. *See* Example 11 for details.

[0051] **Figure 31** shows a FACS trace of T18 cells transfected with plasmids encoding two ZFP nucleases and a plasmid encoding a donor sequence, that were arrested in the G2 phase of the cell cycle 24 hours post-transfection with 0.2 μM vinblastine for 48 hours. The medium was replaced and the cells were allowed to recover for an additional 48 hours, and gene correction was measured by FACS analysis. *See* Example 11 for details.

[0052] **Figure 32** shows the nucleotide sequence of a 1,527 nucleotide eGFP insert in pCR(R)4-TOPO (SEQ ID NO:21). The sequence encodes a non-modified enhanced Green Fluorescent Protein lacking an initiation codon. *See,* Example 13 for details.

[0053] **Figure 33** shows a schematic diagram of an assay used to measure the frequency of editing of the endogenous human IL-2Rγ gene. *See,* Example 14 for details.

[0054] **Figure 34** shows autoradiograms of acrylamide gels used in an assay to measure the frequency of editing of an endogenous cellular gene by targeted cleavage and homologous recombination. The lane labeled "GFP" shows assay results from a control in which cells were transfected with an eGFP-encoding vector; the lane labeled "ZFPs only" shows results from another control experiment in which cells were transfected with the two ZFP/nuclease-encoding plasmids (50 ng of each) but not with a donor sequence. Lanes labeled "donor only" show results from a control experiment in which cells were transfected with 1 μg of donor plasmid but not with the ZFP/nuclease-encoding plasmids. In the experimental lanes, 50Z refers to cells transfected with 50 ng of each ZFP/nuclease expression plasmid, 100Z refers to cells transfected with 100 ng of each ZFP/nuclease expression plasmid, 0.5D refers to cells transfected with 0.5 μg of the donor plasmid, and 1D refers to cells transfected with 1.0 μg of the donor plasmid. "+" refers to cells that were exposed to 0.2 μM vinblastine; "-" refer to cells that were not exposed to vinblastine. "wt" refers to the fragment obtained after *Bsr*BI digestion of amplification products obtained from chromosomes containing the wild-type chromosomal IL-2Rγ gene; "rflp" refers to the two fragments (of approximately equal molecular weight) obtained after *Bsr*BI digestion of amplification products obtained from chromosomes containing sequences from the donor plasmid which had integrated by homologous recombination.

[0055] **Figure 35** shows an autoradiographic image of a four-hour exposure of a gel used in an assay to measure targeted recombination at the human IL-2Rγ locus in K562 cells. "wt" identifies a band that is diagnostic for chromosomal DNA containing the native K562 IL-2Rγ sequence; "rflp" identifies a doublet diagnostic for chromosomal DNA containing the altered IL-2Rγ sequence present in the donor DNA molecule. The symbol "+" above a lane indicates that cells were treated with 0.2 μM vinblastine; the symbol "-" indicates that cells were not treated with vinblastine. The numbers in the "ZFP + donor" lanes indicate the percentage of total chromosomal DNA containing sequence originally present in the donor DNA molecule, calculated using the "peak finder, automatic baseline" function of Molecular Dynamics' ImageQuant v. 5.1 software as described in Ch. 8 of the manufacturer's manual (Molecular Dynamics ImageQuant User's Guide; part 218-415). "Untr" indicates untransfected cells. See Example 15 for additional details.

[0056] **Figure 36** shows an autoradiographic image of a four-hour exposure of a gel used in an assay to measure targeted recombination at the human IL-2Rγ locus in K562 cells. "wt" identifies a band that is diagnostic for chromosomal DNA containing the native K562 IL-2Rγ sequence; "rflp" identifies a band that is diagnostic for chromosomal DNA containing the altered IL-2Rγ sequence present in the donor DNA molecule. The symbol "+" above a lane indicates that cells were treated with 0.2 μM vinblastine; the symbol "-" indicates that cells were not treated with vinblastine. The numbers beneath the "ZFP + donor" lanes indicate the percentage of total chromosomal DNA containing sequence originally present in the donor DNA molecule, calculated as described in Example 35. See Example 15 for additional details.

[0057] **Figure 37** shows an autoradiogram of a four-hour exposure of a DNA blot probed with a fragment specific to the human IL-2Rγ gene. The arrow to the right of the image indicates the position of a band corresponding to genomic DNA whose sequence has been altered by homologous recombination. The symbol "+" above a lane indicates that cells were treated with 0.2 μM vinblastine; the symbol "-" indicates that cells were not treated with vinblastine. The numbers beneath the "ZFP + donor" lanes indicate the percentage of total chromosomal DNA containing sequence originally present in the donor DNA molecule, calculated as described in Example 35. See Example 15 for additional details.

[0058] **Figure 38** shows autoradiographic images of gels used in an assay to measure targeted recombination at the human IL-2Rγ locus in $CD34^+$ human bone marrow cells. The left panel shows a reference standard in which the stated percentage of normal human genomic DNA (containing a *Mae*II site) was added to genomic DNA from Jurkat cells (lacking a *Mae*II site), the mixture was amplified by PCR to generate a radiolabelled amplification product, and the amplification product was digested with *Mae*II. "wt" identifies a band representing undigested DNA, and "rflp" identifies a band resulting from *Mae*II digestion.

[0059] The right panel shows results of an experiment in which $CD34^+$ cells were transfected with donor DNA containing a *Bsr*BI site and plasmids encoding zinc finger-*Fok*I fusion endonucleases. The relevant genomic region was then amplified and labeled, and the labeled amplification product was digested with *Bsr*BI. "GFP" indicates control cells that were transfected with a GFP-encoding plasmid; "Donor only" indicates control cells that were transfected only with donor DNA, and "ZFP + Donor" indicates cells that were transfected with donor DNA and with plasmids encoding the zinc finger/*Fok*I nucleases. "wt" identifies a band that is diagnostic for chromosomal DNA containing the native IL-2Rγ sequence; "rflp" identifies a band that is diagnostic for chromosomal DNA containing the altered IL-2Rγ sequence present in the donor DNA molecule. The rightmost lane contains DNA size markers. See Example 16 for additional details.

[0060] **Figure 39** shows an image of an immunoblot used to test for Ku70 protein levels in cells transfected with Ku70-targeted siRNA. The T7 cell line (Example 9, Figure 27) was transfected with two concentrations each of siRNA from two different siRNA pools (see Example 18). Lane 1: 70 ng of siRNA pool D; Lane 2: 140 ng of siRNA pool D; Lane 3: 70 ng of siRNA pool E; Lane 4: 140 ng of siRNA pool E.. "Ku70" indicates the band representing the Ku70 protein; "TFIIB" indicates a band representing the TFIIB transcription factor, used as a control.

[0061] **Figure 40** shows the amino acid sequences of four zinc finger domains targeted to the human β-globin gene: sca-29b (SEQ ID NO:22); sca-36a (SEQ ID NO:23); sca-36b (SEQ ID NO:24) and sca-36c (SEQ ID NO:25). The target site for the sca-29b domain is on one DNA strand, and the target sites for the sca-36a, sca-36b and sca-36c domains are on the opposite strand. *See* Example 20.

[0062] **Figure 41** shows results of an *in vitro* assay, in which different combinations of zinc finger/*Fok*I fusion nucleases (ZFNs) were tested for sequence-specific DNA cleavage. The lane labeled "U" shows a sample of the DNA template. The next four lanes show results of incubation of the DNA template with each of four β-globin-targeted ZFNs (see Example 20 for characterization of these ZFNs). The rightmost three lanes show results of incubation of template DNA with the sca-29b ZFN and one of the sca-36a, sca-36b or sca-36c ZFNs (all of which are targeted to the strand opposite that to which sca-29b is targeted).

[0063] **Figure 42** shows levels of eGFP mRNA in T18 cells (bars) as a function of doxycycline concentration (provided on the abscissa). The number above each bar represents the percentage correction of the eGFP mutation, in cells transfected with donor DNA and plasmids encoding eGFP-targeted zinc finger nucleases, as a function of doxycycline concentration.

[0064] **Figure 43A-C** show schematic diagrams of different fusion protein configurations. Figure 43A shows two fusion proteins, in which the zinc finger domain is nearest the N-terminus and the *Fok*I cleavage half-domain is nearest the C-terminus, binding to DNA target sites on opposite strands whose 5' ends are proximal to each other. Figure 43B shows two fusion proteins, in which the *Fok*I cleavage half-domain is nearest the N-terminus and the zinc finger domain is nearest the C-terminus, binding to DNA target sites on opposite strands whose 3' ends are proximal to each other. Figure 43C shows a first protein in which the *Fok*I cleavage half-domain is nearest the N-terminus and the zinc finger domain is nearest the C-terminus and a second protein in which the zinc finger domain is nearest the N-terminus and the *Fok*I cleavage half-domain is nearest the C-terminus, binding to DNA target sites on the same strand, in which the target site for the first protein is upstream (*i.e.* to the 5' side) of the binding site for the second protein.

[0065] In all examples, three-finger proteins are shown binding to nine-nucleotide target sites. 5' and 3' polarity of the DNA strands is shown, and the N-termini of the fusion proteins are identified.

[0066] **Figure 44** is an autoradiogram of an acrylamide gel in which cleavage of a model substrate by zinc finger endonucleases was assayed. Lane 1 shows the migration of uncleaved substrate. Lane 2 shows substrate after incubation with the IL2-1R zinc finger/*Fok*I fusion protein. Lane 3 shows substrate after incubation with the5-9DR zinc finger/*Fok*I fusion protein. Lane 4 shows substrate after incubation with both proteins. Approximate sizes (in base pairs) of the substrate and its cleavage products are shown to the right of the image. Below the image, the nucleotide sequence (SEQ ID NO:211) of the portion of the substrate containing the binding sites for the 5-9D and IL2-1 zinc finger binding domains is shown. The binding sites are identified and indicated by underlining.

[0067] **Figure 45** is an autoradiogram of an acrylamide gel in which cleavage of a model substrate by zinc finger endonucleases was assayed. Lane 1 shows the migration of uncleaved substrate. Lane 2 shows substrate after incubation with the IL2-1C zinc finger/*Fok*I fusion protein. Lane 3 shows substrate after incubation with the IL2-1R zinc finger/*Fok*I fusion protein. Lane 4 shows substrate after incubation with the5-9DR zinc finger/*Fok*I fusion protein. Lane 5 shows substrate after incubation with both the IL2-1R and 5-9DR fusion proteins. Lane 6 shows substrate after incubation with both the IL2-1C and 5-9DR proteins. Approximate sizes (in base pairs) of the substrate and its cleavage products are shown to the right of the image. Below the image, the nucleotide sequence (SEQ ID NO:212) of the portion of the substrate containing the binding sites for the 5-9D and IL2-1 zinc finger binding domains is shown. The binding sites are identified and indicated by underlining.

[0068] **Figure 46** is a schematic diagram of a plasmid containing mutant eGFP coding sequences containing an insertion of sequences from exon 5 of the IL-2Rγ gene. See Example 29 for details.

[0069] **Figure 47** shows an autoradiographic image of a gel in which amplification products of DNA from transfected K562 cells were incubated with the restriction enzyme *Stu* I. Headings above the lane indicate DNA from cells transfected with a GFP-encoding plasmid (GFP); DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins (ZFNs); DNA from cells transfected with a plasmid containing a 12-nucleotide pair exogenous sequence (including a *Stu*I recognition site) flanked on either side by 750 nucleotide pairs of sequence homologous to exon 5 of the IL-2Rγ gene, wherein the two exon 5-homologous sequences are adjacent to one another in the wild-type IL-2Rγ gene (donor); and DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins and a plasmid containing a 12-nucleotide pair exogenous sequence (including a *Stu*I recognition site) flanked on either side by 750 nucleotide pairs of sequence homologous to exon 5 and adjacent regions of the IL-2Rγ gene, wherein the two exon 5-homologous sequences are adjacent to one another in the wild-type IL-2Rγ gene (ZFNs + donor). Bands arising from chromosomes containing wild-type IL-2R sequences ("WT") and chromosomes into which exogenous sequences have been integrated ("+patch") are indicated. The rightmost lane contains molecular weight markers. See also Example 33.

[0070] **Figure 48** shows images of gels in which amplification products of DNA from transfected K562 cells were analyzed. Headings above the lane indicate DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins (ZFNs); DNA from cells transfected with a plasmid containing a 720 nucleotide pair open reading frame encoding eGFP (donor 1); DNA from cells transfected with a plasmid containing a 924 nucleotide pair sequence that included an eGFP open reading frame and a downstream polyadenylation signal (donor 2); DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins and with a plasmid containing a 720 nucleotide pair open reading frame encoding eGFP (ZFNs + donor 1) and DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins and with a plasmid containing a 924 nucleotide pair sequence that included an eGFP open reading frame and a downstream polyadenylation signal (ZFNs + donor 2). The leftmost and rightmost lanes of the top panel contains molecular weight markers. The top panel is a photograph of an ethidium bromide-stained gel; the bottom panel shows an autoradiogram of a gel from a separate experiment in which labeled amplification products were analyzed. See also Example 34.

[0071] **Figure 49** is a schematic diagram describing an experiment in which a "therapeutic half-gene" was introduced into the endogenous human IL-2Rγ gene. The top line represents chromosomal IL-2Rγ sequences, and the middle line represents donor sequences, with exons indicated by boxes and introns by horizontal lines. Numbers inside the boxes identify the exons of the IL-2Rγ gene, with "5" representing the fifth exon of the chromosomal IL-2R gene, "5u" representing the upstream portion of the fifth exon, "5d" representing the downstream portion of the fifth exon and "5d(m)" representing the downstream portion of the fifth exon containing several silent sequence changes (*i.e.*, changes that do not alter the encoded amino acid sequence). Diagonal lines demarcate regions of homology between donor and chromosomal sequences. The bottom line shows the expected product of homologous recombination, in which exons 5d(m), 6, 7, and 8 are inserted within the fifth exon of the chromosomal gene. See also Example 35.

[0072] **Figure 50** shows an autoradiogram of a gel in which amplification products of DNA from transfected K562 cells were analyzed. Headings above the lane indicate DNA from control cells transfected with a vector encoding green fluorescent protein ("GFP") and from experimental cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins and a plasmid containing a 720-nucleotide cDNA construct containing part of exon 5 and exons 6, 7 and 8 of the IL-2Rγ gene, flanked on either side by 750 nucleotide pairs of sequence homologous to exon 5 and surrounding regions of the IL-2Rγ gene, wherein the two exon 5-homologous sequences are adjacent to one another in the wild-type IL-2Rγ gene (ZFNs + donor). Bands arising from chromosomes containing wild-type IL-2R sequences ("WT") and chromosomes into which exogenous sequences have been integrated ("+ORF") are indicated. See also Example 35.

[0073] **Figure 51** shows the design and results of an experiment in which a 7.7 kbp antibody expression construct was inserted into the endogenous chromosomal IL-2Rγ gene. The upper portion of the figure is a schematic diagram showing the result of homology-dependent targeted integration of a 7.7 kilobase pair expression construct (shaded) into exon 5 of the endogenous chromosomal IL-2Rγ gene. Arrows indicate the locations and polarities of amplification primers used to detect the junctions between exogenous and endogenous sequences that result from targeted integration.

[0074] The lower portion of the figure shows photographs of ethidium bromide-stained gels in which amplification products were analyzed. The left panel shows products of cellular DNA that was amplified using primers that detect the upstream junction (Primer set A) and the right panel shows products of cellular DNA that was amplified using primers that detect the downstream junction (Primer set B). DNA samples used as templates for amplification are identified below the gel, as follows: DNA from cells transfected with a vector encoding green fluorescent protein (GFP); DNA from cells transfected only with the donor DNA molecule containing the 7.7 kbp expression construct (don.); DNA from cells transfected with a vector encoding the 5-8G and 5-9D ZFP/*Fok*I fusion proteins and with the donor DNA molecule (ZFNs + donor). The topology of the donor DNA (circular or linear) is also indicated. See also Example 36.

[0075] **Figure 52** is an autoradiogram of a gel in which amplification products from the CHO DHFR gene were analyzed for mismatches using a Cel-1 assay. Amplification products were obtained from wild-type CHO cell DNA (W) or DNA from CHO cells that had been treated with zinc finger nucleases (Mu), and were then exposed to Cel-1 nuclease (+) or not (-), as indicated above the gel. To the right of the gel, bands indicative of wild-type DHFR sequences (WT) and mutant DHFR sequences containing a 157-nucleotide insertion (Mutant) are indicated. See Example 37 for details.

[0076] **Figure 53** shows a portion of the nucleotide sequence of the CHO dihydrofolate reductase (DHFR) gene (upper lines) and a portion of the nucleotide sequence of a mutant DHFR gene generated by targeted homology-independent integration of exogenous sequences (lower lines). Target sequences for the zinc finger nucleases described in Table 28 are boxed; changes from wild-type sequence are underlined. See also Example 37.

[0077] **Figure 54** shows the amino acid sequences of the wild-type *Fok*I cleavage half-domain and of several mutant cleavage half-domains containing alterations in the amino acid sequence of the dimerization interface. Positions at which the sequence was altered (amino acids 486, 490 and 538) are underlined.

DETAILED DESCRIPTION

[0078] Disclosed herein are compositions and methods useful for targeted cleavage of cellular chromatin and for targeted alteration of a cellular nucleotide sequence, *e.g.*, by targeted cleavage followed by non-homologous end joining (with or without an exogenous sequence inserted therebetween) or 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. Genomic sequences include those present in chromosomes, episomes, organellar genomes (*e.g.*, mitochondria, chloroplasts), artificial chromosomes and any other type of nucleic acid present in a cell such as, for example, amplified sequences, double minute chromosomes and the genomes of endogenous or infecting bacteria and viruses. Genomic sequences can be normal (*i.e.*, wild-type) or mutant; mutant sequences can comprise, for example, insertions, deletions, translocations, rearrangements, and/or point mutations. A genomic sequence can also comprise one of a number of different alleles.

[0079] Compositions useful for targeted cleavage and recombination include fusion proteins comprising a cleavage domain (or a cleavage half-domain) and a zinc finger binding domain, polynucleotides encoding these proteins and combinations of polypeptides and polypeptide-encoding polynucleotides. A zinc finger binding domain can comprise one or more zinc fingers (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9 or more zinc fingers), and can be engineered to bind to any genomic sequence. Thus, by identifying a target genomic region of interest at which cleavage or recombination is desired, one can, according to the methods disclosed herein, construct one or more fusion proteins comprising a cleavage domain (or cleavage half-domain) and a zinc finger domain engineered to recognize a target sequence in said genomic region. The presence of such a fusion protein (or proteins) in a cell will result in binding of the fusion protein(s) to its (their) binding site(s) and cleavage within or near said genomic region. Moreover, if an exogenous polynucleotide homologous to the genomic region is also present in such a cell, homologous recombination occurs at a high rate between the genomic region and the exogenous polynucleotide.

General

[0080] 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, 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

[0081] 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.

[0082] 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 a corresponding naturally-occurring amino acids.

[0083] "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.

[0084] 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 activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0085] 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.

[0086] Zinc finger binding domains can be "engineered" to bind to a predetermined nucleotide sequence. Non-limiting examples of methods for engineering zinc finger proteins are design and selection. A designed zinc finger 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 designs and binding data. See, for example, US Patents 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.

[0087] A "selected" zinc finger protein 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 and WO 02/099084.

[0088] 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.

[0089] 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 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.

[0090] 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. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, WI) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). A preferred method of establishing percent identity in the context of the present disclosure is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects sequence identity. Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found on the internet. 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.

[0091] 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 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).

[0092] 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.

[0093] 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 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).

[0094] 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.

[0095] 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. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

[0096] "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, 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.

[0097] "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.

[0098] A "cleavage domain" comprises one or more polypeptide sequences which possesses catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides.

[0099] 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).

[0100] "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.

[0101] 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.

[0102] 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.

[0103] 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.

[0104] 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.

[0105] 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 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 or a malfunctioning version of a normally-functioning endogenous molecule.

[0106] 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.

[0107] 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.

[0108] 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, 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.

[0109] 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.

[0110] 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.

[0111] 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.

[0112] "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 a mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0113] "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.

[0114] "Eucaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells.

[0115] 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.

[0116] 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.

[0117] 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 ZFP DNA-binding domain is fused to a cleavage domain, the ZFP DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the ZFP 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.

[0118] 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 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 ore 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.

Target sites

[0119] The disclosed methods and compositions include fusion proteins comprising a cleavage domain (or a cleavage half-domain) and a zinc finger domain, in which the zinc finger domain, by binding to a sequence in cellular chromatin (*e.g.*, a target site or a binding site), directs the activity of the cleavage domain (or cleavage half-domain) to the vicinity of the sequence and, hence, induces cleavage in the vicinity of the target sequence. As set forth elsewhere in this disclosure, a zinc finger domain can be engineered to bind to virtually any desired sequence. Accordingly, after identifying a region of interest containing a sequence at which cleavage or recombination is desired, one or more zinc finger binding domains can be engineered to bind to one or more sequences in the region of interest. Expression of a fusion protein comprising a zinc finger binding domain and a cleavage domain (or of two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain), in a cell, effects cleavage in the region of interest.

[0120] Selection of a sequence in cellular chromatin for binding by a zinc finger domain (*e.g.,* a target site) can be accomplished, for example, according to the methods disclosed in co-owned US Patent No. 6,453,242 (Sept. 17, 2002), which also discloses methods for designing ZFPs to bind to a selected sequence. It will be clear to those skilled in the art that simple visual inspection of a nucleotide sequence can also be used for selection of a target site. Accordingly, any means for target site selection can be used in the claimed methods.

[0121] Target sites are generally composed of a plurality of adjacent target subsites. A target subsite refers to the sequence (usually either a nucleotide triplet, or a nucleotide quadruplet that can overlap by one nucleotide with an adjacent quadruplet) bound by an individual zinc finger. *See*, for example, WO 02/077227. If the strand with which a zinc finger protein makes most contacts is designated the target strand "primary recognition strand," or "primary contact strand," some zinc finger proteins bind to a three base triplet in the target strand and a fourth base on the non-target strand. A target site generally has a length of at least 9 nucleotides and, accordingly, is bound by a zinc finger binding domain comprising at least three zinc fingers. However binding of, for example, a 4-finger binding domain to a 12-nucleotide target site, a 5-finger binding domain to a 15-nucleotide target site or a 6-finger binding domain to an 18-nucleotide target site, is also possible. As will be apparent, binding of larger binding domains (*e.g.*, 7-, 8-, 9-finger and more) to longer target sites is also possible.

[0122] It is not necessary for a target site to be a multiple of three nucleotides. For example, in cases in which cross-strand interactions occur (see, *e.g.*, US Patent 6,453,242 and WO 02/077227), one or more of the individual zinc fingers of a multi-finger binding domain can bind to overlapping quadruplet subsites. As a result, a three-finger protein can bind a 10-nucleotide sequence, wherein the tenth nucleotide is part of a quadruplet bound by a terminal finger, a four-finger protein can bind a 13-nucleotide sequence, wherein the thirteenth nucleotide is part of a quadruplet bound by a terminal finger, *etc.*

[0123] The length and nature of amino acid linker sequences between individual zinc fingers in a multi-finger binding domain also affects binding to a target sequence. For example, the presence of a so-called "non-canonical linker," "long linker" or "structured linker" between adjacent zinc fingers in a multi-finger binding domain can allow those fingers to bind subsites which are not immediately adjacent. Non-limiting examples of such linkers are described, for example, in US Patent No. 6,479,626 and WO 01/53480. Accordingly, one or more subsites, in a target site for a zinc finger binding domain, can be separated from each other by 1, 2, 3, 4, 5 or more nucleotides. To provide but one example, a four-finger binding domain can bind to a 13-nucleotide target site comprising, in sequence, two contiguous 3-nucleotide subsites, an intervening nucleotide, and two contiguous triplet subsites.

[0124] Distance between sequences (*e.g.*, target sites) refers to the number of nucleotides or nucleotide pairs intervening between two sequences, as measured from the edges of the sequences nearest each other.

[0125] In certain embodiments in which cleavage depends on the binding of two zinc finger domain/cleavage half-domain fusion molecules to separate target sites, the two target sites can be on opposite DNA strands. In other embodiments, both target sites are on the same DNA strand.

Zinc finger binding domains

[0126] A zinc finger binding domain comprises one or more zinc fingers. Miller *et al.* (1985) *EMBO J.* **4:**1609-1614; Rhodes (1993) *Scientific American* **Feb.:**56-65; US Patent No. 6,453,242. Typically, a single zinc finger domain is about 30 amino acids in length. Structural studies have demonstrated that each zinc finger domain (motif) contains two beta sheets (held in a beta turn which contains the two invariant cysteine residues) and an alpha helix (containing the two invariant histidine residues), which are held in a particular conformation through coordination of a zinc atom by the two cysteines and the two histidines.

[0127] Zinc fingers include both canonical C_2H_2 zinc fingers (*i.e.*, those in which the zinc ion is coordinated by two cysteine and two histidine residues) and non-canonical zinc fingers such as, for example, C_3H zinc fingers (those in which the zinc ion is coordinated by three cysteine residues and one histidine residue) and C_4 zinc fingers (those in which the zinc ion is coordinated by four cysteine residues). See also WO 02/057293.

[0128] Zinc finger binding domains can be engineered to bind to a sequence 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.

[0129] 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.

[0130] Enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0131] Since an individual zinc finger binds to a three-nucleotide (*i.e.*, triplet) sequence (or a four-nucleotide sequence which can overlap, by one nucleotide, with the four-nucleotide binding site of an adjacent zinc finger), the length of a sequence to which a zinc finger binding domain is engineered to bind (*e.g.*, a target sequence) will determine the number of zinc fingers in an engineered zinc finger binding domain. For example, for ZFPs in which the finger motifs do not bind to overlapping subsites, a six-nucleotide target sequence is bound by a two-finger binding domain; a nine-nucleotide target sequence is bound by a three-finger binding domain, *etc.* As noted herein, binding sites for individual zinc fingers (*i.e.*, subsites) in a target site need not be contiguous, but can be separated by one or several nucleotides, depending on the length and nature of the amino acids sequences between the zinc fingers (*i.e.*, the inter-finger linkers) in a multi-finger binding domain.

[0132] In a multi-finger zinc finger binding domain, adjacent zinc fingers can be separated by amino acid linker sequences of approximately 5 amino acids (so-called "canonical" inter-finger linkers) or, alternatively, by one or more non-canonical linkers. See, *e.g.*, co-owned US Patent Nos. 6,453,242 and 6,534,261. For engineered zinc finger binding domains comprising more than three fingers, insertion of longer ("non-canonical") inter-finger linkers between certain of the zinc fingers may be preferred as it may increase the affinity and/or specificity of binding by the binding domain. See, for example, U.S. Patent No. 6,479,626 and WO 01/53480. Accordingly, multi-finger zinc finger binding domains can also be characterized with respect to the presence and location of non-canonical inter-finger linkers. For example, a six-finger zinc finger binding domain comprising three fingers (joined by two canonical inter-finger linkers), a long linker and three additional fingers (joined by two canonical inter-finger linkers) is denoted a 2x3 configuration. Similarly, a binding domain comprising two fingers (with a canonical linker therebetween), a long linker and two additional fingers (joined by a canonical linker) is denoted a 2x2 protein. A protein comprising three two-finger units (in each of which the two fingers are joined by a canonical linker), and in which each two-finger unit is joined to the adjacent two finger unit by a long linker, is referred to as a 3x2 protein.

[0133] The presence of a long or non-canonical inter-finger linker between two adjacent zinc fingers in a multi-finger binding domain often allows the two fingers to bind to subsites which are not immediately contiguous in the target sequence. Accordingly, there can be gaps of one or more nucleotides between subsites in a target site; *i.e.*, a target site can contain one or more nucleotides that are not contacted by a zinc finger. For example, a 2x2 zinc finger binding domain can bind to two six-nucleotide sequences separated by one nucleotide, *i.e.*, it binds to a 13-nucleotide target site. *See* also Moore *et al.* (2001a) Proc. Natl. Acad. Sci. USA **98:**1432-1436; Moore *et al.* (2001b) Proc. Natl. Acad. Sci. USA **98:**1437-1441 and WO 01/53480.

[0134] As mentioned previously, a target subsite is a three- or four-nucleotide sequence that is bound by a single zinc finger. For certain purposes, a two-finger unit is denoted a binding module. A binding module can be obtained by, for example, selecting for two adjacent fingers in the context of a multi-finger protein (generally three fingers) which bind a particular six-nucleotide target sequence. Alternatively, modules can be constructed by assembly of individual zinc fingers. See also WO 98/53057 and WO 01/53480.

Cleavage domains

[0135] The cleavage domain portion of the fusion proteins disclosed herein 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. *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.

[0136] Similarly, a cleavage half-domain (*e.g.*, fusion proteins comprising a zinc finger binding domain and 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 nucleotides or more). In general, the point of cleavage lies between the target sites.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] Exemplary Type IIS restriction enzymes are listed in Table 1. 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.

Table 1: Some Type IIS Restriction Enzymes

Aar I	BsrB I	SspD5 I
Ace III	BsrD I	Sth132 I
Aci I	BstF5 I	Sts I
Alo I	Btr I	TspDT I
Bae I	Bts I	TspGW I
Bbr7 I	Cdi I	Tth111 II
Bbv I	CjeP I	UbaP I
Bbv II	Drd II	Bsa I
BbvC I	Eci I	BsmB I
Bcc I	Eco31 I	

Bce83 I	Eco57 I
BceA I	Eco57M I
Bcef I	Esp3 I
Bcg I	Fau I
BciV I	Fin I
Bfi I	Fok I
Bin I	Gdi II
Bmg I	Gsu I
Bpu10 I	Hga I
BsaX I	Hin4 II
Bsb I	Hph I
BscA I	Ksp632 I
BscG I	Mbo II
BseR I	Mly I
BseY I	Mme I
Bsi I	Mnl I
Bsm I	Pfl1108 I
BsmA I	Ple I
BsmF I	Ppi I
Bsp24 I	Psr I
BspG I	RleA I
BspM I	Sap I
BspNC I	SfaN I
Bsr I	Sim I

Zinc finger domain-cleavage domain fusions

[0141] Methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art. For example, methods for the design and construction of fusion protein comprising zinc finger proteins (and polynucleotides encoding same) are described in co-owned US Patents 6,453,242 and 6,534,261. In certain embodiments, polynucleotides encoding such fusion proteins are constructed. These polynucleotides can be inserted into a vector and the vector can be introduced into a cell (see below for additional disclosure regarding vectors and methods for introducing polynucleotides into cells).

[0142] In certain embodiments of the methods described herein, a fusion protein comprises a zinc finger binding domain and a cleavage half-domain from the *Fok* I restriction enzyme, and two such fusion proteins are expressed in a cell. Expression of two fusion proteins in a cell can result from delivery of the two proteins to the cell; delivery of one protein and one nucleic acid encoding one of the proteins to the cell; delivery of two nucleic acids, each encoding one of the proteins, to the cell; or by delivery of a single nucleic acid, encoding both proteins, to the cell. In additional embodiments, a fusion protein comprises a single polypeptide chain comprising two cleavage half domains and a zinc finger binding domain. In this case, a single fusion protein is expressed in a cell and, without wishing to be bound by theory, is believed to cleave DNA as a result of formation of an intramolecular dimer of the cleavage half-domains.

[0143] In certain embodiments, the components of the fusion proteins (*e.g.*, ZFP-*Fok* I fusions) are arranged such that the zinc finger domain is nearest the amino terminus of the fusion protein, and the cleavage half-domain is nearest the carboxy-terminus. This mirrors the relative orientation of the cleavage domain in naturally-occurring dimerizing cleavage domains such as those derived from the *Fok* I enzyme, in which the DNA-binding domain is nearest the amino terminus and the cleavage half-domain is nearest the carboxy terminus. In these embodiments, dimerization of the cleavage half-domains to form a functional nuclease is brought about by binding of the fusion proteins to sites on opposite DNA strands, with the 5' ends of the binding sites being proximal to each other. *See* Figure 43A.

[0144] In additional embodiments, the components of the fusion proteins (*e.g.*, ZFP-*Fok* I fusions) are arranged such that the cleavage half-domain is nearest the amino terminus of the fusion protein, and the zinc finger domain is nearest the carboxy-terminus. In these embodiments, dimerization of the cleavage half-domains to form a functional nuclease is brought about by binding of the fusion proteins to sites on opposite DNA strands, with the 3' ends of the binding sites being proximal to each other. *See* Figure 43B.

[0145] In yet additional embodiments, a first fusion protein contains the cleavage half-domain nearest the amino terminus of the fusion protein, and the zinc finger domain nearest the carboxy-terminus, and a second fusion protein is arranged such that the zinc finger domain is nearest the amino terminus of the fusion protein, and the cleavage half-domain is nearest the carboxy-terminus. In these embodiments, both fusion proteins bind to the same DNA strand, with the binding site of the first fusion protein containing the zinc finger domain nearest the carboxy terminus located to the 5' side of the binding site of the second fusion protein containing the zinc finger domain nearest the amino terminus. *See* Figure 43C.

[0146] In the disclosed fusion proteins, the amino acid sequence between the zinc finger domain and the cleavage domain (or cleavage half-domain) is denoted the "ZC linker." The ZC linker is to be distinguished from the inter-finger linkers discussed above. For the purposes of determining the length of a ZC linker, the zinc finger structure described by Pabo *et al.* (2001) *Ann. Rev. Biochem.* **70:**313-340 is used:

$$X\text{-}X\text{-}C\text{-}X_{2\text{-}4}\text{-}C\text{-}X_{12}\text{-}H\text{-}X_{3\text{-}5}\text{-}H \text{ (SEQ ID NO: 201)}$$

[0147] In this structure, the first residue of a zinc finger is the amino acid located two residues amino-terminal to the first conserved cysteine residue. In the majority of naturally-occurring zinc finger proteins, this position is occupied by a hydrophobic amino acid (usually either phenylalanine or tyrosine). In the disclosed fusion proteins, the first residue of a zinc finger will thus often be a hydrophobic residue, but it can be any amino acid. The final amino acid residue of a zinc finger, as shown above, is the second conserved histidine residue.

[0148] Thus, in the disclosed fusion proteins having a polarity in which the zinc finger binding domain is amino-terminal to the cleavage domain (or cleavage half-domain), the ZC linker is the amino acid sequence between the second conserved histidine residue of the C-terminal-most zinc finger and the N-terminal-most amino acid of the cleavage domain (or cleavage half-domain). For example, in certain fusion proteins whose construction is exemplified in the Examples section, the N-terminal-most amino acid of a cleavage half-domain is a glutamine (Q) residue corresponding to amino acid number 384 in the *Fok*I sequence of Looney *et al.* (1989) *Gene* **80:**193-208.

[0149] For fusion proteins having a polarity in which the cleavage domain (or cleavage half-domain) is amino-terminal to the zinc finger binding domain, the ZC linker is the amino acid sequence between the C-terminal-most amino acid residue of the cleavage domain (or half-domain) and the first residue of the N-terminal-most zinc finger of the zinc finger binding domain (*i.e.*, the residue located two residues upstream of the first conserved cysteine residue). In certain exemplary fusion proteins, the C-terminal-most amino acid of a cleavage half-domain is a phenylalanine (F) residue corresponding to amino acid number 579 in the *Fok*I sequence of Looney *et al.* (1989) *Gene* **80:**193-208.

[0150] The ZC linker can be any amino acid sequence. To obtain optimal cleavage, the length of the ZC linker and the distance between the target sites (binding sites) are interrelated. *See*, for example, Smith *et al.* (2000) *Nucleic Acids Res.* **28:**3361-3369; Bibikova *et al.* (2001) *Mol. Cell. Biol.* **21:**289-297, noting that their notation for linker length differs from that given here. For example, for ZFP-*Fok* I fusions in which the zinc finger binding domain is amino-terminal to the cleavage half-domain, and having a ZC linker length of four amino acids as defined herein (and denoted L0 by others), optimal cleavage occurs when the binding sites for the fusion proteins are located 6 or 16 nucleotides apart (as measured from the near edge of each binding site). *See* Example 4.

Methods for targeted cleavage

[0151] The disclosed methods and compositions can be used to cleave DNA 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). For such targeted DNA cleavage, a zinc finger binding domain is engineered to bind a target site at or near the predetermined cleavage site, and a fusion protein comprising the engineered zinc finger binding domain and a cleavage domain is expressed in a cell. Upon binding of the zinc finger portion of the fusion protein to the target site, the DNA is cleaved near the target site by the cleavage domain. The exact site of cleavage can depend on the length of the ZC linker.

[0152] Alternatively, two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain, are expressed in a cell, and bind to target sites which are juxtaposed in such a way that a functional cleavage domain is reconstituted and DNA is cleaved in the vicinity of the target sites. In one embodiment, cleavage occurs between the target sites of the two zinc finger binding domains. One or both of the zinc finger binding domains can be engineered.

[0153] For targeted cleavage using a zinc finger binding domain-cleavage domain fusion polypeptide, the binding site can encompass the cleavage site, or the near edge of the binding site can be 1, 2, 3, 4, 5, 6, 10, 25, 50 or more nucleotides (or any integral value between 1 and 50 nucleotides) from the cleavage site. The exact location of the binding site, with respect to the cleavage site, will depend upon the particular cleavage domain, and the length of the ZC linker. For methods in which two fusion polypeptides, each comprising a zinc finger binding domain and a cleavage half-domain, are used, the binding sites generally straddle the cleavage site. Thus the near edge of the first binding site can be 1, 2, 3, 4, 5, 6, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides) on one side of the cleavage site, and the near edge of the second binding site can be 1, 2, 3, 4, 5, 6, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides) on the other side of the cleavage site. Methods for mapping cleavage sites *in vitro* and *in vivo* are known to those of skill in the art.

[0154] Thus, the methods described herein can employ an engineered zinc finger binding domain fused to a cleavage domain. In these cases, the binding domain is engineered to bind to a target sequence, at or near which cleavage is desired. The fusion protein, or a polynucleotide encoding same, is introduced into a cell. Once introduced into, or expressed in, the cell, the fusion protein binds to the target sequence and cleaves at or near the target sequence. The exact site of cleavage depends on the nature of the cleavage domain and/or the presence and/or nature of linker sequences between the binding and cleavage domains. In cases where two fusion proteins, each comprising a cleavage half-domain, are used, the distance between the near edges of the binding sites can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 25 or more nucleotides (or any integral value between 1 and 50 nucleotides). Optimal levels of cleavage can also depend on both the distance between the binding sites of the two fusion proteins (*See*, for example, Smith *et al.* (2000) *Nucleic Acids Res.* **28**:3361-3369; Bibikova *et al.* (2001) *Mol. Cell. Biol.* **21**:289-297) and the length of the ZC linker in each fusion protein.

[0155] For ZFP-*Fok*I fusion nucleases, the length of the linker between the ZFP and the *Fok*I cleavage half-domain (*i.e.*, the ZC linker) can influence cleavage efficiency. In one experimental system utilizing a ZFP-*Fok*I fusion with a ZC linker of 4 amino acid residues, optimal cleavage was obtained when the near edges of the binding sites for two ZFP-*Fok*I nucleases were separated by 6 base pairs. This particular fusion nuclease comprised the following amino acid sequence between the zinc finger portion and the nuclease half-domain:

<u>H</u>QRT<u>H</u>QNKK<u>QLV</u> (SEQ ID NO:26)

in which the two conserved histidines in the C-terminal portion of the zinc finger and the first three residues in the *Fok*I cleavage half-domain are underlined. Accordingly, the ZC linker sequence in this construct is QNKK. Bibikova *et al.* (2001) *Mol. Cell. Biol.* **21**:289-297. The present inventors have constructed a number of ZFP-*Fok*I fusion nucleases having a variety of ZC linker lengths and sequences, and analyzed the cleavage efficiencies of these nucleases on a series of substrates having different distances between the ZFP binding sites. *See* Example 4.

[0156] In certain embodiments, the cleavage domain comprises two cleavage half-domains, both of which are part of a single polypeptide comprising a binding domain, a first cleavage half-domain and a second cleavage half-domain. The cleavage half-domains can have the same amino acid sequence or different amino acid sequences, so long as they function to cleave the DNA.

[0157] Cleavage half-domains may also be provided in separate molecules. For example, two fusion polypeptides may be introduced into a cell, wherein each polypeptide comprises a binding domain and a cleavage half-domain. The cleavage half-domains can have the same amino acid sequence or different amino acid sequences, so long as they function to cleave the DNA. Further, the binding domains bind to target sequences which are typically disposed in such a way that, upon binding of the fusion polypeptides, the two cleavage half-domains are presented in a spatial orientation to each other that allows reconstitution of a cleavage domain (*e.g.*, by dimerization of the half-domains), thereby positioning the half-domains relative to each other to form a functional cleavage domain, resulting in cleavage of cellular chromatin in a region of interest. Generally, cleavage by the reconstituted cleavage domain occurs at a site located between the two target sequences. One or both of the proteins can be engineered to bind to its target site.

[0158] The two fusion proteins can bind in the region of interest in the same or opposite polarity, and their binding sites (*i.e.*, target sites) can be separated by any number of nucleotides, *e.g.*, from 0 to 200 nucleotides or any integral value therebetween. In certain embodiments, the binding sites for two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain, can be located between 5 and 18 nucleotides apart, for example, 5-8 nucleotides apart, or 15-18 nucleotides apart, or 6 nucleotides apart, or 16 nucleotides apart, as measured from the edge of each binding site nearest the other binding site, and cleavage occurs between the binding sites.

[0159] The site at which the DNA is cleaved generally lies between the binding sites for the two fusion proteins. Double-strand breakage of DNA often results from two single-strand breaks, or "nicks," offset by 1, 2, 3, 4, 5, 6 or more nucleotides, (for example, cleavage of double-stranded DNA by native *Fok* I results from single-strand breaks offset by 4 nucleotides). Thus, cleavage does not necessarily occur at exactly opposite sites on each DNA strand. In addition, the structure of the fusion proteins and the distance between the target sites can influence whether cleavage occurs adjacent a single nucleotide pair, or whether cleavage occurs at several sites. However, for many applications, including targeted recombination and targeted mutagenesis (see *infra*) cleavage within a range of nucleotides is generally sufficient, and cleavage between particular base pairs is not required.

[0160] As noted above, the fusion protein(s) can be introduced as polypeptides and/or polynucleotides. For example, two polynucleotides, each comprising sequences encoding one of the aforementioned polypeptides, can be introduced into a cell, and when the polypeptides are expressed and each binds to its target sequence, cleavage occurs at or near the target sequence. Alternatively, a single polynucleotide comprising sequences encoding both fusion polypeptides is introduced into a cell. Polynucleotides can be DNA, RNA or any modified forms or analogues or DNA and/or RNA.

[0161] To enhance cleavage specificity, additional compositions may also be employed in the methods described herein. For example, single cleavage half-domains can exhibit limited double-stranded cleavage activity. In methods in which two fusion proteins, each containing a three-finger zinc finger domain and a cleavage half-domain, are introduced into the cell, either protein specifies an approximately 9-nucleotide target site. Although the aggregate target sequence of 18 nucleotides is likely to be unique in a mammalian genome, any given 9-nucleotide target site occurs, on average, approximately 23,000 times in the human genome. Thus, non-specific cleavage, due to the site-specific binding of a single half-domain, may occur. Accordingly, the methods described herein contemplate the use of a dominant-negative mutant of a cleavage half-domain such as *Fok* I (or a nucleic acid encoding same) that is expressed in a cell along with the two fusion proteins. The dominant-negative mutant is capable of dimerizing but is unable to cleave, and also blocks the cleavage activity of a half-domain to which it is dimerized. By providing the dominant-negative mutant in molar excess to the fusion proteins, only regions in which both fusion proteins are bound will have a high enough local concentration of functional cleavage half-domains for dimerization and cleavage to occur. At sites where only one of the two fusion proteins is bound, its cleavage half-domain forms a dimer with the dominant negative mutant half-domain, and undesirable, non-specific cleavage does not occur.

[0162] Three catalytic amino acid residues in the *Fok* I cleavage half-domain have been identified: Asp 450, Asp 467 and Lys 469. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95:** 10,570-10,575. Thus, one or more mutations at one of these residues can be used to generate a dominant negative mutation. Further, many of the catalytic amino acid residues of other Type IIS endonucleases are known and/or can be determined, for example, by alignment with *Fok* I sequences and/or by generation and testing of mutants for catalytic activity.

Dimerization domain mutations in the cleavage half-domain

[0163] Methods for targeted cleavage which involve the use of fusions between a ZFP and a cleavage half-domain (such as, *e.g.*, a ZFP/*Fok*I fusion) require the use of two such fusion molecules, each generally directed to a distinct target sequence. Target sequences for the two fusion proteins can be chosen so that targeted cleavage is directed to a unique site in a genome, as discussed above. A potential source of reduced cleavage specificity could result from homodimerization of one of the two ZFP/cleavage half-domain fusions. This might occur, for example, due to the presence, in a genome, of inverted repeats of the target sequences for one of the two ZFP/cleavage half-domain fusions, located so as to allow two copies of the same fusion protein to bind with an orientation and spacing that allows formation of a functional dimer.

[0164] One approach for reducing the probability of this type of aberrant cleavage at sequences other than the intended target site involves generating variants of the cleavage half-domain that minimize or prevent homodimerization. Preferably, one or more amino acids in the region of the half-domain involved in its dimerization are altered. In the crystal structure of the *Fok*I protein dimer, the structure of the cleavage half-domains is reported to be similar to the arrangement of the cleavage half-domains during cleavage of DNA by *Fok*I. Wah *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95:**10564-10569. This structure indicates that amino acid residues at positions 483 and 487 play a key role in the dimerization of the *Fok*I cleavage half-domains. The structure also indicates that amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 are all close enough to the dimerization interface to influence dimerization. Accordingly, amino acid sequence alterations at one or more of the aforementioned positions will likely alter the dimerization properties of the cleavage half-domain. Such changes can be introduced, for example, by constructing a library containing (or encoding) different amino acid residues at these positions and selecting variants with the desired properties, or by rationally designing individual mutants. In addition to preventing homodimerization, it is also possible that some of these mutations may increase the cleavage efficiency above that obtained with two wild-type cleavage half-domains.

[0165] Accordingly, alteration of a *Fok*I cleavage half-domain at any amino acid residue which affects dimerization can be used to prevent one of a pair of ZFP/*Fok*I fusions from undergoing homodimerization which can lead to cleavage at undesired sequences. Thus, for targeted cleavage using a pair of ZFP/*Fok*I fusions, one or both of the fusion proteins can comprise one or more amino acid alterations that inhibit self-dimerization, but allow heterodimerization of the two fusion proteins to occur such that cleavage occurs at the desired target site. In certain embodiments, alterations are present in both fusion proteins, and the alterations have additive effects; *i.e.*, homodimerization of either fusion, leading to aberrant cleavage, is minimized or abolished, while heterodimerization of the two fusion proteins is facilitated compared to that obtained with wild-type cleavage half-domains. *See* Example 5.

Methods for targeted alteration of genomic sequences and targeted recombination

[0166] Also described herein are methods of replacing a genomic sequence (*e.g.*, a region of interest in cellular chromatin) with a homologous non-identical sequence (*i.e.*, targeted recombination). Previous attempts to replace particular sequences have involved contacting a cell with a polynucleotide comprising sequences bearing homology to a chromosomal region (*i.e.*, a donor DNA), followed by selection of cells in which the donor DNA molecule had undergone homologous recombination into the genome. The success rate of these methods is low, due to poor efficiency of homologous recombination and a high frequency of non-specific insertion of the donor DNA into regions of the genome other than the target site.

[0167] The present disclosure provides methods of targeted sequence alteration characterized by a greater efficiency of targeted recombination and a lower frequency of non-specific insertion events. The methods involve making and using engineered zinc finger binding domains fused to cleavage domains (or cleavage half-domains) to make one or more targeted double-stranded breaks in cellular DNA. Because double-stranded breaks in cellular DNA stimulate cellular repair mechanisms several thousand-fold in the vicinity of the cleavage site, such targeted cleavage allows for the alteration or replacement (*via* homology-directed repair) of sequences at virtually any site in the genome.

[0168] In addition to the fusion molecules described herein, targeted replacement of a selected genomic sequence also requires the introduction of the replacement (or donor) sequence. The donor sequence can be introduced into the cell prior to, concurrently with, or subsequent to, expression of the fusion protein(s). The donor polynucleotide contains sufficient homology to a genomic sequence to support homologous recombination (or homology-directed repair) between it and the genomic sequence to which it bears homology. Approximately 25, 50 100, 200, 500, 750, 1,000, 1,500, 2,000 nucleotides or more of sequence homology between a donor and a genomic sequence (or any integral value between 10 and 2,000 nucleotides, or more) will support homologous recombination therebetween. Donor sequences can range in length from 10 to 5,000 nucleotides (or any integral value of nucleotides therebetween) or longer. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence that it replaces. For example, the sequence of the donor polynucleotide can contain one or more single base changes, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology with chromosomal sequences is present. Alternatively, a donor sequence can contain a non-homologous sequence flanked by two regions of homology. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous to the region of interest in cellular chromatin. Generally, the homologous region(s) of a donor sequence will have at least 50% sequence identity to a genomic sequence with which recombination is desired. In certain embodiments, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.9% sequence identity is present. Any value between 1% and 100% sequence identity can be present, depending upon the length of the donor polynucleotide.

[0169] A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. For example, 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.

[0170] To simplify assays (*e.g.*, hybridization, PCR, restriction enzyme digestion) for determining successful insertion of the donor sequence, certain sequence differences may be present in the donor sequence as compared to the genomic sequence. Preferably, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (*i.e.*, changes which do not affect the structure or function of the protein). The donor polynucleotide can optionally contain changes in sequences corresponding to the zinc finger domain binding sites in the region of interest, to prevent cleavage of donor sequences that have been introduced into cellular chromatin by homologous recombination.

[0171] The donor polynucleotide can be DNA or RNA, single-stranded or double-stranded and can be introduced into a cell in linear or circular 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 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. A polynucleotide can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor polynucleotides can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (*e.g.*, adenovirus, AAV, herpesvirus, retrovirus, lentivirus).

[0172] Without being bound by one theory, it appears that the presence of a double-stranded break in a cellular sequence, coupled with the presence of an exogenous DNA molecule having homology to a region adjacent to or surrounding the break, activates cellular mechanisms which repair the break by transfer of sequence information from the donor molecule into the cellular (*e.g.*, genomic or chromosomal) sequence; *i.e.*, by a processes of homology-directed repair, also known as "gene conversion." Applicants' methods advantageously combine the powerful targeting capabilities of engineered ZFPs with a cleavage domain (or cleavage half-domain) to specifically target a double-stranded break to the region of the genome at insertion of exogenous sequences is desired.

[0173] For alteration of a chromosomal sequence, it is not necessary for the entire sequence of the donor to be copied into the chromosome, as long as enough of the donor sequence is copied to effect the desired sequence alteration.

[0174] The efficiency of insertion of donor sequences by homologous recombination is inversely related to the distance, in the cellular DNA, between the double-stranded break and the site at which recombination is desired. In other words, higher homologous recombination efficiencies are observed when the double-stranded break is closer to the site at which recombination is desired. In cases in which a precise site of recombination is not predetermined (*e.g.*, the desired recombination event can occur over an interval of genomic sequence), the length and sequence of the donor nucleic acid, together with the site(s) of cleavage, are selected to obtain the desired recombination event. In cases in which the desired event is designed to change the sequence of a single nucleotide pair in a genomic sequence, cellular chromatin is cleaved within 10,000 nucleotides on either side of that nucleotide pair. In certain embodiments, cleavage occurs within 1,000, 500, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, or 2 nucleotides, or any integral value between 2 and 1,000 nucleotides, on either side of the nucleotide pair whose sequence is to be changed.

[0175] As detailed above, the binding sites for two fusion proteins, each comprising a zinc finger binding domain and a cleavage half-domain, can be located 5-8 or 15-18 nucleotides apart, as measured from the edge of each binding site nearest the other binding site, and cleavage occurs between the binding sites. Whether cleavage occurs at a single site or at multiple sites between the binding sites is immaterial, since the cleaved genomic sequences are replaced by the donor sequences. Thus, for efficient alteration of the sequence of a single nucleotide pair by targeted recombination, the midpoint of the region between the binding sites is within 10,000 nucleotides of that nucleotide pair, preferably within 1,000 nucleotides, or 500 nucleotides, or 200 nucleotides, or 100 nucleotides, or 50 nucleotides, or 20 nucleotides, or 10 nucleotides, or 5 nucleotide, or 2 nucleotides, or one nucleotide, or at the nucleotide pair of interest.

[0176] In certain embodiments, a homologous chromosome can serve as the donor polynucleotide. Thus, for example, correction of a mutation in a heterozygote can be achieved by engineering fusion proteins which bind to and cleave the mutant sequence on one chromosome, but do not cleave the wild-type sequence on the homologous chromosome. The double-stranded break on the mutation-bearing chromosome stimulates a homology-based "gene conversion" process in which the wild-type sequence from the homologous chromosome is copied into the cleaved chromosome, thus restoring two copies of the wild-type sequence.

[0177] Methods and compositions are also provided that may enhance levels of targeted recombination including, but not limited to, the use of additional ZFP-functional domain fusions to activate expression of genes involved in homologous recombination, such as, for example, members of the RAD52 epistasis group (*e.g.*, *Rad50, Rad51, Rad51B, Rad51C, Rad51D, Rad52, Rad54, Rad54B, Mre11, XRCC2, XRCC3*), genes whose products interact with the aforementioned gene products (*e.g.*, BRCA1, BRCA2) and/or genes in the NBS1 complex. Similarly ZFP-functional domain fusions can be used, in combination with the methods and compositions disclosed herein, to repress expression of genes involved in non-homologous end joining (*e.g.*, Ku70/80, XRCC4, poly(ADP ribose) polymerase, DNA ligase 4). *See*, for example, Yanez *et al.* (1998) *Gene Therapy* **5**:149-159; Hoeijmakers (2001) *Nature* **411**:366-374; Johnson *et al.* (2001) *Biochem. Soc. Trans.* **29**:196-201; Tauchi *et al.* (2002) *Oncogene* **21**:8967-8980. Methods for activation and repression of gene expression using fusions between a zinc finger binding domain and a functional domain are disclosed, for example, in co-owned US Patents 6,534,261; 6,824,978 and 6,933,113. Additional repression methods include the use of antisense oligonucleotides and/or small interfering RNA (siRNA or RNAi) targeted to the sequence of the gene to be repressed.

[0178] As an alternative to or, in addition to, activating expression of gene products involved in homologous recombination, fusions of these protein (or functional fragments thereof) with a zinc finger binding domain targeted to the region of interest, can be used to recruit these proteins (recombination proteins) to the region of interest, thereby increasing their local concentration and further stimulating homologous recombination processes. Alternatively, a polypeptide involved in homologous recombination as described above (or a functional fragment thereof) can be part of a triple fusion protein comprising a zinc finger binding domain, a cleavage domain (or cleavage half-domain) and the recombination protein (or functional fragment thereof). Additional proteins involved in gene conversion and recombination-related chromatin remodeling, which can be used in the aforementioned methods and compositions, include histone acetyltransferases (*e.g.*, Esa1p, Tip60), histone methyltransferases (*e.g.*, Dot1p), histone kinases and histone phosphatases.

[0179] The p53 protein has been reported to play a central role in repressing homologous recombination (HR). *See*, for example, Valerie *et al.*, (2003) *Oncogene* **22:**5792-5812; Janz, *et al.* (2002) *Oncogene* **21:**5929-5933. For example, the rate of HR in p53-deficient human tumor lines is 10,000-fold greater than in primary human fibroblasts, and there is a 100-fold increase in HR in tumor cells with a non-functional p53 compared to those with functional p53. Mekeel *et al.* (1997) *Oncogene* **14:**1847-1857. In addition, overexpression of p53 dominant negative mutants leads to a 20-fold increase in spontaneous recombination. Bertrand *et al.* (1997) *Oncogene* **14:**1117-1122. Analysis of different p53 mutations has revealed that the roles of p53 in transcriptional transactivation and G1 cell cycle checkpoint control are separable from its involvement in HR. Saintigny *et al.* (1999) *Oncogene* **18:**3553-3563; Boehden *et al.* (2003) *Oncogene* **22:**4111-4117. Accordingly, downregulation of p53 activity can serve to increase the efficiency of targeted homologous recombination using the methods and compositions disclosed herein. Any method for downregulation of p53 activity can be used, including but not limited to cotransfection and overexpression of a p53 dominant negative mutant or targeted repression of p53 gene expression according to methods disclosed, *e.g.*, in co-owned U.S. Patent No. 6,534,261.

[0180] Further increases in efficiency of targeted recombination, in cells comprising a zinc finger/nuclease fusion molecule and a donor DNA molecule, are achieved by blocking the cells in the G_2 phase of the cell cycle, when homology-driven repair processes are maximally active. Such arrest can be achieved in a number of ways. For example, cells can be treated with *e.g.*, drugs, compounds and/or small molecules which influence cell-cycle progression so as to arrest cells in G_2 phase. Exemplary molecules of this type include, but are not limited to, compounds which affect microtubule polymerization (*e.g.*, vinblastine, nocodazole, Taxol), compounds that interact with DNA (*e.g.*, *cis*-platinum(II) diamine dichloride, Cisplatin, doxorubicin) and/or compounds that affect DNA synthesis (*e.g.*, thymidine, hydroxyurea, L-mimosine, etoposide, 5-fluorouracil). Additional increases in recombination efficiency are achieved by the use of histone deacetylase (HDAC) inhibitors (*e.g.*, sodium butyrate, trichostatin A) which alter chromatin structure to make genomic DNA more accessible to the cellular recombination machinery.

[0181] Additional methods for cell-cycle arrest include overexpression of proteins which inhibit the activity of the CDK cell-cycle kinases, for example, by introducing a cDNA encoding the protein into the cell or by introducing into the cell an engineered ZFP which activates expression of the gene encoding the protein. Cell-cycle arrest is also achieved by inhibiting the activity of cyclins and CDKs, for example, using RNAi methods (*e.g.*, U.S. Patent No. 6,506,559) or by introducing into the cell an engineered ZFP which represses expression of one or more genes involved in cell-cycle progression such as, for example, cyclin and/or CDK genes. See, *e.g.*, co-owned U.S. Patent No. 6,534,261 for methods for the synthesis of engineered zinc finger proteins for regulation of gene expression.

[0182] Alternatively, in certain cases, targeted cleavage is conducted in the absence of a donor polynucleotide (preferably in S or G_2 phase), and recombination occurs between homologous chromosomes.

Methods to screen for cellular factors that facilitate homologous recombination

[0183] Since homologous recombination is a multi-step process requiring the modification of DNA ends and the recruitment of several cellular factors into a protein complex, the addition of one or more exogenous factors, along with donor DNA and vectors encoding zinc finger-cleavage domain fusions, can be used to facilitate targeted homologous recombination. An exemplary method for identifying such a factor or factors employs analyses of gene expression using microarrays (*e.g.*, Affymetrix Gene Chip® arrays) to compare the mRNA expression patterns of different cells. For example, cells that exhibit a higher capacity to stimulate double strand break-driven homologous recombination in the presence of donor DNA and zinc finger-cleavage domain fusions, either unaided or under conditions known to increase the level of gene correction, can be analyzed for their gene expression patterns compared to cells that lack such capacity. Genes that are upregulated or downregulated in a manner that directly correlates with increased levels of homologous recombination are thereby identified and can be cloned into any one of a number of expression vectors. These expression constructs can be co-transfected along with zinc finger-cleavage domain fusions and donor constructs to yield improved methods for achieving high-efficiency homologous recombination. Alternatively, expression of such genes can be appropriately regulated using engineered zinc finger proteins which modulate expression (either activation or repression) of one or more these genes. See, *e.g.*, co- owned U.S. Patent No. 6,534,261 for methods for the synthesis of engineered zinc finger proteins for regulation of gene expression.

[0184] As an example, it was observed that the different clones obtained in the experiments described in Example 9 and Figure 27 exhibited a wide-range of homologous recombination frequencies, when transfected with donor DNA and plasmids encoding zinc finger-cleavage domain fusions. Gene expression in clones showing a high frequency of targeted recombination can thus be compared to that in clones exhibiting a low frequency, and expression patterns unique to the former clones can be identified.

[0185] As an additional example, studies using cell cycle inhibitors (*e.g.*, nocodazole or vinblastine, see *e.g.*, Examples 11, 14 and 15) showed that cells arrested in the G2 phase of the cell cycle carried out homologous recombination at higher rates, indicating that cellular factors responsible for homologous recombination may be preferentially expressed or active in G2. One way to identify these factors is to compare the mRNA expression patterns between the stably transfected HEK 293 cell clones that carry out gene correction at high and low levels (*e.g.*, clone T18 *vs.* clone T7). Similar comparisons are made between these cell lines in response to compounds that arrest the cells in G2 phase. Candidate genes that are differentially expressed in cells that carry out homologous recombination at a higher rate, either unaided or in response to compounds that arrest the cells in G2, are identified, cloned, and re-introduced into cells to determine whether their expression is sufficient to re-capitulate the improved rates. Alternatively, expression of said candidate genes is activated using engineered zinc finger transcription factors as described, for example, in co-owned U.S. Patent No. 6,534,261.

Expression vectors

[0186] A nucleic acid encoding one or more ZFPs or ZFP fusion proteins can be cloned into a vector for transformation into prokaryotic or eukaryotic cells for replication and/or expression. Vectors can be prokaryotic vectors, e.g., plasmids, or shuttle vectors, insect vectors, or eukaryotic vectors. A nucleic acid encoding a ZFP can also be cloned into an expression vector, for administration to a plant cell, animal cell, preferably a mammalian cell or a human cell, fungal cell, bacterial cell, or protozoal cell.

[0187] To obtain expression of a cloned gene or nucleic acid, sequences encoding a ZFP or ZFP fusion protein are typically subcloned into an expression vector that contains a promoter to direct transcription. Suitable bacterial and eukaryotic promoters are well known in the art and described, e.g., in Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989; 3rd ed., 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, *supra.* Bacterial expression systems for expressing the ZFP are available in, e.g., *E. coli*, *Bacillus sp.*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983)). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known by those of skill in the art and are also commercially available.

[0188] The promoter used to direct expression of a ZFP-encoding nucleic acid depends on the particular application. For example, a strong constitutive promoter is typically used for expression and purification of ZFP. In contrast, when a ZFP is administered *in vivo* for gene regulation, either a constitutive or an inducible promoter is used, depending on the particular use of the ZFP. In addition, a preferred promoter for administration of a ZFP can be a weak promoter, such as HSV TK or a promoter having similar activity. The promoter typically can also include elements that are responsive to transactivation, *e.g.*, hypoxia response elements, Gal4 response elements, lac repressor response element, and small molecule control systems such as tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *PNAS* 89:5547 (1992); Oligino *et al., Gene Ther.* 5:491-496 (1998); Wang *et al., Gene Ther.* 4:432-441 (1997); Neering *et al., Blood* 88:1147-1155 (1996); and Rendahl *et al., Nat. Biotechnol.* 16:757-761 (1998)). The MNDU3 promoter can also be used, and is preferentially active in CD34$^+$ hematopoietic stem cells.

[0189] In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the nucleic acid in host cells, either prokaryotic or eukaryotic. A typical expression cassette thus contains a promoter operably linked, *e.g.*, to a nucleic acid sequence encoding the ZFP, and signals required, *e.g.*, for efficient polyadenylation of the transcript, transcriptional termination, ribosome binding sites, or translation termination. Additional elements of the cassette may include, *e.g.*, enhancers, and heterologous splicing signals.

[0190] The particular expression vector used to transport the genetic information into the cell is selected with regard to the intended use of the ZFP, *e.g.*, expression in plants, animals, bacteria, fungus, protozoa, *etc.* (see expression vectors described below). Standard bacterial expression vectors include plasmids such as pBR322-based plasmids, pSKF, pET23D, and commercially available fusion expression systems such as GST and LacZ. An exemplary fusion protein is the maltose binding protein, "MBP." Such fusion proteins are used for purification of the ZFP. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, for monitoring expression, and for monitoring cellular and subcellular localization, e.g., c-myc or FLAG.

[0191] Expression vectors containing regulatory elements from eukaryotic viruses are often used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 late promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

[0192] Some expression systems have markers for selection of stably transfected cell lines such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. High yield expression systems are also suitable, such as using a baculovirus vector in insect cells, with a ZFP encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

[0193] The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of recombinant sequences.

[0194] Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of protein, which are then purified using standard techniques (*see, e.g.,* Colley *et al.*, *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification, in Methods in Enzymology,* vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (*see, e.g.,* Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds., 1983)..

[0195] Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, ultrasonic methods (*e.g.*, sonoporation), liposomes, microinjection, naked DNA, plasmid vectors, viral vectors, both episomal and integrative, and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.,* Sambrook *et al., supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the protein of choice.

Nucleic acids encoding fusion proteins and delivery to cells

[0196] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding engineered ZFPs in cells (*e.g.*, mammalian cells) and target tissues. Such methods can also be used to administer nucleic acids encoding ZFPs to cells *in vitro*. In certain embodiments, nucleic acids encoding ZFPs are administered for *in vivo* or *ex vivo* gene therapy uses. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds.) (1995); and Yu *et al., Gene Therapy* 1:13-26 (1994).

[0197] Methods of non-viral delivery of nucleic acids encoding engineered ZFPs include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked

DNA, artificial virions, and agent-enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

[0198] Additional exemplary nucleic acid delivery systems include those provided by Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland) and BTX Molecular Delivery Systems (Holliston, MA).

[0199] Lipofection is described in e.g., US 5,049,386, US 4,946,787; and US 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and 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. Delivery can be to cells (*ex vivo* administration) or target tissues (*in vivo* administration).

[0200] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (*see, e.g.*, Crystal, *Science* 270:404-410 (1995); Blaese *et al., Cancer Gene Ther.* 2:291-297 (1995); Behr *et al., Bioconjugate Chem.* 5:382-389 (1994); Remy *et al., Bioconjugate Chem.* 5:647-654 (1994); Gao *et al., Gene Therapy* 2:710-722 (1995); Ahmad *et al., Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0201] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered ZFPs 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 patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of ZFPs 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.

[0202] 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 (*see, e.g.,* Buchscher *et al., J. Virol.* 66:2731-2739 (1992); Johann *et al., J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al., Virol.* 176:58-59 (1990); Wilson *et al., J. Virol.* 63:2374-2378 (1989); Miller *et al., J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0203] In applications in which transient expression of a ZFP fusion protein 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).

[0204] 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.

[0205] 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).

[0206] 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 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 Ther.* 9:748-55 (1996)).

[0207] 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 nondividing, 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 antitumor 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).

[0208] 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 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.

[0209] 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.

[0210] Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (e.g., intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (e.g., lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0211] *Ex vivo* cell transfection for diagnostics, research, or for gene therapy (e.g., via re-infusion of the transfected cells into the host organism) is well known to those of skill in the art. In a preferred embodiment, cells are isolated from the subject organism, transfected with a ZFP nucleic acid (gene or cDNA), and re-infused back into the subject organism (*e.g.*, patient). Various cell types suitable for *ex vivo* transfection are well known to those of skill in the art (*see, e.g.*, Freshney *et al.*, *Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994)) and the references cited therein for a discussion of how to isolate and culture cells from patients).

[0212] In one embodiment, stem cells are used in *ex vivo* procedures for cell transfection and gene therapy. The advantage to using stem cells is that they can be differentiated into other cell types *in vitro*, or can be introduced into a mammal (such as the donor of the cells) where they will engraft in the bone marrow. Methods for differentiating CD34+ cells *in vitro* into clinically important immune cell types using cytokines such a GM-CSF, IFN-γ and TNF-α are known (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

[0213] Stem cells are isolated for transduction and differentiation using known methods. For example, stem cells are isolated from bone marrow cells by panning the bone marrow cells with antibodies which bind unwanted cells, such as CD4+ and CD8+ (T cells), CD45+ (panB cells), GR-1 (granulocytes), and Iad (differentiated antigen presenting cells) (*see* Inaba *et al., J. Exp. Med.* 176:1693-1702 (1992)).

[0214] Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, etc.) containing therapeutic ZFP nucleic acids can also be administered directly to an organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0215] Methods for introduction of DNA into hematopoietic stem cells are disclosed, for example, in U.S. Patent No. 5,928,638. Vectors useful for introduction of transgenes into hematopoietic stem cells, *e.g.*, CD34$^+$ cells, include adenovirus Type 35.

[0216] Vectors suitable for introduction of transgenes into immune cells (*e.g.*, T-cells) include non-integrating lentivirus vectors. *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.

[0217] 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).

[0218] DNA constructs may be introduced into the genome of a desired plant host by a variety of conventional techniques. For reviews of such techniques see, for example, Weissbach & Weissbach *Methods for Plant Molecular Biology* (1988, Academic Press, N.Y.) Section VIII, pp. 421-463; and Grierson & Corey, *Plant Molecular Biology* (1988, 2d Ed.), Blackie, London, Ch. 7-9. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see, e.g., Klein et al (1987) *Nature* 327:70-73). Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al (1984) *Science* 233:496-498, and Fraley et al (1983) *Proc. Nat'l. Acad. Sci. USA* 80:4803. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria using binary T DNA vector (Bevan (1984) *Nuc. Acid Res.* 12:8711-8721) or the co-cultivation procedure (Horsch et al (1985) *Science* 227:1229-1231). Generally, the *Agrobacterium* transformation system is used to engineer dicotyledonous plants (Bevan et al (1982) *Ann. Rev. Genet* 16:357-384; Rogers et al (1986) *Methods Enzymol.* 118:627-641). The *Agrobacterium* transformation system may also be used to transform, as well as transfer, DNA to monocotyledonous plants and plant cells. *See* Hernalsteen et al (1984) *EMBO J* 3:3039-3041; Hooykass-Van Slogteren et al (1984) *Nature* 311:763-764; Grimsley et al (1987) *Nature* 325:1677-179; Boulton et al (1989) *Plant Mol. Biol.* 12:31-40.; and Gould et al (1991) *Plant Physiol.* 95:426-434.

[0219] Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-, polyethylene glycol (PEG)- or electroporation-mediated uptake of naked DNA (see Paszkowski et al. (1984) *EMBO J* 3:2717-2722, Potrykus et al. (1985) *Molec. Gen. Genet.* 199:169-177; Fromm et al. (1985) *Proc. Nat. Acad. Sci.* USA 82:5824-5828; and Shimamoto (1989) *Nature* 338:274-276) and electroporation of plant tissues (D'Halluin et al. (1992) *Plant Cell* 4:1495-1505). Additional methods for plant cell transformation include microinjection, silicon carbide mediated DNA uptake (Kaeppler et al. (1990) *Plant Cell Reporter* 9:415-418), and microprojectile bombardment (see Klein et al. (1988) *Proc. Nat. Acad. Sci. USA* 85:4305-4309; and Gordon-Kamm et al. (1990) *Plant Cell* 2:603-618).

[0220] The disclosed methods and compositions can be used to insert exogenous sequences into a predetermined location in a plant cell genome. This is useful inasmuch as expression of an introduced transgene into a plant genome depends critically on its integration site. Accordingly, genes encoding, *e.g.*, nutrients, antibiotics or therapeutic molecules can be inserted, by targeted recombination, into regions of a plant genome favorable to their expression.

[0221] Transformed plant cells which are produced by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, et al., "Protoplasts Isolation and Culture" in *Handbook of Plant Cell Culture,* pp. 124-176, Macmillian Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, pollens, embryos or parts thereof. Such regeneration techniques are described generally in Klee et al (1987) *Ann. Rev. of Plant Phys.* 38:467-486.

[0222] Nucleic acids introduced into a plant cell can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., *Arabidopsis*). Thus, the disclosed methods and compositions have use over a broad range of plants, including, but not limited to, species from the genera Asparagus, Avena, Brassica, Citrus, Citrullus, Capsicum, Cucurbita, Daucus, Glycine, Hordeum, Lactuca, Lycopersicon, Malus, Manihot, Nicotiana, Oryza, Persea, Pisum, Pyrus, Prunus, Raphanus, Secale, Solanum, Sorghum, Triticum, Vitis, Vigna, and Zea.

[0223] One of skill in the art will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0224] A transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for traits encoded by the marker genes present on the transforming DNA. For instance, selection may be performed by growing the engineered plant material on media containing an inhibitory amount of the antibiotic or herbicide to which the transforming gene construct confers resistance. Further, transformed plants and plant cells may also be identified by screening for the activities of any visible marker genes (e.g., the β-glucuronidase, luciferase, B or C1 genes) that may be present on the recombinant nucleic acid constructs. Such selection and screening methodologies are well known to those skilled in the art.

[0225] Physical and biochemical methods also may be used to identify plant or plant cell transformants containing inserted gene constructs. These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the recombinant DNA insert; 2) Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, Western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the gene construct products are proteins. Additional techniques, such as *in situ* hybridization, enzyme staining, and immunostaining, also may be used to detect the presence or expression of the recombinant construct in specific plant organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

[0226] Effects of gene manipulation using the methods disclosed herein can be observed by, for example, northern blots of the RNA (*e.g.,* mRNA) isolated from the tissues of interest. Typically, if the amount of mRNA has increased, it can be assumed that the corresponding endogenous gene is being expressed at a greater rate than before. Other methods of measuring gene and/or CYP74B activity can be used. Different types of enzymatic assays can be used, depending on the substrate used and the method of detecting the increase or decrease of a reaction product or by-product. In addition, the levels of and/or CYP74B protein expressed can be measured immunochemically, i.e., ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art, such as by electrophoretic detection assays (either with staining or western blotting). The transgene may be selectively expressed in some tissues of the plant or at some developmental stages, or the transgene may be expressed in substantially all plant tissues, substantially along its entire life cycle. However, any combinatorial expression mode is also applicable.

[0227] The present disclosure also encompasses seeds of the transgenic plants described above wherein the seed has the transgene or gene construct. The present disclosure further encompasses the progeny, clones, cell lines or cells of the transgenic plants described above wherein said progeny, clone, cell line or cell has the transgene or gene construct.

Delivery vehicles

[0228] An important factor in the administration of polypeptide compounds, such as ZFP fusion proteins, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as ZFPs across a cell membrane.

[0229] For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (*see, e.g.*, Prochiantz, *Current Opinion in Neurobiology* 6:629-634 (1996)). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (*see, e.g.*, Lin *et al., J. Biol. Chem.* 270:1 4255-14258 (1995)).

[0230] Examples of peptide sequences which can be linked to a protein, for facilitating uptake of the protein into cells, include, but are not limited to: an 11 amino acid peptide of the tat protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (*see* Fahraeus *et al., Current Biology* 6:84 (1996)); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al., J. Biol. Chem.* 269:10444 (1994)); the h region of a signal peptide such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin *et al., supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, *Cell* 88:223-233 (1997)). Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to ZFPs. Membrane translocation domains (*i.e.*, internalization domains) can also be selected from libraries of randomized peptide sequences. *See*, for example, Yeh *et al.* (2003) *Molecular Therapy* **7(5):**S461, Abstract #1191.

[0231] Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules (called "binary toxins") are composed of at least two parts: a translocation/binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used to deliver peptides to the cell cytosol as internal or amino-terminal fusions (Arora *et al., J. Biol. Chem.*, 268:3334-3341 (1993); Perelle *et al., Infect. Immun.*, 61:5147-5156 (1993); Stenmark *et al., J.*

Cell Biol. 113:1025-1032 (1991); Donnelly *et al.*, *PNAS* 90:3530-3534 (1993); Carbonetti *et al.*, *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295 (1995); Sebo *et al.*, *Infect. Immun.* 63:3851-3857 (1995); Klimpel *et al.*, *PNAS U.S.A.* 89:10277-10281 (1992); and Novak *et al.*, *J. Biol. Chem.* 267:17186-17193 1992)).

[0232] Such peptide sequences can be used to translocate ZFPs across a cell membrane. ZFPs can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the ZFP and the translocation sequence. Any suitable linker can be used, e.g., a peptide linker.

[0233] The ZFP can also be introduced into an animal cell, preferably a mammalian cell, via a liposomes and liposome derivatives such as immunoliposomes. The term "liposome" refers to vesicles comprised of one or more concentrically ordered lipid bilayers, which encapsulate an aqueous phase. The aqueous phase typically contains the compound to be delivered to the cell, i.e., a ZFP.

[0234] The liposome fuses with the plasma membrane, thereby releasing the drug into the cytosol. Alternatively, the liposome is phagocytosed or taken up by the cell in a transport vesicle. Once in the endosome or phagosome, the liposome either degrades or fuses with the membrane of the transport vesicle and releases its contents.

[0235] In current methods of drug delivery via liposomes, the liposome ultimately becomes permeable and releases the encapsulated compound (in this case, a ZFP) at the target tissue or cell. For systemic or tissue specific delivery, this can be accomplished, for example, in a passive manner wherein the liposome bilayer degrades over time through the action of various agents in the body. Alternatively, active drug release involves using an agent to induce a permeability change in the liposome vesicle. Liposome membranes can be constructed so that they become destabilized when the environment becomes acidic near the liposome membrane (*see, e.g.*, *PNAS* 84:7851 (1987); *Biochemistry* 28:908 (1989)). When liposomes are endocytosed by a target cell, for example, they become destabilized and release their contents. This destabilization is termed fusogenesis. Dioleoylphosphatidylethanolamine (DOPE) is the basis of many "fusogenic" systems.

[0236] Such liposomes typically comprise a ZFP and a lipid component, e.g., a neutral and/or cationic lipid, optionally including a receptor-recognition molecule such as an antibody that binds to a predetermined cell surface receptor or ligand (*e.g.*, an antigen). A variety of methods are available for preparing liposomes as described in, *e.g.*, Szoka *et al.*, *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, 4,946,787, PCT Publication No. WO 91\17424, Deamer & Bangham, *Biochim. Biophys. Acta* 443:629-634 (1976); Fraley, *et al.*, *PNAS* 76:3348-3352 (1979); Hope *et al.*, *Biochim. Biophys. Acta* 812:55-65 (1985); Mayer *et al.*, *Biochim. Biophys. Acta* 858:161-168 (1986); Williams *et al.*, *PNAS* 85:242-246 (1988); *Liposomes* (Ostro (ed.), 1983, Chapter 1); Hope *et al.*, *Chem. Phys. Lip.* 40:89 (1986); Gregoriadis, *Liposome Technology* (1984) and Lasic, *Liposomes: from Physics to Applications* (1993)). Suitable methods include, for example, sonication, extrusion, high pressure/homogenization, microfluidization, detergent dialysis, calcium-induced fusion of small liposome vesicles and ether-fusion methods, all of which are known to those of skill in the art.

[0237] In certain embodiments, it is desirable to target liposomes using targeting moieties that are specific to a particular cell type, tissue, and the like. Targeting of liposomes using a variety of targeting moieties (*e.g.*, ligands, receptors, and monoclonal antibodies) has been described. *See, e.g.*, U.S. Patent Nos. 4,957,773 and 4,603,044.

[0238] Examples of targeting moieties include monoclonal antibodies specific to antigens associated with neoplasms, such as prostate cancer specific antigen and MAGE. Tumors can also be diagnosed by detecting gene products resulting from the activation or over-expression of oncogenes, such as ras or c-erbB2. In addition, many tumors express antigens normally expressed by fetal tissue, such as the alphafetoprotein (AFP) and carcinoembryonic antigen (CEA). Sites of viral infection can be diagnosed using various viral antigens such as hepatitis B core and surface antigens (HBVc, HBVs) hepatitis C antigens, Epstein-Barr virus antigens, human immunodeficiency type-1 virus (HIV1) and papilloma virus antigens. Inflammation can be detected using molecules specifically recognized by surface molecules which are expressed at sites of inflammation such as integrins (*e.g.*, VCAM-1), selectin receptors (*e.g.*, ELAM-1) and the like.

[0239] Standard methods for coupling targeting agents to liposomes can be used. These methods generally involve incorporation into liposomes of lipid components, *e.g.*, phosphatidylethanolamine, which can be activated for attachment of targeting agents, or derivatized lipophilic compounds, such as lipid derivatized bleomycin. Antibody targeted liposomes can be constructed using, for instance, liposomes which incorporate protein A (*see* Renneisen *et al.*, *J. Biol. Chem.*, 265:16337-16342 (1990) and Leonetti *et al.*, *PNAS* 87:2448-2451 (1990).

Dosages

[0240] For therapeutic applications, the dose administered to a patient, or to a cell which will be introduced into a patient, in the context of the present disclosure, should be sufficient to effect a beneficial therapeutic response in the patient over time. In addition, particular dosage regimens can be useful for determining phenotypic changes in an experimental setting, *e.g.*, in functional genomics studies, and in cell or animal models. The dose will be determined by the efficacy and K_d of the particular ZFP employed, the nuclear volume of the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular patient.

[0241] The maximum therapeutically effective dosage of ZFP for approximately 99% binding to target sites is calculated to be in the range of less than about $1.5\text{x}10^5$ to $1.5\text{x}10^6$ copies of the specific ZFP molecule per cell. The number of ZFPs per cell for this level of binding is calculated as follows, using the volume of a HeLa cell nucleus (approximately 1000 μm^3 or 10^{-12} L; *Cell Biology*, (Altman & Katz, eds. (1976)). As the HeLa nucleus is relatively large, this dosage number is recalculated as needed using the volume of the target cell nucleus. This calculation also does not take into account competition for ZFP binding by other sites. This calculation also assumes that essentially all of the ZFP is localized to the nucleus. A value of 100x K_d is used to calculate approximately 99% binding of to the target site, and a value of 10x K_d is used to calculate approximately 90% binding of to the target site. For this example, K_d = 25 nM

ZFP + target site ↔ complex

i.e., DNA + protein ↔ DNA:protein complex

$$K_d = \frac{[\text{DNA}]\,[\text{protein}]}{[\text{DNA:protein complex}]}$$

When 50% of ZFP is bound, K_d = [protein]

So when [protein] = 25 nM and the nucleus volume is 10^{-12} L [protein] = ($25x10^{-9}$ moles/L) (10^{-12} L/nucleus) ($6x10^{23}$ molecules/mole)

= 15,000 molecules/nucleus for 50% binding

When 99% target is bound; 100x K_d = [protein]

100x K_d = [protein] = 2.5 μM

($2.5x10^{-6}$ moles/L) (10^{-12}L/nucleus) ($6x10^{23}$ molecules/mole)

= about 1,500,000 molecules per nucleus for 99% binding of target site.

[0242] The appropriate dose of an expression vector encoding a ZFP can also be calculated by taking into account the average rate of ZFP expression from the promoter and the average rate of ZFP degradation in the cell. In certain embodiments, a weak promoter such as a wild-type or mutant HSV TK promoter is used, as described above. The dose of ZFP in micrograms is calculated by taking into account the molecular weight of the particular ZFP being employed.

[0243] In determining the effective amount of the ZFP to be administered in the treatment or prophylaxis of disease, the physician evaluates circulating plasma levels of the ZFP or nucleic acid encoding the ZFP, potential ZFP toxicities, progression of the disease, and the production of anti-ZFP antibodies. Administration can be accomplished via single or divided doses.

Pharmaceutical compositions and administration

[0244] ZFPs and expression vectors encoding ZFPs can be administered directly to the patient for targeted cleavage and/or recombination, and for therapeutic or prophylactic applications, for example, cancer, ischemia, diabetic retinopathy, macular degeneration, rheumatoid arthritis, psoriasis, HIV infection, sickle cell anemia, Alzheimer's disease, muscular dystrophy, neurodegenerative diseases, vascular disease, cystic fibrosis, stroke, and the like. Examples of microorganisms that can be inhibited by ZFP gene therapy include pathogenic bacteria, e.g., chlamydia, rickettsial bacteria, mycobacteria, staphylococci, streptococci, pneumococci, meningococci and conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis, and Lyme disease bacteria; infectious fungus, e.g., *Aspergillus*, *Candida* species; protozoa such as sporozoa (e.g., *Plasmodia*), rhizopods (e.g., *Entamoeba*) and flagellates (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.);viral diseases, e.g., hepatitis (A, B, or C), herpes virus (*e.g.*, VZV, HSV-1, HSV-6, HSV-II, CMV, and EBV), HIV, Ebola, adenovirus, influenza virus, flaviviruses, echovirus, rhinovirus, coxsackie virus, coronavirus, respiratory syncytial virus, mumps virus, rotavirus, measles virus, rubella virus, parvovirus, vaccinia virus, HTLV virus, dengue virus, papillomavirus, poliovirus, rabies virus, and arboviral encephalitis virus, *etc.*

[0245] Administration of therapeutically effective amounts is by any of the routes normally used for introducing ZFP into ultimate contact with the tissue to be treated. The ZFPs are administered in any suitable manner, preferably with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0246] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions that are available (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

[0247] The ZFPs, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[0248] Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The disclosed compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

Applications

[0249] The disclosed methods and compositions for targeted cleavage can be used to induce mutations in a genomic sequence, *e.g.*, by cleaving at two sites and deleting sequences in between, by cleavage at a single site followed by non-homologous end joining, cleaving at one or two sites with insertion of an exogenous sequence between the breaks and/or by cleaving at a site so as to remove one or two or a few nucleotides. Targeted cleavage can also be used to create gene knock-outs (*e.g.*, for functional genomics or target validation) and to facilitate targeted insertion of a sequence into a genome (*i.e.*, gene knock-in); *e.g.*, for purposes of cell engineering or protein overexpression. Insertion can be by means of replacements of chromosomal sequences through homologous recombination or by targeted integration, in which a new sequence (*i.e.*, a sequence not present in the region of interest), flanked by sequences homologous to the region of interest in the chromosome, is inserted at a predetermined target site.

[0250] The same methods can also be used to replace a wild-type sequence with a mutant sequence, or to convert one allele to a different allele.

[0251] 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. 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;

Rhabodoviridae (*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.

[0252] In similar fashion, the genome of an infecting bacterium can be mutagenized by targeted DNA cleavage followed by non-homologous end joining, to block or ameliorate bacterial infections.

[0253] The disclosed methods for targeted recombination can be used to replace any genomic sequence with a homologous, non-identical sequence. For example, a mutant genomic sequence can be replaced by its wild-type counterpart, thereby providing 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.

[0254] Exemplary genetic diseases include, but are not limited to, achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia 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 6^{th} 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, 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).

[0255] 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.

[0256] In certain cases, alteration of a genomic sequence in a pluripotent cell (*e.g.*, a hematopoietic stem cell) is desired. Methods for mobilization, enrichment and culture of hematopoietic stem cells are known in the art. *See* for example, U.S. Patents 5,061,620; 5,681,559; 6,335,195; 6,645,489 and 6,667,064. Treated stem cells can be returned to a patient for treatment of various diseases including, but not limited to, SCID and sickle-cell anemia.

[0257] In many of these cases, a region of interest comprises a mutation, and the donor polynucleotide comprises the corresponding wild-type sequence. Similarly, a wild-type genomic sequence can be replaced by a mutant sequence, if such is desirable. For example, overexpression of an oncogene can be reversed either by mutating the gene or by replacing its control sequences with sequences that support a lower, non-pathologic level of expression. As another example, the wild-type allele of the ApoAI gene can be replaced by the ApoAI Milano allele, to treat atherosclerosis. Indeed, any pathology dependent upon a particular genomic sequence, in any fashion, can be corrected or alleviated using the methods and compositions disclosed herein.

[0258] Targeted cleavage and targeted recombination can also be used to alter non-coding sequences (*e.g.*, regulatory sequences such as promoters, enhancers, initiators, terminators, splice sites) to alter the levels of expression of a gene product. Such methods can be used, for example, for therapeutic purposes, functional genomics and/or target validation studies.

[0259] The compositions and methods described herein also allow for novel approaches and systems to address immune reactions of a host to allogeneic grafts. In particular, a major problem faced when allogeneic stem cells (or any type of allogeneic cell) are grafted into a host recipient is the high risk of rejection by the host's immune system, primarily mediated through recognition of the Major Histocompatibility Complex (MHC) on the surface of the engrafted cells. The MHC comprises the HLA class I protein(s) that function as heterodimers that are comprised of a common β subunit and variable α subunits. It has been demonstrated that tissue grafts derived from stem cells that are devoid of HLA escape the host's immune response. *See, e.g.*, Coffman et al. *J Immunol* **151**, 425-35. (1993); Markmann et al. *Transplantation* **54**, 1085-9. (1992); Koller et al. *Science* **248**, 1227-30. (1990). Using the compositions and methods described herein, genes encoding HLA proteins involved in graft rejection can be cleaved, mutagenized or altered by recombination, in either their coding or regulatory sequences, so that their expression is blocked or they express a non-functional product. For example, by inactivating the gene encoding the common β subunit gene (β2 microglobulin) using ZFP fusion proteins as described herein, HLA class I can be removed from the cells to rapidly and reliably generate HLA class I *null* stem cells from any donor, thereby reducing the need for closely matched donor/recipient MHC haplotypes during stem cell grafting.

[0260] Inactivation of any gene (*e.g.*, the β2 microglobulin gene) 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, or by targeted recombination of an irrelevant sequence (*i.e.*, a "stuffer" sequence) into the gene or its regulatory region, so as to disrupt the gene or regulatory region.

[0261] Targeted modification of chromatin structure, as disclosed in co-owned WO 01/83793, can be used to facilitate the binding of fusion proteins to cellular chromatin.

[0262] In additional embodiments, one or more fusions between a zinc finger binding domain and a recombinase (or functional fragment thereof) can be used, in addition to or instead of the zinc finger-cleavage domain fusions disclosed herein, to facilitate targeted recombination. See, for example, co-owned US patent No. 6,534,261 and Akopian *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100:**8688-8691.

[0263] In additional embodiments, the disclosed methods and compositions are used to provide fusions of ZFP binding domains with transcriptional activation or repression domains that require dimerization (either homodimerization or heterodimerization) for their activity. In these cases, a fusion polypeptide comprises a zinc finger binding domain and a functional domain monomer (*e.g.*, a monomer from a dimeric transcriptional activation or repression domain). Binding of two such fusion polypeptides to properly situated target sites allows dimerization so as to reconstitute a functional transcription activation or repression domain.

Targeted Integration

[0264] As disclosed above, the methods and compositions set forth herein can be used for targeted integration of exogenous sequences into a region of interest in the genome of a cell. Targeted integration of an exogenous sequence at a double-strand break in a genome can occur by both homology-dependent and homology-independent mechanisms.

[0265] As noted above, in certain embodiments, targeted integration by both homology-dependent and homology-independent mechanisms involves insertion of an exogenous sequence between the ends generated by cleavage. The exogenous sequence inserted can be any length, for example, a relatively short "patch" sequence of between 1 and 50 nucleotides in length (*e.g.*, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45 or 50 nucleotide sequence).

[0266] In cases in which targeted integration is homology-dependent, a donor nucleic acid or donor sequence comprises an exogenous sequence together with one or more sequences that are either identical, or homologous but non-identical, with a predetermined genomic sequence (*i.e.*, a target site). In certain embodiments two of the identical sequences or two of the homologous but non-identical sequences (or one of each) are present, flanking the exogenous sequence. An exogenous sequence (or exogenous nucleic acid or exogenous polynucleotide) is one that contains a nucleotide sequence that is not normally present in the region of interest.

[0267] Exemplary exogenous sequences include, but are not limited to, 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.

[0268] Protein expression constructs include, but are not limited to, cDNAs and transcriptional control sequences in operative linkage with cDNA sequences. Transcriptional control sequences include promoters, enhancers and insulators. Additional transcriptional and translational regulatory sequences which can be included in expression constructs include, *e.g.*, internal ribosome entry sites, sequences encoding 2A peptides and polyadenylation signals. An exemplary protein expression construct is an antibody expression construct comprising a sequence encoding an antibody heavy chain and a sequence encoding an antibody light chain, each sequence operatively linked to a promoter (the promoters being the same or different) and either or both sequences optionally operatively linked to an enhancer (and, in the case of both coding sequences being linked to enhancers, the enhancers being the same or different).

[0269] Cleavage enzyme recognition sites include, for example, sequences recognized by restriction endonucleases, homing endonucleases and/or meganucleases. Targeted integration of a cleavage enzyme recognition site (by either homology-dependent or homology-independent mechanisms) is useful for generating cells whose genome contains only a single site that can be cleaved by a particular enzyme. Contacting such cells with an enzyme that recognizes and cleaves at the single site facilitates subsequent targeted integration of exogenous sequences (by either homology-dependent or homology-independent mechanisms) and/or targeted mutagenesis at the site that is cleaved.

[0270] One example of a cleavage enzyme recognition site is that recognized by the homing endonuclease I-*Sce*I, which has the following sequence:

TAGGGATAACAGGGTAAT (SEQ ID NO:213)

See, for example, U.S. Patent No. 6,833,252. Additional exemplary homing endonucleases include 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; 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.

[0271] Although the cleavage specificity of most homing endonucleases is not absolute with respect to their recognition sites, the sites are of sufficient length that a single cleavage event per mammalian-sized genome can be obtained by expressing a homing endonuclease in a cell containing a single copy of its recognition site. It has also been reported that cleavage enzymes 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.

[0272] Previous methods for obtaining targeted recombination and integration using homing endonucleases suffered from the problem that targeted insertion of the recognition site is extremely inefficient, requiring laborious screening to identify cells that contained the recognition site inserted at the desired location. The present methods surmount these problems by allowing highly-efficient targeted integration (either homology-dependent or homology-independent) of a recognition site for a DNA-cleaving enzyme.

[0273] In certain embodiments, targeted integration is used to insert a RNA expression construct, *e.g.*, sequences responsible for regulated expression of micro RNA or siRNA. Promoters, enhancers and additional transcription regulatory sequences, as described above, can also be incorporated in a RNA expression construct.

[0274] In embodiments in which targeted integration occurs by a homology-dependent mechanism, the donor sequence contains sufficient homology, in the regions flanking the exogenous sequence, to support homology-directed repair of a double-strand break in a genomic sequence, thereby inserting the exogenous sequence at the genomic target site. Therefore, the donor nucleic acid can be of any size sufficient to support integration of the exogenous sequence by homology-dependent repair mechanisms (*e.g.*, homologous recombination). Without wishing to be bound by any particular theory, the regions of homology flanking the exogenous sequence are thought to provide the broken chromosome ends with a template for re-synthesis of the genetic information at the site of the double-stranded break.

[0275] Targeted integration of exogenous sequences, as disclosed herein, can be used to generate cells and cell lines for protein expression. *See*, for example, co-owned U.S. Patent Application Publication No. 2006/0063231. For optimal expression of one or more proteins encoded by exogenous sequences integrated into a genome, the chromosomal integration site should be compatible with high-level transcription of the integrated sequences, preferably in a wide range of cell types and developmental states. However, it has been observed that transcription of integrated sequences varies depending on the integration site due to, among other things, the chromatin structure of the genome at the integration site. Accordingly, genomic target sites that support high-level transcription of integrated sequences are desirable. In certain embodiments, it will also be desirable that integration of exogenous sequences not result in ectopic activation of one or more cellular genes (*e.g.*, oncogenes). On the other hand, in the case of integration of promoter and/or enhancer sequences, ectopic expression may be desired.

[0276] For certain embodiments, it is desirable that an integration site is not present in an essential gene (*e.g.*, a gene essential for cell viability), so that inactivation of said essential gene does not result from integration of the exogenous sequences. On the other hand, if the intent is to disable gene function (*i.e.*, create a gene "knock-out") targeted integration of an exogenous sequence to disrupt an endogenous gene is an effective method. In these cases, the exogenous sequence can be any sequence capable of blocking transcription of the endogenous gene or of generating a non-functional translation product, for example a short patch of amino acid sequence, which is optionally detectable (see above). In certain embodiments, the exogenous sequences can comprise a marker gene (described above), allowing selection of cells that have undergone targeted integration.

[0277] Non-limiting examples of chromosomal regions that do not encode an essential gene and support high-level transcription of sequences integrated therein ("safe harbor" integration sites) include the Rosa26 and CCR5 loci.

[0278] The Rosa26 locus has been identified in the murine genome. Zambrowicz *et al.* (1997) *Proc. Natl. Acad. Sci. USA* **94:**3789-3794. The sequence of mouse Rosa26 mRNA was compared to data from a human cDNA screen (Strausberg *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99:**16899-16903), and a homologous human transcript was detected by the present inventors. Accordingly, the human homologue of Rosa26 can be used as a target site for integration of exogenous sequences into the genome of human cells and cell lines, using the methods and compositions disclosed herein.

[0279] CCR5 genomic sequences (including allelic variants such as CCR5-Δ32) are well known in the art. *See, e.g.,* Liu et al. (1996) *Cell* 367-377.

[0280] Additional genomic target sites supporting high-level transcription of integrated sequences can be identified as regions of open chromatin or 'accessible regions" as described, for example in co-owned U.S. Patent Application Publications 2002/0064802 (May 30, 2002) and 2002/0081603 (June 27, 2002).

[0281] The presence of a double-stranded break in a genomic sequence facilitates not only homology-dependent integration of exogenous sequences (*i.e.*, homologous recombination) but also homology-independent integration of exogenous sequences into the genome at the site of the double-strand break. Accordingly, the compositions and methods disclosed herein can be used for targeted cleavage of a genomic sequence, followed by non-homology-dependent integration of an exogenous sequence at or near the targeted cleavage site. For example, a cell can be contacted with one or more ZFP-cleavage domain (or cleavage half-domain) fusion proteins engineered to cleave in a region of interest in a genome as described herein (or one or more polynucleotides encoding such fusion proteins), and a polynucleotide comprising an exogenous sequence lacking homology to the region of interest, to obtain a cell in which all or a portion of the exogenous sequence is integrated in the region of interest.

[0282] The methods of targeted integration (*i.e.*, insertion of an exogenous sequence into a genome), both homology-dependent and -independent, disclosed herein can be used for a number of purposes. These include, but are not limited to, insertion of a gene or cDNA sequence into the genome of a cell to enable expression of the transcription and/or translation products of the gene or cDNA by the cell. For situations in which a disease or pathology can result from one of a plurality of mutations (*e.g.*, multiple point mutations spread across the sequence of the gene), targeted integration (either homology-dependent or homology-independent) of a cDNA copy of the wild-type gene is particularly effective. For example, such a wild-type cDNA is inserted into an untranslated leader sequence or into the first exon of a gene upstream of all known mutations. In certain integrants, in which translational reading frame is preserved, the result is that the wild-type cDNA is expressed and its expression is regulated by the appropriate endogenous transcriptional regulatory sequences. In additional embodiments, such integrated cDNA sequences can include transcriptional (and/or translational) termination signals disposed downstream of the wild-type cDNA and upstream of the mutant endogenous gene. In this way, a wild-type copy of the disease-causing gene is expressed, and the mutant endogenous gene is not expressed. In other embodiments, a portion of a wild-type cDNA is inserted into the appropriate region of a gene (for example, a gene in which disease-causing mutations are clustered).

EXAMPLES

Example 1: Editing of a Chromosomal hSMC1L1 Gene by Targeted Recombination

[0283] The hSMC1L1 gene is the human orthologue of the budding yeast gene *structural maintenance of chromosomes 1*. A region of this gene encoding an amino-terminal portion of the protein which includes the Walker ATPase domain was mutagenized by targeted cleavage and recombination. Cleavage was targeted to the region of the methionine initiation codon (nucleotides 24-26, Figure 1), by designing chimeric nucleases, comprising a zinc finger DNA-binding domain and a *Fok*I cleavage half-domain, which bind in the vicinity of the codon. Thus, two zinc finger binding domains were designed, one of which recognizes nucleotides 23-34 (primary contacts along the top strand as shown in Figure 1), and the other of which recognizes nucleotides 5-16 (primary contacts along the bottom strand). Zinc finger proteins were designed as described in co-owned US Patents 6,453,242 and 6,534,261. *See* Table 2 for the amino acid sequences of the recognition regions of the zinc finger proteins.

[0284] Sequences encoding each of these two ZFP binding domains were fused to sequences encoding a *Fok*I cleavage half-domain (amino acids 384-579 of the native *Fok*I sequence; Kita *et al.* (1989) *J. Biol. Chem.* **264:**5751-5756), such that the encoded protein contained *Fok*I sequences at the carboxy terminus and ZFP sequences at the amino terminus. Each of these fusion sequences was then cloned in a modified mammalian expression vector pcDNA3 (Figure 2).

Table 2: Zinc Finger Designs for the hSMC1L1 Gene

Target sequence	F1	F2	F3	F4
CATGGGGTTCCT (SEQ ID NO: 27)	RSHDLIE (SEQ ID NO: 28)	TSSSLSR (SEQ ID NO: 29)	RSDHLST (SEQ ID NO: 30)	TNSNRIT (SEQ ID NO: 31)
GCGGCGCCGGCG (SEQ ID NO: 32)	RSDDLSR (SEQ ID NO: 33)	RSDDRKT (SEQ ID NO: 34)	RSEDLIR (SEQ ID NO: 35)	RSDTLSR (SEQ ID NO: 36)

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein, and Finger F4 is closest to the carboxy terminus.

[0285] A donor DNA molecule was obtained as follows. First, a 700 base pair fragment of human genomic DNA representing nucleotides 52415936-52416635 of the "-" strand of the X chromosome (UCSC human genome release July, 2003), which includes the first exon of the human hSMC1L1 gene, was amplified, using genomic DNA from HEK293 cells as template. Sequences of primers used for amplification are shown in Table 3 ("Initial amp 1" and "Initial amp 2"). The PCR product was then altered, using standard overlap extension PCR methodology (see, *e.g.*, Ho, *et al.* (1989) *Gene* 77:51-59), resulting in replacement of the sequence ATGGGG (nucleotides 24-29 in Figure 1) to ATAAGAAGC. This change resulted in conversion of the ATG codon (methionine) to an ATA codon (isoleucine) and replacement of GGG (nucleotides 27-29 in Figure 1) by the sequence AGAAGC, allowing discrimination between donor-derived sequences and endogenous chromosomal sequences following recombination. A schematic diagram of the hSMC1 gene, including sequences of the chromosomal DNA in the region of the initiation codon, and sequences in the donor DNA that differ from the chromosomal sequence, is given in Figure 3. The resulting 700 base pair donor fragment was cloned into pCR4BluntTopo, which does not contain any sequences homologous to the human genome. *See* Figure 4.

[0286] For targeted mutation of the chromosomal hSMC1L1 gene, the two plasmids encoding ZFP-*Fok*I fusions and the donor plasmid were introduced into 1×10^6 HEK293 cells by transfection using Lipofectamine 2000® (Invitrogen). Controls included cells transfected only with the two plasmids encoding the ZFP-*Fok*I fusions, cells transfected only with the donor plasmid and cells transfected with a control plasmid (pEGFP-N1, Clontech). Cells were cultured in 5% CO_2 at 37°C. At 48 hours after transfection, genomic DNA was isolated from the cells, and 200 ng was used as template for PCR amplification, using one primer complementary to a region of the gene outside of its region of homology with the donor sequences (nucleotides 52416677-52416701 on the "-" STRAND of the X chromosome; UCSC July 2003), and a second primer complementary to a region of the donor molecule into which distinguishing mutations were introduced. Using these two primers, an amplification product of 400 base pairs will be obtained from genomic DNA if a targeted recombination event has occurred. The sequences of these primers are given in Table 3 (labeled "chromosome-specific" and "donor-specific," respectively). Conditions for amplification were: 94°C, 2 min, followed by 40 cycles of 94°C, 30 sec, 60°C, 1 min, 72°C, 1 min; and a final step of 72°C, 7min.

[0287] The results of this analysis (Figure 5) indicate that a 400 base pair amplification product (labeled "Chimeric DNA" in the Figure) was obtained only with DNA extracted from cells which had been transfected with the donor plasmid and both ZFP-*Fok*I plasmids.

Table 3: Amplification Primers for the hSMC1L1 Gene

Initial amp 1	AGCAACAACTCCTCCGGGGATC (SEQ ID NO: 37)
Initial amp 2	TTCCAGACGCGACTCTTTGGC (SEQ ID NO: 38)
Chromosome-specific	CTCAGCAAGCGTGAGCTCAGGTCTC (SEQ ID NO: 39)
Donor-specific	CAATCAGTTTCAGGAAGCTTCTT (SEQ ID NO: 40)
Outside 1	CTCAGCAAGCGTGAGCTCAGGTCTC (SEQ ID NO: 41)
Outside 2	GGGGTCAAGTAAGGCTGGGAAGC (SEQ ID NO: 42)

[0288] To confirm this result, two additional experiments were conducted. First, the amplification product was cloned into pCR4Blunt-Topo (Invitrogen) and its nucleotide sequence was determined. As shown in Figure 6 (SEQ ID NO: 6), the amplified sequence obtained from chromosomal DNA of cells transfected with the two ZFP-*FokI*-encoding plasmids and the donor plasmid contains the AAGAAGC sequence that is unique to the donor (nucleotides 395-401 of the sequence presented in Figure 6) covalently linked to chromosomal sequences not present in the donor molecule (nucleotides 32-97 of Figure 6), indicating that donor sequences have been recombined into the chromosome. In particular, the G→A mutation converting the initiation codon to an isoleucine codon is observed at position 395 in the sequence.

[0289] In a second experiment, chromosomal DNA from cells transfected only with donor plasmid, cells transfected with both ZFP-*Fok*I fusion plasmids, cells transfected with the donor plasmid and both ZFP-*Fok*I fusion plasmids or cells transfected with the EGFP control plasmid was used as template for amplification, using primers complementary to sequences outside of the 700-nucleotide region of homology between donor and chromosomal sequences (identified as "Outside 1" and "Outside 2" in Table 3). The resulting amplification product was purified and used as template for a second amplification reaction using the donor-specific and chromosome-specific primers described above (Table 3). This amplification yielded a 400 nucleotide product only from cells transfected with the donor construct and both ZFP-*Fok*I fusion constructs, a result consistent with the replacement of genomic sequences by targeted recombination in these cells.

Example 2: Editing of a Chromosomal IL2Rγ Gene by Targeted Recombination

[0290] The IL-2Rγ gene encodes a protein, known as the "common cytokine receptor gamma chain," that functions as a subunit of several interleukin receptors (including IL-2R, IL-4R, IL-7R, IL-9R, IL-15R and IL-21R). Mutations in this gene, including those surrounding the 5' end of the third exon (*e.g.* the tyrosine 91 codon), can cause X-linked severe combined immunodeficiency (SCID). See, for example, Puck et al. (1997) Blood 89:1968-1977. A mutation in the tyrosine 91 codon (nucleotides 23-25 of SEQ ID NO: 7; Figure 7), was introduced into the IL2Rγ gene by targeted cleavage and recombination. Cleavage was targeted to this region by designing two pairs of zinc finger proteins. The first pair (first two rows of Table 4) comprises a zinc finger protein designed to bind to nucleotides 29-40 (primary contacts along the top strand as shown in Figure 7) and a zinc finger protein designed to bind to nucleotides 8-20 (primary contacts along the bottom strand). The second pair (third and fourth rows of Table 4) comprises two zinc finger proteins, the first of which recognizes nucleotides 23-34 (primary contacts along the top strand as shown in Figure 7) and the second of which recognizes nucleotides 8-16 (primary contacts along the bottom strand). Zinc finger proteins were designed as described in co-owned US Patents 6,453,242 and 6,534,261. *See* Table 4 for the amino acid sequences of the recognition regions of the zinc finger proteins.

[0291] Sequences encoding the ZFP binding domains were fused to sequences encoding a *Fok*I cleavage half-domain (amino acids 384-579 of the native *Fok*I sequence, Kita *et al., supra*), such that the encoded protein contained *Fok*I sequences at the carboxy terminus and ZFP sequences at the amino terminus. Each of these fusion sequences was then cloned in a modified mammalian expression vector pcDNA3. See Figure 8 for a schematic diagram of the constructs.

Table 4: Zinc Finger Designs for the IL2Rγ Gene

Target sequence	F1	F2	F3	F4
AACTCGGATA AT (SEQ ID NO: 43)	DRSTLIE (SEQ ID NO:44)	SSSNLSR (SEQ ID NO:45)	RSDDLSK (SEQ ID NO:46)	DNSNRIK (SEQ ID NO:47)
TAGAGGaGAAA GG (SEQ ID NO:48)	RSDNLSN (SEQ ID NO:49)	TSSSRIN (SEQ ID NO:50)	RSDHLSQ (SEQ ID NO:51)	RNADRKT (SEQ ID NO:52)
TACAAGAACT CG (SEQ ID NO:53)	RSDDLSK (SEQ ID NO:54)	DNSNRIK (SEQ ID NO:55)	RSDALSV (SEQ ID NO:56)	DNANRTK (SEQ ID NO:57)
GGAGAAAGG (SEQ ID NO:58)	RSDHLTQ (SEQ ID NO:59)	QSGNLAR (SEQ ID NO:60)	RSDHLSR (SEQ ID NO:61)	

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein.

[0292] A donor DNA molecule was obtained as follows. First, a 700 base pair fragment of human DNA corresponding to positions 69196910-69197609 on the "-" strand of the X chromosome (UCSC, July 2003), which includes exon 3 of the of the IL2Rγ gene, was amplified, using genomic DNA from K562 cells as template. *See* Figure 9. Sequences of primers used for amplification are shown in Table 5 (labeled initial amp 1 and initial amp 2). The PCR product was then altered via standard overlap extension PCR methodology (Ho, *et al., supra*) to replace the sequence TACAAGAACTCGGATAAT (SEQ ID NO:62) with the sequence TAAAAGAATTCCGACAAC (SEQ ID NO:63). This replacement results in the introduction of a point mutation at nucleotide 25 (Figure 7), converting the tyrosine 91 codon TAC to a TAA termination codon and enables discrimination between donor-derived and endogenous chromosomal sequences following recombination, because of differences in the sequences downstream of codon 91. The resulting 700 base pair fragment was cloned into pCR4BluntTopo which does not contain any sequences homologous to the human genome. *See* Figure 10.

[0293] For targeted mutation of the chromosomal IL2Rγ gene, the donor plasmid, along with two plasmids each encoding one of a pair of ZFP-*Fok*I fusions, were introduced into 2×10^6 K652 cells using mixed lipofection/electroporation (Amaxa). Each of the ZFP/*Fok*I pairs (see Table 4) was tested in separate experiments. Controls included cells transfected only with two plasmids encoding ZFP-*Fok*I fusions, and cells transfected only with the donor plasmid. Cells were cultured in 5% CO_2 at 37°C. At 48 hours after transfection, genomic DNA was isolated from the cells, and 200 ng was used as template for PCR amplification, using one primer complementary to a region of the gene outside of its region of homology with the donor sequences (nucleotides 69196839-69196863 on the "+" strand of the X chromosome; UCSC, July 2003), and a second primer complementary to a region of the donor molecule into which distinguishing mutations were introduced (see above) and whose sequence therefore diverges from that of chromosomal DNA. *See* Table 5 for primer sequences, labeled "chromosome-specific" and "donor-specific," respectively. Using these two primers, an amplification product of 500 bp is obtained from genomic DNA in which a targeted recombination event has occurred. Conditions for amplification were: 94°C, 2 min, followed by 35 cycles of 94°C, 30 sec, 62°C, 1 min, 72°C, 45 sec; and a final step of 72°C, 7min.

[0294] The results of this analysis (Figure 11) indicate that an amplification product of the expected size (500 base pairs) is obtained with DNA extracted from cells which had been transfected with the donor plasmid and either of the pairs of ZFP-*Fok*I-encoding plasmids. DNA from cells transfected with plasmids encoding a pair of ZFPs only (no donor plasmid) did not result in generation of the 500 bp product, nor did DNA from cells transfected only with the donor plasmid.

Table 5: Amplification Primers for the IL2Rγ Gene

Initial amp 1	TGTCGAGTACATGAATTGCACTTGG (SEQ ID NO:64)
Initial amp 2	TTAGGTTCTCTGGAGCCCAGGG (SEQ ID NO:65)
Chromosome-specific	CTCCAAACAGTGGTTCAAGAATCTG (SEQ ID NO:66)
Donor-specific	TCCTCTAGGTAAAGAATTCCGACAAC (SEQ ID NO:67)

[0295] To confirm this result, the amplification product obtained from the experiment using the second pair of ZFP/*Fok*I fusions was cloned into pCR4Blunt-Topo (Invitrogen) and its nucleotide sequence was determined. As shown in Figure 12 (SEQ ID NO:12), the sequence consists of a fusion between chromosomal sequences and sequences from the donor plasmid. In particular, the G to A mutation converting tyrosine 91 to a stop codon is observed at position 43 in the sequence. Positions 43-58 contain nucleotides unique to the donor; nucleotides 32-42 and 59-459 are sequences common to the donor and the chromosome, and nucleotides 460-552 are unique to the chromosome. The presence of donor-unique sequences covalently linked to sequences present in the chromosome but not in the donor indicates that DNA from the donor plasmid was introduced into the chromosome by homologous recombination.

Example 3: Editing of a Chromosomal β-globin Gene by Targeted Recombination

[0296] The human beta globin gene is one of two gene products responsible for the structure and function of hemoglobin in adult human erythrocytes. Mutations in the beta-globin gene can result in sickle cell anemia. Two zinc finger proteins were designed to bind within this sequence, near the location of a nucleotide which, when mutated, causes sickle cell anemia. Figure 13 shows the nucleotide sequence of a portion of the human beta-globin gene, and the target sites for the two zinc finger proteins are underlined in the sequence presented in Figure 13. Amino acid sequences of the recognition regions of the two zinc finger proteins are shown in Table 6. Sequences encoding each of these two ZFP binding domains were fused to sequences encoding a *Fok*I cleavage half-domain, as described above, to create engineered ZFP-nucleases that targeted the endogenous beta globin gene. Each of these fusion sequences was then cloned in the mammalian expression vector pcDNA3.1 (Figure 14).

Table 6: Zinc Finger Designs for the beta-globin Gene

Target sequence	F1	F2	F3	F4
GGGCAGTAACGG (SEQ ID NO: 68)	RSDHLSE (SEQ ID NO: 69)	QSANRTK (SEQ ID NO: 70)	RSDNLSA (SEQ ID NO: 71)	RSQNRTR (SEQ ID NO: 72)
AAGGTGAACGTG (SEQ ID NO: 73)	RSDSLSR (SEQ ID NO: 74)	DSSNRKT (SEQ ID NO: 75)	RSDSLSA (SEQ ID NO: 76)	RNDNRKT (SEQ ID NO: 77)

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein, and Finger F4 is closest to the carboxy terminus.

[0297] A donor DNA molecule was obtained as follows. First, a 700 base pair fragment of human genomic DNA corresponding to nucleotides 5212134 - 5212833 on the "-" strand of Chromosome 11 (BLAT, UCSC Human Genome site) was amplified by PCR, using genomic DNA from K562 cells as template. Sequences of primers used for amplification are shown in Table 7 (labeled initial amp 1 and initial amp 2). The resulting amplified fragment contains sequences corresponding to the promoter, the first two exons and the first intron of the human beta globin gene. *See* Figure 15 for a schematic illustrating the locations of exons 1 and 2, the first intron, and the primer binding sites in the beta globin sequence. The cloned product was then further modified by PCR to introduce a set of sequence changes between nucleotides 305-336 (as shown in Figure 13), which replaced the sequence CCGTTACTGCCCTGTGGGGCAAGGTGAACGTG (SEQ ID NO: 78) with **g**CGTTA**g**TGCCC**GAATTCCG**A**tc**GT**c**AAC**cac** (SEQ ID NO: 79) (changes in bold). Certain of these changes (shown in lowercase) were specifically engineered to prevent the ZFP/*Fok*I fusion proteins from binding to and cleaving the donor sequence, once integrated into the chromosome. In addition, all of the sequence changes enable discrimination between donor and endogenous chromosomal sequences following recombination. The resulting 700 base pair fragment was cloned into pCR4-TOPO, which does not contain any sequences homologous to the human genome (Figure 16).

[0298] For targeted mutation of the chromosomal beta globin gene, the two plasmids encoding ZFP-*Fok*I fusions and the donor plasmid (pCR4-TOPO-HBBdonor) were introduced into 1 X 10^6 K562 cells by transfection using Nucleofector™ Solution (Amaxa Biosystems). Controls included cells transfected only with 100 ng (low) or 200 ng (high) of the two plasmids encoding the ZFP-*Fok*I fusions, cells transfected only with 200 ng (low) or 600 ng (high) of the donor plasmid, cells transfected with a GFP-encoding plasmid, and mock transfected cells. Cells were cultured in RPMI Medium 1640 (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 2 mM L-glutamine. Cells were maintained at 37°C in an atmosphere of 5% CO_2. At 72 hours after transfection, genomic DNA was isolated from the cells, and 200 ng was used as template for PCR amplification, using one primer complementary to a region of the gene outside of its region of homology with the donor sequences (nucleotides 5212883-5212905 on the "-" strand of chromosome 11), and a second primer complementary to a region of the donor molecule into which distinguishing mutations were introduced into the donor sequence (see *supra*). The sequences of these primers are given in Table 7 (labeled "chromosome-specific" and "donor-specific," respectively). Using these two primers, an amplification product of 415 base pairs will be obtained from genomic DNA if a targeted recombination event has occurred. As a control for DNA loading, PCR reactions were also carried out using the Initial amp 1 and Initial amp 2 primers to ensure that similar levels of genomic DNA were added to each PCR reaction. Conditions for amplification were: 95°C, 2 min, followed by 40 cycles of 95°C, 30 sec, 60°C, 45 sec, 68°C, 2 min; and a final step of 68°C, 10 min.

[0299] The results of this analysis (Figure 17) indicate that a 415 base pair amplification product was obtained only with DNA extracted from cells which had been transfected with the "high" concentration of donor plasmid and both ZFP-*Fok*I plasmids, consistent with targeted recombination of donor sequences into the chromosomal beta-globin locus.

Table 7: Amplification Primers for the human beta globin gene

Initial amp 1	TACTGATGGTATGGGGCCAAGAG (SEQ ID NO:80)
Initial amp 2	CACGTGCAGCTTGTCACAGTGC (SEQ ID NO:81)
Chromosome-specific	TGCTTACCAAGCTGTGATTCCA (SEQ ID NO:82)
Donor-specific	GGTTGACGATCGGAATTC (SEQ ID NO:83)

[0300] To confirm this result, the amplification product was cloned into pCR4-TOPO (Invitrogen) and its nucleotide sequence was determined. As shown in Figure 18 (SEQ ID NO: 14), the sequence consists of a fusion between chromosomal sequences not present on the donor plasmid and sequences unique to the donor plasmid. For example, two C→G mutations which disrupt ZFP-binding are observed at positions 377 and 383 in the sequence. Nucleotides 377-408 represent sequence obtained from the donor plasmid containing the sequence changes described above; nucleotides 73-376 are sequences common to the donor and the chromosome, and nucleotides 1-72 are unique to the chromosome. The covalent linkage of donor-specific and chromosome-specific sequences in the genome confirms the successful recombination of the donor sequence at the correct locus within the genome of K562 cells.

Example 4: ZFP-FokI linker (ZC linker) optimization

[0301] In order to test the effect of ZC linker length on cleavage efficiency, a four-finger ZFP binding domain was fused to a *Fok*I cleavage half-domain, using ZC linkers of various lengths. The target site for the ZFP is 5'-AACTCGGATAAT-3' (SEQ ID NO:84) and the amino acid sequences of the recognition regions (positions -1 through +6 with respect to the start of the alpha-helix) of each of the zinc fingers were as follows (wherein F1 is the N-most, and F4 is the C-most zinc finger):

F1: DRSTLIE (SEQ ID NO:85)
F2: SSSNLSR (SEQ ID NO:86)
F3: RSDDLSK (SEQ ID NO:87)
F4: DNSNRIK (SEQ ID NO:88)

[0302] ZFP-*Fok*I fusions, in which the aforementioned ZFP binding domain and a *Fok*I cleavage half-domain were separated by 2, 3, 4, 5, 6, or 10 amino acid residues, were constructed. Each of these proteins was tested for cleavage of substrates having an inverted repeat of the ZFP target site, with repeats separated by 4, 5, 6, 7, 8, 9, 12, 15, 16, 17, 22, or 26 basepairs.

[0303] The amino acid sequences of the fusion constructs, in the region of the ZFP-*Fok*I junction (with the ZC linker sequence underlined), are as follows:

10-residue linker	HTKIH<u>LRQKDAARGS</u>QLV (SEQ ID NO:89)
6-residue linker	HTKIH<u>LRQKGS</u>QLV (SEQ ID NO:90)
5-residue linker	HTKIH<u>LRQGS</u>QLV (SEQ ID NO:91)
4-residue linker	HTKIH<u>LRGS</u>QLV (SEQ ID NO:92)
3-residue linker	HTKIH<u>LGS</u>QLV (SEQ ID NO:93)
2-residue linker	HTKIH<u>GS</u>QLV (SEQ ID NO:94)

[0304] The sequences of the various cleavage substrates, with the ZFP target sites underlined, are as follows:

4bp separation
CTAGCATTATCCGAGTTACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAATGTGTTGAGCCTATTACGATC
(SEQ ID NO:95)

5bp separation
CTAGCATTATCCGAGTTCACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGTGTTGAGCCTATTACGATC
(SEQ ID NO:96)

6bp separation
CTAGGCATTATCCGAGTTCACCACAACTCGGATAATGACTAG
GATCCGTAATAGGCTCAAGTGGTGTTGAGCCTATTACTGATC
(SEQ ID NO:97)

7bp separation
CTAGCATTATCCGAGTTCACACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGTGTGTTGAGCCTATTACGATC
(SEQ ID NO:98)

8bp separation
CTAGCATTATCCGAGTTCACCACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGGTGTGTTGAGCCTATTACGATC
(SEQ ID NO:99)

9bp separation
CTAGCATTATCCGAGTTCACACACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGTGTGTGTTGAGCCTATTACGATC
(SEQ ID NO:100)

12bp separation
CTAGCATTATCCGAGTTCACCACCAACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGGTGGTTGTGTTGAGCCTATTACGATC
(SEQ ID NO:101)

15bp separation
CTAGCATTATCCGAGTTCACCACCAACCACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGGTGGTTGGTGTGTTGAGCCTATTACGATC
(SEQ ID NO:102)

16bp separation
CTAGCATTATCCGAGTTCACCACCAACCACACCAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTGGTGGTTGGTGTGGTTGAGCCTATTACGATC
(SEQ ID NO:103)

17bp separation
CTAGCATTATCCGAGTTCAACCACCAACCACACCAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTTGGTGGTTGGTGTGGTTGAGCCTATTACGATC
(SEQ ID NO:104)

22bp separation
CTAGCATTATCCGAGTTCAACCACCAACCACACCAACACAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTTGGTGGTTGGTGTGGTTGTGTTGAGCCTATTACGATC
(SEQ ID NO:105)

26bp separation
CTAGCATTATCCGAGTTCAACCACCAACCACACCAACACCACCAACTCGGATAATGCTAG
GATCGTAATAGGCTCAAGTTGGTGGTTGGTGTGGTTGTGGTGGTTGAGCCTATTACGATC
(SEQ ID NO:106)

[0305] Plasmids encoding the different ZFP-*Fok*I fusion proteins (see above) were constructed by standard molecular biological techniques, and an *in vitro* coupled transcription/translation system was used to express the encoded proteins. For each construct, 200 ng linearized plasmid DNA was incubated in 20 μL TnT mix and incubated at 30° C for 1 hour and 45 minutes. TnT mix contains 100 μl TnT lysate (Promega, Madison, WI) with 4 μl T7 RNA polymerase (Promega) + 2 μl Methionine (1 mM) + 2.5 μl $ZnCl_2$ (20 mM).

[0306] For analysis of DNA cleavage by the different ZFP-*Fok*I fusions, 1 ul of the coupled transcription/translation reaction mixture was combined with approximately 1 ng DNA substrate (end-labeled with ^{32}P using T4 polynucleotide kinase), and the mixture was diluted to a final volume of 19 μl with *Fok*I Cleavage Buffer. *Fok*I Cleavage buffer contains 20 mM Tris-HCl pH 8.5, 75 mM NaCl, 10 μM $ZnCl_2$, 1 mM DTT, 5% glycerol, 500 μg/ml BSA. The mixture was incubated for 1 hour at 37° C. 6.5 μl of FokI buffer, also containing 8 mM $MgCl_2$, was then added and incubation was continued for one hour at 37° C. Protein was extracted by adding 10 μl phenol-chloroform solution to each reaction, mixing, and centrifuging to separate the phases. Ten microliters of the aqueous phase from each reaction was analyzed by electrophoresis on a 10% polyacrylamide gel.

[0307] The gel was subjected to autoradiography, and the cleavage efficiency for each ZFP-*Fok*I fusion/substrate pair was calculated by quantifying the radioactivity in bands corresponding to uncleaved and cleaved substrate, summing to obtain total radioactivity, and determining the percentage of the total radioactivity present in the bands representing cleavage products.

[0308] The results of this experiment are shown in Table 8. This data allows the selection of a ZC linker that provides optimum cleavage efficiency for a given target site separation. This data also allows the selection of linker lengths that allow cleavage at a selected pair of target sites, but discriminate against cleavage at the same or similar ZFP target sites that have a separation that is different from that at the intended cleavage site.

Table 8: DNA cleavage efficiency for various ZC linker lengths and various binding site separations*

	2-residue	3-residue	4-residue	5-residue	6-residue	10-residue
4 bp	74%	81%	74%	12%	6%	4%
5 bp	61%	89%	92%	80%	53%	40%
6 bp	78%	89%	95%	91%	93%	76%
7 bp	15%	55%	80%	80%	70%	80%
8 bp	0%	0%	8%	11%	22%	63%
9 bp	2%	6%	23%	9%	13%	51%
12 bp	8%	12%	22%	40%	69%	84%
15 bp	73%	78%	97%	92%	95%	88%
16 bp	59%	89%	100%	97%	90%	86%
17 bp	5%	22%	77%	71%	85%	82%
22 bp	1%	3%	5%	8%	18%	58%
26 bp	1%	2%	35%	36%	84%	78%

* The columns represent different ZFP-*Fok*I fusion constructs with the indicated number of residues separating the ZFP and the *Fok*I cleavage half-domain. The rows represent different DNA substrates with the indicated number of basepairs separating the inverted repeats of the ZFP target site.

[0309] For ZFP-*Fok*I fusions with four residue linkers, the amino acid sequence of the linker was also varied. In separate constructs, the original LRGS linker sequence (SEQ ID NO:107) was changed to LGGS (SEQ ID NO:108), TGGS (SEQ ID NO:109), GGGS (SEQ ID NO:110), LPGS (SEQ ID NO:111), LRKS (SEQ ID NO:112), and LRWS (SEQ ID NO:113); and the resulting fusions were tested on substrates having a six-basepair separation between binding sites. Fusions containing the LGGS (SEQ ID NO:108) linker sequence were observed to cleave more efficiently than those containing the original LRGS sequence(SEQ ID NO:107). Fusions containing the LRKS(SEQ ID NO:112) and LRWS(SEQ ID NO:113) sequences cleaved with less efficiency than the LRGS sequence(SEQ ID NO:107), while the cleavage efficiencies of the remaining fusions were similar to that of the fusion comprising the original LRGS sequence(SEQ ID NO:107).

Example 5: Increased cleavage specificity resulting from alteration of the *Fok*I cleavage half-domain in the dimerization interface

[0310] A pair of ZFP/*Fok*I fusion proteins (denoted 5-8 and 5-10) were designed to bind to target sites in the fifth exon of the IL-2Rγ gene, to promote cleavage in the region between the target sites. The relevant region of the gene, including the target sequences of the two fusion proteins, is shown in Figure 19. The amino acid sequence of the 5-8 protein is shown in Figure 20, and the amino acid sequence of the 5-10 protein is shown in Figure 21. Both proteins contain a 10 amino acid ZC linker. With respect to the zinc finger portion of these proteins, the DNA target sequences, as well as amino acid sequences of the recognition regions in the zinc fingers, are given in Table 9.

Table 9: Zinc Finger Designs for the IL2Rγ Gene

Fusion	Target sequence	F1	F2	F3	F4
5-8	ACTCTGTGGA AG (SEQ ID NO:114)	RSDNLSE (SEQ ID NO:115)	RNAHRIN (SEQ ID NO:116)	RSDTLSE (SEQ ID NO:117)	ARSTRTT (SEQ ID NO:118)
5-10	AACACGaAAC GTG (SEQ ID NO:119)	RSDSLSR (SEQ ID NO:120)	DSSNRKT (SEQ ID NO:121)	RSDSLSV (SEQ ID NO:122)	DRSNRIT (SEQ ID NO:123)

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein.

[0311] The ability of this pair of fusion proteins to catalyze specific cleavage of DNA between their target sequences (see Figure 19) was tested *in vitro* using a labeled DNA template containing the target sequence and assaying for the presence of diagnostic digestion products. Specific cleavage was obtained when both proteins were used (Table 10, first row). However, the 5-10 fusion protein (comprising a wild-type *Fok*I cleavage half-domain) was also capable of aberrant cleavage at a non-target site in the absence of the 5-8 protein (Table 10, second row), possibly due to self-dimerization.

[0312] Accordingly, 5-10 was modified in its *Fok*I cleavage half-domain by converting amino acid residue 490 from glutamic acid (E) to lysine (K). (Numbering of amino acid residues in the *Fok*I protein is according to Wah *et al., supra.*) This modification was designed to prevent homodimerization by altering an amino acid residue in the dimerization interface. The 5-10 (E490K) mutant, unlike the parental 5-10 protein, was unable to cleave at aberrant sites in the absence of the 5-8 fusion protein (Table 10, Row 3). However, the 5-10 (E490K) mutant, together with the 5-8 protein, catalyzed specific cleavage of the substrate (Table 10, Row 4). Thus, alteration of a residue in the cleavage half-domain of 5-10, that is involved in dimerization, prevented aberrant cleavage by this fusion protein due to self-dimerization. An E490R mutant also exhibits lower levels of homodimerization than the parent protein.

[0313] In addition, the 5-8 protein was modified in its dimerization interface by replacing the glutamine (Q) residue at position 486 with glutamic acid (E). This 5-8 (Q486E) mutant was tested for its ability to catalyze targeted cleavage in the presence of either the wild-type 5-10 protein or the 5-10 (E490K) mutant. DNA cleavage was not observed when the labeled substrate was incubated in the presence of both 5-8 (Q486E) and wild-type 5-10 (Table 10, Row 5). However, cleavage was obtained when the 5-8 (Q486E) and 5-10 (E490K) mutants were used in combination (Table 10, Row 6).

[0314] These results indicate that DNA cleavage by a ZFP/*Fok*I fusion protein pair, at regions other than that defined by the target sequences of the two fusion proteins, can be minimized or abolished by altering the amino acid sequence of the cleavage half-domain in one or both of the fusion proteins.

Table 10: DNA cleavage by ZFP/*Fok*I fusion protein pairs containing wild-type and mutant cleavage half-domains

	ZFP 5-8 binding domain	ZFP 5-10 binding domain	DNA cleavage
1	Wild-type *Fok*I	Wild-type *Fok*I	Specific
2	Not present	Wild-type *Fok*I	Non-specific
3	Not present	*Fok*I E490K	None
4	Wild-type *Fok*I	*Fok*I E490K	Specific
5	*Fok*I Q486E	Wild-type *Fok*I	None
6	*Fok*I Q486E	*Fok*I E490K	Specific

Note: Each row of the table presents results of a separate experiment in which ZFP/*Fok*I fusion proteins were tested for cleavage of a labeled DNA substrate. One of the fusion proteins contained the 5-8 DNA binding domain, and the other fusion protein contained the 5-10 DNA binding domain (See Table 9 and Figure 19). The cleavage half-domain portion of the fusion proteins was as indicated in the Table. Thus, the entries in the ZFP 5-8 column indicate the type of *Fok*I cleavage domain fused to ZFP 5-8; and the entries in the ZFP 5-10 column indicates the type of *Fok*I cleavage domain fused to ZFP 5-10. For the *Fok*I cleavage half-domain mutants, the number refers to the amino acid residue in the *Fok*I protein; the letter preceding the number refers to the amino acid present in the wild-type protein and the letter following the number denotes the amino acid to which the wild-type residue was changed in generating the modified protein.
'Not present' indicates that the entire ZFP/*Fok*I fusion protein was omitted from that particular experiment.
The DNA substrate used in this experiment was an approximately 400 bp PCR product containing the target sites for both ZFP 5-8 and ZFP 5-10. See Figure 19 for the sequences and relative orientation of the two target sites.

Example 6: Generation of a defective enhanced Green Fluorescent Protein (eGFP) gene

[0315] The enhanced Green Fluorescent Protein (eGFP) is a modified form of the Green Fluorescent Protein (GFP; *see, e.g.,* Tsien (1998) *Ann. Rev. Biochem.* **67**:509-544) containing changes at amino acid 64 (phe to leu) and 65 (ser to thr). Heim *et al.* (1995) *Nature* **373**:663-664; Cormack *et al.* (1996) *Gene* **173**:33-38. An eGFP-based reporter system was constructed by generating a defective form of the eGFP gene, which contained a stop codon and a 2-bp frameshift mutation. The sequence of the eGFP gene is shown in Figure 22. The mutations were inserted by overlapping PCR mutagenesis, using the Platinum® *Taq* DNA Polymerase High Fidelity kit (Invitrogen) and the oligonucleotides GFP-Bam, GFP-Xba, stop sense2, and stop anti2 as primers (oligonucleotide sequences are listed below in Table 11). GFP-Bam and GFP-Xba served as the external primers, while the primers stop sense2 and stop anti2 served as the internal primers encoding the nucleotide changes. The peGFP-NI vector (BD Biosciences), encoding a full-length eGFP gene, was used as the DNA template in two separate amplification reactions, the first utilizing the GFP-Bam and stop anti2 oligonucleotides as primers and the second using the GFP-Xba and stop sense2 oligonucleotides as primers. This generated two amplification products whose sequences overlapped. These products were combined and used as template in a third amplification reaction, using the external GFP-Bam and GFP-Xba oligonucleotides as primers, to regenerate a modified eGFP gene in which the sequence GACCACAT (SEQ ID NO: 124) at nucleotides 280-287 was replaced with the sequence TAACAC (SEQ ID NO: 125). The PCR conditions for all amplification reactions were as follows: the template was initially denatured for 2 minutes at 94 degrees and followed by 25 cycles of amplification by incubating the reaction for 30 sec. at 94 degrees C, 45 sec. at 46 degrees C, and 60 sec. at 68 degrees C. A final round of extension was carried out at 68 degrees C for 10 minutes. The sequence of the final amplification product is shown in Figure 23. This 795 bp fragment was cloned into the pCR(R)4-TOPO vector using the TOPO-TA cloning kit (Invitrogen) to generate the pCR(R)4-TOPO-GFPmut construct.

Table 11: Oligonucleotide sequences for GFP

Oligo	sequence 5'-3'
GFP-Bam	CGAATTCTGCAGTCGAC (SEQ ID NO:126)
GFP-Xba	GATTATGATCTAGAGTCG (SEQ ID NO:127)
stop sense2	AGCCGCTACCCCTAACACGAAGCAG (SEQ ID NO:128)
stop anti2	CTGCTTCGTGTTAGGGGTAGCGGCT (SEQ ID NO:129)

Example 7: Design and assembly of Zinc Finger Nucleases targeting eGFP

[0316] Two three-finger ZFPs were designed to bind a region of the mutated GFP gene (Example 6) corresponding to nucleotides 271-294 (numbering according to Figure 23). The binding sites for these proteins occur in opposite orientation with 6 base pairs separating the two binding sites. *See* Figure 23. ZFP 287A binds nucleotides 271-279 on the non-coding strand, while ZFP 296 binds nucleotides 286-294 on the coding strand. The DNA target and amino acid sequence for the recognition regions of the ZFPs are listed below, and in Table 12:

287A:

F1 (GCGg) RSDDLTR (SEQ ID NO: 130)
F2 (GTA) QSGALAR (SEQ ID NO:131)
F3 (GGG) RSDHLSR (SEQ ID NO:132)

296S:

F1 (GCA) QSGSLTR (SEQ ID NO:133)
F2 (GCA) QSGDLTR (SEQ ID NO:134)
F3 (GAA) QSGNLAR (SEQ ID NO:135)

Table 12: Zinc finger designs for the GFP gene

Protein	Target sequence	F1	F2	F3
287A	GGGGTAGCGg (SEQ ID NO:136)	RSDDLTR (SEQ ID NO:137)	QSGALAR (SEQ ID NO:138)	RSDHLSR (SEQ ID NO:139)
296S	GAAGCAGCA (SEQ ID NO:140)	QSGSLTR (SEQ ID NO:141)	QSGDLTR (SEQ ID NO:142)	QSGNLAR (SEQ ID NO:143)

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein, and Finger F3 is closest to the carboxy terminus.

[0317] Sequences encoding these proteins were generated by PCR assembly (*e.g.*, U.S. Patent No. 6,534,261), cloned between the *Kpn*I and *Bam*HI sites of the pcDNA3.1 vector (Invitrogen), and fused in frame with the catalytic domain of the

*Fok*I endonuclease (amino acids 384-579 of the sequence of Looney *et al.* (1989) *Gene* **80:**193-208). The resulting constructs were named pcDNA3.1-GFP287-FokI and pcDNA3.1-GFP296-FokI (Figure 24).

Example 8: Targeted *in vitro* DNA cleavage by designed Zinc Finger Nucleases

[0318] The pCR(R)4-TOPO-GFPmut construct (Example 6) was used to provide a template for testing the ability of the 287 and 296 zinc finger proteins to specifically recognize their target sites and cleave this modified form of eGFP *in vitro*.

[0319] A DNA fragment containing the defective eGFP-encoding insert was obtained by PCR amplification, using the T7 and T3 universal primers and pCR(R)4-TOPO-GFPmut as template. This fragment was end-labeled using γ–^{32}P-ATP and T4 polynucleotide kinase. Unincorporated nucleotide was removed using a microspin G-50 column (Amersham).

[0320] An *in vitro* coupled transcription/translation system was used to express the 287 and 296 zinc finger nucleases described in Example 7. For each construct, 200 ng linearized plasmid DNA was incubated in 20 μL TnT mix and incubated at 30° C for 1 hour and 45 minutes. TnT mix contains 100 μl TnT lysate (which includes T7 RNA polymerase, Promega, Madison, WI) supplemented with 2 μl Methionine (1 mM) and 2.5 μl $ZnCl_2$ (20 mM).

[0321] For analysis of DNA cleavage, aliquots from each of the 287 and 296 coupled transcription/translation reaction mixtures were combined, then serially diluted with cleavage buffer. Cleavage buffer contains 20 mM Tris-HCl pH 8.5, 75 mM NaCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT, 5% glycerol, 500 μg/ml BSA. 5μl of each dilution was combined with approximately 1 ng DNA substrate (end-labeled with ^{32}P using T4 polynucleotide kinase as described above), and each mixture was further diluted to generate a 20 μl cleavage reaction having the following composition: 20 mM Tris-HCl pH 8.5, 75 mM NaCl, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT, 5% glycerol, 500 μg/ml BSA. Cleavage reactions were incubated for 1 hour at 37°C. Protein was extracted by adding 10 μl phenol-chloroform solution to each reaction, mixing, and centrifuging to separate the phases. Ten microliters of the aqueous phase from each reaction was analyzed by electrophoresis on a 10% polyacrylamide gel.

[0322] The gel was subjected to autoradiography, and the results of this experiment are shown in Figure 25. The four left-most lanes show the results of reactions in which the final dilution of each coupled transcription/translation reaction mixture (in the cleavage reaction) was 1/156.25, 1/31.25, 1/12.5 and 1/5, respectively, resulting in effective volumes of 0.032, 0.16, 04. and 1 ul, respectively of each coupled transcription/translation reaction. The appearance of two DNA fragments having lower molecular weights than the starting fragment (lane labeled "uncut control" in Figure 25) is correlated with increasing amounts of the 287 and 296 zinc finger endonucleases in the reaction mixture, showing that DNA cleavage at the expected target site was obtained.

Example 9: Generation of stable cell lines containing an integrated defective eGFP gene

[0323] A DNA fragment encoding the mutated eGFP, eGFPmut, was cleaved out of the pCR(R)4-TOPO-GFPmut vector (Example 6) and cloned into the *Hind*III and *Not*I sites of pcDNA4/TO, thereby placing this gene under control of a tetracycline-inducible CMV promoter. The resulting plasmid was named pcDNA4/TO/GFPmut (Figure 26). T-Rex 293 cells (Invitrogen) were grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% Tet-free fetal bovine serum (FBS) (HyClone). Cells were plated into a 6-well dish at 50% confluence, and two wells were each transfected with pcDNA4/TO/GFPmut. The cells were allowed to recover for 48 hours, then cells from both wells were combined and split into 10x15-cm^2 dishes in selective medium, *i.e.*, medium supplemented with 400 ug/ml Zeocin (Invitrogen). The medium was changed every 3 days, and after 10 days single colonies were isolated and expanded further. Each clonal line was tested individually for doxycycline(dox)-inducible expression of the eGFPmut gene by quantitative RT-PCR (TaqMan®).

[0324] For quantitative RT-PCR analysis, total RNA was isolated from dox-treated and untreated cells using the High Pure Isolation Kit (Roche Molecular Biochemicals), and 25 ng of total RNA from each sample was subjected to real time quantitative RT-PCR to analyze endogenous gene expression, using TaqMan® assays. Probe and primer sequences are shown in Table 13. Reactions were carried out on an ABI 7700 SDS machine (PerkinElmer Life Sciences) under the following conditions. The reverse transcription reaction was performed at 48° C for 30 minutes with

MultiScribe reverse transcriptase (PerkinElmer Life Sciences), followed by a 10-minute denaturation step at 95°C. Polymerase chain reaction (PCR) was carried out with AmpliGold DNA polymerase (PerkinElmer Life Sciences) for 40 cycles at 95°C for 15 seconds and 60°C for 1 minute. Results were analyzed using the SDS version 1.7 software and are shown in Figure 27, with expression of the eGFPmut gene normalized to the expression of the human GAPDH gene. A number of cell lines exhibited doxycycline-dependent expression of eGFP; line 18 (T18) was chosen as a model cell line for further studies.

Table 13: Oligonucleotides for mRNA analysis

Oligonucleotide	Sequence
eGFP primer 1 (5T)	CTGCTGCCCGACAACCA (SEQ ID NO:144)
eGFP primer 2 (3T)	CCATGTGATCGCGCTTCTC (SEQ ID NO:145)
eGFP probe	CCCAGTCCGCCCTGAGCAAAGA (SEQ ID NO:146)
GAPDH primer 1	CCATGTTCGTCATGGGTGTGA (SEQ ID NO:147)
GAPDH primer 2	CATGGACTGTGGTCATGAGT (SEQ ID NO:148)
GAPDH probe	TCCTGCACCACCAACTGCTTAGCA (SEQ ID NO:149)

Example 10: Generation of a donor sequence for correction of a defective chromosomal eGFP gene

[0325] A donor construct containing the genetic information for correcting the defective eGFPmut gene was constructed by PCR. The PCR reaction was carried out as described above, using the peGFP-NI vector as the template. To prevent background expression of the donor construct in targeted recombination experiments, the first 12 bp and start codon were removed from the donor by PCR using the primers GFPnostart and GFP-Xba (sequences provided in Table 14). The resulting PCR fragment (734 bp) was cloned into the pCR(R)4-TOPO vector, which does not contain a mammalian cell promoter, by TOPO-TA cloning to create pCR(R)4-TOPO-GFPdonor5 (Figure 28). The sequence of the eGFP insert of this construct (corresponding to nucleotides 64-797 of the sequence shown in Figure 22) is shown in Figure 29 (SEQ ID NO:20).

Table 14: Oligonucleotides for construction of donor molecule

Oligonucleotide	Sequence 5'-3'
GFPnostart	GGCGAGGAGCTGTTCAC (SEQ ID NO:150)
GFP-Xba	GATTATGATCTAGAGTCG (SEQ ID NO:151)

Example 11: Correction of a mutation in an integrated chromosomal eGFP gene by targeted cleavage and recombination

[0326] The T18 stable cell line (Example 9) was transfected with one or both of the ZFP-FokI expression plasmid (pcDNA3.1-GFP287-FokI and pcDNA3.1-GFP296-FokI, Example 7) and 300 ng of the donor plasmid pCR(R)4-TOPO-GFPdonor5 (Example 10) using LipofectAMINE 2000 Reagent (Invitrogen) in Opti-MEM I reduced serum medium, according to the manufacturer's protocol. Expression of the defective chromosomal eGFP gene was induced 5-6 hours after transfection by the addition of 2 ng/ml doxycycline to the culture medium. The cells were arrested in the G2 phase of the cell cycle by the addition, at 24 hours post-transfection, of 100 ng/ml Nocodazole (Figure 30) or 0.2 μM Vinblastine (Figure 31). G2 arrest was allowed to continue for 24-48 hours, and was then released by the removal of the medium. The cells were washed with PBS and the medium was replaced with DMEM containing tetracycline-free FBS and 2 ng/ml doxycycline. The cells were allowed to recover for 24-48 hours, and gene correction efficiency was measured by monitoring the number of cells exhibiting eGFP fluorescence, by fluorescence-activated cell sorting (FACS) analysis. FACS analysis was carried out using a Beckman-Coulter EPICS XL-MCL instrument and System II Data Acquisition and Display software, version 2.0. eGFP fluorescence was detected by excitation at 488 nm with an argon laser and monitoring emissions at 525 nm (x-axis). Background or autofluorescence was measured by monitoring emissions at 570 nm (y-axis). Cells exhibiting high fluorescent emission at 525 nm and low emission at 570 nm (region E) were scored positive for gene correction.

[0327] The results are summarized in Table 15 and Figures 30 and 31. Figures 30 and 31 show results in which T18 cells were transfected with the pcDNA3.1-GFP287-FokI and pcDNA3.1-GFP296-FokI plasmids encoding ZFP nucleases and the pCR(R)4-TOPO-GFPdonor5 plasmid, eGFP expression was induced with doxycycline, and cells were arrested in G2 with either nocodazole (Figure 30) or vinblastine (Figure 31). Both figures show FACS traces, in which cells exhibiting eGFP fluorescence are represented in the lower right-hand portion of the trace (identified as Region E, which is the portion of Quadrant 4 underneath the curve). For transfected cells that had been treated with nocodazole, 5.35% of the cells exhibited GFP fluorescence, indicative of correction of the mutant chromosomal eGFP gene (Figure 30), while 6.7% of cells treated with vinblastine underwent eGFP gene correction (Figure 31). These results are summarized, along with additional control experiments, in Rows 1-8 of Table 15.

[0328] In summary, these experiments show that, in the presence of two ZFP nucleases and a donor sequence, approximately 1% of treated cells underwent gene correction, and that this level of correction was increased 4-5 fold by arresting treated cells in the G2 phase of the cell cycle.

Table 15: Correction of a defective chromosomal eGFP gene

Expt.	Treatment[1]	Percent cells with corrected eGFP gene[2]
1	300 ng donor only	0.01
2	100 ng ZFP 287 + 300 ngdonor	0.16
3	100 ng ZFP 296 + 300 ng donor	0.6
4	50 ng ZFP 287 + 50 ng ZFP 296 + 300 ng donor	1.2
5	as 4 + 100 ng/ml nocodazole	5.35
6	as 4 + 0.2 μM vinblastine	6.7
7	no donor, no ZFP, 100 ng/ml nocodazole	0.01
8	no donor, no ZFP, 0.2 μM vinblastine	0.0
9	100 ng ZFP287/Q486E + 300 ng donor	0.0
10	100 ng ZFP296/E490K + 300 ng donor	0.01
11	50 ng 287/Q486E + 50 ng 296/E490K + 300 ng donor	0.62
12	as 11 + 100 ng/ml nocodazole	2.37
13	as 11 + 0.2 μM vinblastine	2.56

Notes:

1: T18 cells, containing a defective chromosomal eGFP gene, were transfected with plasmids encoding one or two ZFP nucleases and/or a donor plasmid encoding a nondefective eGFP sequence, and expression of the chromosomal eGFP gene was induced with doxycycline. Cells were optionally arrested in G2 phase of the cell cycle after eGFP induction. FACS analysis was conducted 5 days after transfection.

2: The number is the percent of total fluorescence exhibiting high emission at 525 nm and low emission at 570 nm (region E of the FACS trace).

Example 12: Correction of a defective chromosomal gene using zinc finger nucleases with sequence alterations in the dimerization interface

[0329] Zinc finger nucleases whose sequences had been altered in the dimerization interface were tested for their ability to catalyze correction of a defective chromosomal eGFP gene. The protocol described in Example 11 was used, except that the nuclease portion of the ZFP nucleases (*i.e.*, the *Fok*I cleavage half-domains) were altered as described in Example 5. Thus, an E490K cleavage half-domain was fused to the GFP296 ZFP domain (Table 12), and a Q486E cleavage half-domain was fused to the GFP287 ZFP (Table 12).

[0330] The results are shown in Rows 9-11 of Table 15 and indicate that a significant increase in the frequency of gene correction was obtained in the presence of two ZFP nucleases having alterations in their dimerization interfaces, compared to that obtained in the presence of either of the nucleases alone. Additional experiments, in which T18 cells were transfected with donor plasmid and plasmids encoding the 287/Q486E and 296/E490K zinc finger nucleases, then arrested in G2 with nocodazole or vinblastine, showed a further increase in frequency of gene correction, with over 2% of cells exhibiting eGFP fluorescence, indicative of a corrected chromosomal eGFP gene (Table 15, Rows 12 and 13).

Example 13: Effect of donor length on frequency of gene correction

[0331] In an experiment similar to those described in Example 11, the effect of the length of donor sequence on frequency of targeted recombination was tested. T18 cells were transfected with the two ZFP nucleases, and eGFP expression was induced with doxycycline, as in Example 11. Cells were also transfected with either the pCR(R)4-TOPO-GFPdonor5 plasmid (Figure 28) containing a 734 bp eGFP insert (Figure 29) as in Example 11, or a similar plasmid containing a 1527 bp sequence insert (Figure 32) homologous to the mutated chromosomal eGFP gene. Additionally, the effect of G2 arrest with nocodazole on recombination frequency was assessed.

[0332] In a second experiment, donor lengths of 0.7, 1.08 and 1.5 kbp were compared. T18 cells were transfected with 50 ng of the 287-FokI and 296-FokI expression plasmids (Example 7, Table 12) and 500ng of a 0.7 kbp , 1.08 kbp, or 1.5 kbp donors, as described in Example 11. Four days after transfection, cells were assayed for correction of the defective eGFP gene by FACS, monitoring GFP fluorescence.

[0333] The results of these two experiments, shown in Table 16, show that longer donor sequence increases the frequency of targeted recombination (and, hence, of gene correction) and confirm that arrest of cells in the G2 phase of the cell cycle also increases the frequency of targeted recombination.

Table 16: Effect of donor length and cell-cycle arrest on targeted recombination frequency

Donor length (kb)	Experiment 1 Nocodazole concentration: 0 ng/ml	100 ng/ml	Experiment 2 -
0.7	1.41	5.84	1.2
1.08	not done	not done	2.2
1.5	2.16	8.38	2.3

Note: Numbers represent percentage of total fluorescence in Region E of the FACS trace (see Example 11) which is an indication of the fraction of cells that have undergone targeted recombination to correct the defective chromosomal eGFP gene.

Example 14: Editing of the endogenous human IL-2Rγ gene by targeted cleavage and recombination using zinc finger nucleases

[0334] Two expression vectors, each encoding a ZFP-nuclease targeted to the human IL-2Rγ gene, were constructed. Each ZFP-nuclease contained a zinc finger protein-based DNA binding domain (see Table 17) fused to the nuclease domain of the type IIS restriction enzyme *Fok*I (amino acids 384-579 of the sequence of Looney *et al.* (1989) *Gene* **80:**193-208) via a four amino acid ZC linker (see Example 4). The nucleases were designed to bind to positions in exon 5 of the chromosomal IL-2Rγ gene surrounding codons 228 and 229 (a mutational hotspot in the gene) and to introduce a double-strand break in the DNA between their binding sites.

Table 17: Zinc Finger Designs for exon 5 of the IL2Rγ Gene

Target sequence	F1	F2	F3	F4
ACTCTGTGGAAG (SEQ ID NO:152) 5-8G	RSDNLSV (SEQ ID NO:153)	RNAHRIN (SEQ ID NO:154)	RSDTLSE (SEQ ID NO:155)	ARSTRTN (SEQ ID NO:156)
AAAGCGGCTCCG (SEQ ID NO:157) 5-9D	RSDTLSE (SEQ ID NO:158)	ARSTRTT (SEQ ID NO:159)	RSDSLSK (SEQ ID NO:160)	QRSNLKV (SEQ ID NO:161)

Note: The zinc finger amino acid sequences shown above (in one-letter code) represent residues -1 through +6, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein.

[0335] The complete DNA-binding portion of each of the chimeric endonucleases was as follows:

Nuclease targeted to ACTCTGTGGAAG (SEQ ID NO:152)

MAERPFQCRICMRNFSRSDNLSVHIRTHTGEKPFACDICGRKFARNAHRINHTKIHTGSQKPFQCRICMRNFSRSDTLSEHIRTHTGEKPFACDICGRKFAARSTRTNHTKIHLRGS (SEQ ID NO:162)

Nuclease targeted to AAAGCGGCTCCG (SEQ ID NO:157)

MAERPFQCRICMRNFSRSDTLSEHIRTHTGEKPFACDICGRKFAARSTRTTHTKIHTGSQKPFQCRICMRNFSRSDSLSKHIRTHTGEKPFACDICGRKFAQRSNLKVHTKIHLRGS (SEQ ID NO:163)

[0336] Human embryonic kidney 293 cells were transfected (Lipofectamine 2000; Invitrogen) with two expression constructs, each encoding one of the ZFP-nucleases described in the preceding paragraph. The cells were also transfected with a donor construct carrying as an insert a 1,543 bp fragment of the IL2Rγ locus corresponding to positions 69195166-69196708 of the "minus" strand of the X chromosome (UCSC human genome release July 2003), in the pCR4Blunt Topo (Invitrogen) vector. The IL-2Rγ insert sequence contained the following two point mutations in the sequence of exon 5 (underlined):

```
 F  R  V  R  S  R  F  N  P  L  C  G  S  (SEQ ID NO:164)
TTTCGTGTTCGGAGCCGGTTTAACCCGCTCTGTGGAAGT (SEQ ID NO:165)
```

[0337] The first mutation (CGC→CGG) does not change the amino acid sequence (upper line) and serves to adversely affect the ability of the ZFP-nuclease to bind to the donor DNA, and to chromosomal DNA following recombination. The second mutation (CCA→CCG) does not change the amino acid sequence and creates a recognition site for the restriction enzyme *Bsr*BI.

[0338] Either 50 or 100 nanograms of each ZFP-nuclease expression construct and 0.5 or 1 microgram of the donor construct were used in duplicate transfections. The following control experiments were also performed: transfection with an expression plasmid encoding the eGFP protein; transfection with donor construct only; and transfection with plasmids expressing the ZFP nucleases only. Twenty four hours after transfection, vinblastine (Sigma) was added to 0.2 μM final concentration to one sample in each set of duplicates, while the other remained untreated. Vinblastine affects the cell's ability to assemble the mitotic spindle and therefore acts as a potent G_2 arresting agent. This treatment was performed to enhance the frequency of targeting because the homology-directed double-stranded break repair pathway is more active than non-homologous end-joining in the G_2 phase of the cell cycle. Following a 48 hr period of treatment with 0.2 μM vinblastine, growth medium was replaced, and the cells were allowed to recover from vinblastine treatment for an additional 24 hours. Genomic DNA was then isolated from all cell samples using the DNEasy Tissue Kit (Qiagen). Five hundred nanograms of genomic DNA from each sample was then assayed for frequency of gene targeting, by testing for the presence of a new *Bsr*BI site in the chromosomal IL-2Rγ locus, using the assay described schematically in Figure 33.

[0339] In brief, 20 cycles of PCR were performed using the primers shown in Table 18, each of which hybridizes to the chromosomal IL-2Rγ locus immediately outside of the region homologous to the 1.5 kb donor sequence. Twenty microcuries each of α-^{32}P-dCTP and α-^{32}P-dATP were included in each PCR reaction to allow detection of PCR products. The PCR reactions were desalted on a G-50 column (Amersham), and digested for 1 hour with 10 units of *Bsr*BI (New England Biolabs). The digestion products were resolved on a 10% non-denaturing polyacrylamide gel (BioRad), and the gel was dried and autoradiographed (Figure 34). In addition to the major PCR product, corresponding to the 1.55 kb amplififed fragment of the IL2Rγ locus ("wt" in Figure 34), an additional band ("rflp" in Figure 34) was observed in lanes corresponding to samples from cells that were transfected with the donor DNA construct and both ZFP-nuclease constructs. This additional band did not appear in any of the control lanes, indicating that ZFP nuclease-facilitated recombination of the *Bsr*BI RFLP-containing donor sequence into the chromosome occurred in this experiment.

[0340] Additional experiments, in which trace amounts of a RFLP-containing IL-2Rγ DNA sequence was added to human genomic DNA (containing the wild-type IL-2Rγ gene), and the resultant mixture was amplified and subjected to digestion with a restriction enzyme which cleaves at the RFLP, have indicated that as little as 0.5% RFLP-containing sequence can be detected quantitatively using this assay.

Table 18: Oligonucleotides for analysis of the human IL-2Rγ gene

Oligonucleotide	Sequence
Ex5_1.5detF1	GATTCAACCAGACAGATAGAAGG (SEQ ID NO:166)
Ex5_1.5detR1	TTACTGTCTCATCCTTTACTCC (SEQ ID NO:167)

Example 15: Targeted recombination at the IL-2Rγ locus in K562 cells

[0341] K562 is a cell line derived from a human chronic myelogenous leukemia. The proteins used for targeted cleavage were *Fok*I fusions to the 5-8G and 5-9D zinc finger DNA-binding domains (Example 14, Table 17). The donor sequence was the 1.5 kbp fragment of the human IL-2Rγ gene containing a *Bsr*BI site introduced by mutation, described in Example 14.

[0342] K562 cells were cultured in RPMI Medium 1640 (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 2 mM L-glutamine. All cells were maintained at 37°C in an atmosphere of 5% CO_2. These cells were transfected by Nucleofection™ (Solution V, Program T16) (Amaxa Biosystems), according to the manufacturers' protocol, transfecting 2 million cells per sample. DNAs for transfection, used in various combinations as described below, were a plasmid encoding the 5-8G ZFP-*Fok*I fusion endonuclease, a plasmid encoding the 5-9D ZFP-*Fok*I fusion endonuclease, a plasmid containing the donor sequence (described above and in Example 14) and the peGFP-N1 vector (BD Biosciences) used as a control.

[0343] In the first experiment, cells were transfected with various plasmids or combinations of plasmids as shown in Table 19.

Table 19

Sample #	p-eGFP-N1	p5-8G	p5-9D	donor	vinblastine
1	5 μg	-	-	-	-
2	-	-	-	50 μg	-
3	-	-	-	50 μg	yes
4	-	10 μg	10 μg	-	-
5	-	5 μg	5 μg	25 μg	-
6	-	5 μg	5 μg	25 μg	yes
7	-	7.5 μg	7.5 μg	25 μg	-
8	-	7.5 μg	7.5 μg	25 μg	yes
9	-	7.5 μg	7.5 μg	50 μg	-
10	-	7.5 μg	7.5 μg	50 μg	yes

[0344] Vinblastine-treated cells were exposed to 0.2 μM vinblastine at 24 hours after transfection for 30 hours. The cells were collected, washed twice with PBS, and re-plated in growth medium. Cells were harvested 4 days after transfection for analysis of genomic DNA.

[0345] Genomic DNA was extracted from the cells using the DNEasy kit (Qiagen). One hundred nanograms of genomic DNA from each sample were used in a PCR reaction with the following primers:

[0346] Exon 5 forward: GCTAAGGCCAAGAAAGTAGGGCTAAAG (SEQ ID NO:168)

[0347] Exon 5 reverse: TTCCTTCCATCACCAAACCCTCTTG (SEQ ID NO:169)

[0348] These primers amplify a 1,669 bp fragment of the X chromosome corresponding to positions 69195100-69196768 on the "-" strand (UCSC human genome release July 2003) that contain exon 5 of the IL2Rγ gene. Amplification of genomic DNA which has undergone homologous recombination with the donor DNA yields a product containing a *Bsr*BI site; whereas the amplification product of genomic DNA which has not undergone homologous recombination with donor DNA will not contain this restriction site.

[0349] Ten microcuries each of α-^{32}PdCTP and α-^{32}PdATP were included in each amplification reaction to allow visualization of reaction products. Following 20 cycles of PCR, the reaction was desalted on a Sephadex G-50 column (Pharmacia), and digested with 10 Units of *Bsr*BI (New England Biolabs) for 1 hour at 37°C. The reaction was then resolved on a 10% non-denaturing PAGE, dried, and exposed to a PhosphorImager screen.

[0350] The results of this experiment are shown in Figure 35. When cells were transfected with the control GFP plasmid, donor plasmid alone or the two ZFP-encoding plasmids in the absence of donor, no BsrBI site was present in the amplification product, as indicated by the absence of the band marked "rflp" in the lanes corresponding to these samples in Figure 35. However, genomic DNA of cells that were transfected with the donor plasmid and both ZFP-encoding plasmids contained the *Bsr*BI site introduced by homologous recombination with the donor DNA (band labeled "rflp"). Quantitation of the percentage of signal represented by the RFLP-containing DNA, shown in Figure 35, indicated that, under optimal conditions, up to 18% of all IL-2Rγ genes in the transfected cell population were altered by homologous recombination.

[0351] A second experiment was conducted according to the protocol just described, except that the cells were expanded for 10 days after transfection. DNAs used for transfection are shown in Table 20.

Table 20

Sample #	p-eGFP-N1	p5-8G	p5-9D	donor	vinblastine
1	50 μg	-	-	-	-
2	-	-	-	50 μg	-
3	-	-	-	50 μg	yes
4	-	7.5 μg	7.5 μg	-	-
5	-	5 μg	5 μg	25 μg	-
6	-	5 μg	5 μg	25 μg	yes
7	-	7.5 μg	7.5 μg	50 μg	-
8	-	7.5 μg	7.5 μg	50 μg	yes

[0352] Analysis of *Bsr*BI digestion of amplified DNA, shown in Figure 36, again demonstrated that up to 18% of IL-2Rγ genes had undergone sequence alteration through homologous recombination, after multiple rounds of cell division. Thus, the targeted recombination events are stable.

[0353] In addition, DNA from transfected cells in this second experiment was analyzed by Southern blotting. For this analysis, twelve micrograms of genomic DNA from each sample were digested with 100 units *Eco*RI, 50 units *Bsr*BI, and 40 units of

*Dpn*I (all from New England Biolabs) for 12 hours at 37°C. This digestion generates a 7.7 kbp *Eco* RI fragment from the native IL-2Rγ gene (lacking a *Bsr*BI site) and fragments of 6.7 and 1.0 kbp from a chromosomal IL-2Rγ gene whose sequence has been altered, by homologous recombination, to include the *Bsr*BI site. *Dpn*I, a methylation-dependent restriction enzyme, was included to destroy the *dam*-methylated donor DNA. Unmethylated K562 cell genomic DNA is resistant to *Dpn*I digestion.

[0354] Following digestion, genomic DNA was purified by phenol-chloroform extraction and ethanol precipitation, resuspended in TE buffer, and resolved on a 0.8% agarose gel along with a sample of genomic DNA digested with EcoRI and SphI to generate a size marker. The gel was processed for alkaline transfer following standard procedure and DNA was transferred to a nylon membrane (Schleicher and Schuell). Hybridization to the blot was then performed by using a radiolabelled fragment of the IL-2Rγ locus corresponding to positions 69198428-69198769 of the "-" strand of the X chromosome (UCSC human genome July 2003 release). This region of the gene is outside of the region homologous to donor DNA. After hybridization, the membrane was exposed to a PhosphorImager plate and the data quantitated using Molecular Dynamics software. Alteration of the chromosomal IL-2Rγ sequence was measured by analyzing the intensity of the band corresponding to the *Eco*RI-*Bsr*BI fragment (arrow next to autoradiograph; *Bsr*BI site indicated by filled triangle in the map above the autoradiograph).

[0355] The results, shown in Figure 37, indicate up to 15% of chromosomal IL-2Rγ sequences were altered by homologous recombination, thereby confirming the results obtained by PCR analysis that the targeted recombination event was stable through multiple rounds of cell division. The Southern blot results also indicate that the results shown in Figure 36 do not result from an amplification artifact.

Example 16: Targeted recombination at the IL-2Rγ locus in CD34-positive hematopoietic stem cells

[0356] Genetic diseases (*e.g.*, severe combined immune deficiency (SCID) and sickle cell anemia) can be treated by homologous recombination-mediated correction of the specific DNA sequence alteration responsible for the disease. In certain cases, maximal efficiency and stability of treatment would result from correction of the genetic defect in a pluripotent cell. To this end, this example demonstrates alteration of the sequence of the IL-2Rγ gene in human CD34-positive bone marrow cells. $CD34^+$ cells are pluripotential hematopoietic stem cells which give rise to the erythroid, myeloid and lymphoid lineages.

[0357] Bone marrow-derived human CD34 cells were purchased from AllCells, LLC and shipped as frozen stocks. These cells were thawed and allowed to stand for 2 hours at 37°C in an atmosphere of 5% CO_2 in RPMI Medium 1640 (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 2 mM L-glutamine. Cell samples ($1x10^6$ or $2x10^6$ cells) were transfected by Nucleofection™ (amaxa biosystems) using the Human CD34 Cell Nucleofector™ Kit, according to the manufacturers' protocol. After transfection, cells were cultured in RPMI Medium 1640 (Invitrogen), supplemented with 10% FBS, 2 mM L-glutamine, 100ng/ml granulocyte-colony stimulating factor (G-CSF), 100ng/ml stem cell factor (SCF), 100ng/ml thrombopoietin (TPO), 50ng/ml Flt3 Ligand, and 20ng/ml Interleukin-6 (IL-6). The caspase inhibitor zVAD-FMK (Sigma-Aldrich) was added to a final concentration of 40 μM in the growth medium immediately after transfection to block apoptosis. Additional caspase inhibitor was added 48 hours later to a final concentration of 20 μM to further prevent apoptosis. These cells were maintained at 37°C in an atmosphere of 5% CO_2 and were harvested 3 days post-transfection.

[0358] Cell numbers and DNAs used for transfection are shown in Table 21.

Table 21

Sample	# cells	p-eGFP-N1[1]	Donor[2]	p5-8G[3]	p5-9D[3]
1	$1x10^6$	5 μg	-	-	-
2	$2x10^6$	-	50 μg	-	-
3	$2x10^6$	-	50 μg	7.5 μg	7.5 μg

1. This is a control plasmid encoding an enhanced green fluorescent protein.
2. The donor DNA is a 1.5 kbp fragment containing sequences from exon 5 of the IL-2Rγ gene with an introduced *Bsr*BI site (see Example 14).
3. These are plasmids encoding *Fok*I fusions with the 5-8 G and 5-9D zinc finger DNA binding domains (see Table 17).

[0359] Genomic DNA was extracted from the cells using the MasterPure DNA Purification Kit (Epicentre). Due to the presence of glycogen in the precipitate, accurate quantitation of this DNA used as input in the PCR reaction is impossible; estimates using analysis of ethidium bromide-stained agarose gels indicate that ca. 50 ng genomic DNA was used in each sample. Thirty cycles of PCR were then performed using the following primers, each of which hybridizes to the chromosomal IL-2Rγ locus immediately outside of the region homologous to the 1.5 kb donor:

ex5_1.5detF3 GCTAAGGCCAAGAAAGTAGGGCTAAAG (SEQ ID NO:170)
ex5_1.5detR3 TTCCTTCCATCACCAAACCCTCTTG (SEQ ID NO:171)

[0360] Twenty microcuries each of α-^{32}PdCTP and α-^{32}PdATP were included in each PCR reaction to allow detection of PCR products. To provide an in-gel quantitation reference, the existence of a spontaneously occurring SNP in exon 5 of the IL-2Rgamma gene in Jurkat cells was exploited: this SNP creates a RFLP by destroying a MaeII site that is present in normal human DNA. A reference standard was therefore created by adding 1 or 10 nanograms of normal human genomic DNA (obtained from Clontech, Palo Alto, CA) to 100 or 90 ng of Jurkat genomic DNA, respectively, and performing the PCR as described above. The PCR reactions were desalted on a G-50 column (Amersham), and digested for 1 hour with restriction enzyme: experimental samples were digested with 10 units of *Bsr*BI (New England Biolabs); the "reference standard" reactions were digested with *Mae*II. The digestion products were resolved on a 10% non-denaturing PAGE (BioRad), the gel dried and analyzed by exposure to a PhosphorImager plate (Molecular Dynamics).

[0361] The results are shown in Figure 38. In addition to the major PCR product, corresponding to the 1.6 kb fragment of the IL2Rγ locus ("wt" in the right-hand panel of Figure 38), an additional band (labeled "rflp") was observed in lanes corresponding to samples from cells that were transfected with plasmids encoding both ZFP-nucleases and the donor DNA construct. This additional band did not appear in the control lanes, consistent with the idea that ZFP-nuclease assisted gene targeting of exon 5 of the common gamma chain gene occurred in this experiment.

[0362] Although accurate quantitation of the targeting rate is complicated by the proximity of the RFLP band to the wild-type band; the targeting frequency was estimated, by comparison to the reference standard (left panel), to be between 1-5%.

Example 17: Donor-target homology effects

[0363] The effect, on frequency of homologous recombination, of the degree of homology between donor DNA and the chromosomal sequence with which it recombines was examined in T18 cell line, described in Example 9. This line contains a chromosomally integrated defective eGFP gene, and the donor DNA contains sequence changes, with respect to the chromosomal gene, that correct the defect.

[0364] Accordingly, the donor sequence described in Example 10 was modified, by PCR mutagenesis, to generate a series of ~700 bp donor constructs with different degrees of non-homology to the target. All of the modified donors contained sequence changes that corrected the defect in the chromosomal eGFP gene and contained additional silent mutations (DNA mutations that do not change the sequence of the encoded protein) inserted into the coding region surrounding the cleavage site. These silent mutations were intended to prevent the binding to, and cleavage of, the donor sequence by the zinc finger-cleavage domain fusions, thereby reducing competition between the intended chromosomal target and the donor plasmid for binding by the chimeric nucleases. In addition, following homologous recombination, the ability of the chimeric nucleases to bind and re-cleave the newly-inserted chromosomal sequences (and possibly stimulating another round of recombination, or causing non-homologous end joining or other double-strand break-driven alterations of the genome) would be minimized.

[0365] Four different donor sequences were tested. Donor 1 contains 8 mismatches with respect to the chromosomal defective eGFP target sequence, Donor 2 has 10 mismatches, Donor 3 has 6 mismatches, and Donor 5 has 4 mismatches. Note that the sequence of donor 5 is identical to wild-type eGFP sequence, but contains 4 mismatches with respect to the defective chromosomal eGFP sequence in the T18 cell line. Table 22 provides the sequence of each donor between nucleotides 201-242. Nucleotides that are divergent from the sequence of the defective eGFP gene integrated into the genome of the T18 cell line are shown in bold and underlined. The corresponding sequences of the defective chromosomal eGFP gene (GFP mut) and the normal eGFP gene (GFP wt) are also shown.

Table 22

Donor	Sequence	SEQ ID NO.
Donor1	CTTCAGCCGCTATCCAGACCACATGAAACAACACGACTTCTT	172
Donor2	CTTCAGCCGGTATCCAGACCACATGAAACAACATGACTTCTT	173
Donor3	CTTCAGCCGCTACCCAGACCACATGAAACAGCACGACTTCTT	174
Donor 5	CTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT	175
GFP mut	CTTCAGCCGCTACCCCTAACAC--GAAGCAGCACGACTTCTT	176
GFP wt	CTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTT	177

[0366] The T18 cell line was transfected, as described in Example 11, with 50 ng of the 287-*Fok*I and 296-*Fok*I expression constructs (Example 7 and Table 12) and 500 ng of each donor construct. FACS analysis was conducted as described in Example 11.

[0367] The results, shown in Table 23, indicate that a decreasing degree of mismatch between donor and chromosomal target sequence (*i.e.*, increased homology) results in an increased frequency of homologous recombination as assessed by restoration of GFP function.

Table 23[1]

Donor	# mismatches	Percent cells with corrected eGFP gene[2]
Donor 2	10	0.45%
Donor 1	8	0.53%
Donor 3	6	0.89%
Donor 5	4	1.56%

1: T18 cells, containing a defective chromosomal eGFP gene, were transfected with plasmids encoding two ZFP nucleases and with donor plasmids encoding a nondefective eGFP sequence having different numbers of sequence mismatches with the chromosomal target sequence. Expression of the chromosomal eGFP gene was induced with doxycycline and FACS analysis was conducted 5 days after transfection.

2: The number is the percent of total fluorescence exhibiting high emission at 525 nm and low emission at 570 nm (region E of the FACS trace).

[0368] The foregoing results show that levels of homologous recombination are increased by decreasing the degree of target-donor sequence divergence. Without wishing to be bound by any particular theory or to propose a particular mechanism, it is noted that greater homology between donor and target could facilitate homologous recombination by increasing the efficiency by which the cellular homologous recombination machinery recognizes the donor molecule as a suitable template.

Alternatively, an increase in donor homology to the target could also lead to cleavage of the donor by the chimeric ZFP nucleases. A cleaved donor could help facilitate homologous recombination by increasing the rate of strand invasion or could aid in the recognition of the cleaved donor end as a homologous stretch of DNA during homology search by the homologous recombination machinery. Moreover, these possibilities are not mutually exclusive.

Example 18: Preparation of siRNA

[0369] To test whether decreasing the cellular levels of proteins involved in non-homologous end joining (NHEJ) facilitates targeted homologous recombination, an experiment in which levels of the Ku70 protein were decreased through siRNA inhibition was conducted. siRNA molecules targeted to the Ku70 gene were generated by transcription of Ku70 cDNA followed by cleavage of double-stranded transcript with Dicer enzyme.

[0370] Briefly, a cDNA pool generated from 293 and U2OS cells was used in five separate amplification reactions, each using a different set of amplification primers specific to the Ku70 gene, to generate five pools of cDNA fragments (pools A-E), ranging in size from 500-750 bp. Fragments in each of these five pools were then re-amplified using primers containing the bacteriophage T7 RNA polymerase promoter element, again using a different set of primers for each cDNA pool. cDNA generation and PCR reactions were performed using the Superscript Choice cDNA system and Platinum Taq High Fidelity Polymerase (both from Invitrogen, Carlsbad, CA), according to manufacturers protocols and recommendations.

[0371] Each of the amplified DNA pools was then transcribed *in vitro* with bacteriophage T7 RNA polymerase to generate five pools (A-E) of double stranded RNA (dsRNA), using the RNAMAXX *in vitro* transcription kit (Stratagene, San Diego, CA) according to the manufacturer's instructions. After precipitation with ethanol, the RNA in each of the pools was resuspended and cleaved *in vitro* using recombinant Dicer enzyme (Stratagene, San Diego, CA) according to the manufacturer's instructions. 21-23 bp siRNA products in each of the five pools were purified by a two-step method, first using a Microspin G-25 column (Amershan), followed by a Microcon YM-100 column (Amicon). Each pool of siRNA products was transiently transfected into the T7 cell line using Lipofectamone 2000®.

[0372] Western blots to assay the relative effectiveness of the siRNA pools in suppressing Ku70 expression were performed approximately 3 days post-transfection. Briefly, cells were lysed and disrupted using RIPA buffer (Santa Cruz Biotechnology), and homogenized by passing the lysates through a QIAshredder (Qiagen, Valencia, CA). The clarified lysates were then treated with SDS PAGE sample buffer (with β mercaptoethanol used as the reducing agent) and boiled for 5 minutes. Samples were then resolved on a 4-12% gradient NUPAGE gel and transferred onto a PVDF membrane. The upper portion of the blot was exposed to an anti-Ku70 antibody (Santa Cruz sc-5309) and the lower portion exposed to an anti-TF IIB antibody (Santa Cruz sc-225, used as an input control). The blot was then exposed to horseradish peroxidase-conjugated goat anti-mouse secondary antibody and processed for electrochemiluminescent (ECL) detection using a kit from Pierce Chemical Co. according to the manufacturer's instructions.

[0373] Figure 39 shows representative results following transfection of two of the siRNA pools (pools D and E) into T7 cells. Transfection with 70 ng of siRNA E results in a significant decrease in Ku70 protein levels (Figure 39, lane 3).

Example 19: Increasing the Frequency of Homologous Recombination by Inhibition of Expression of a Protein Involved in Non-Homologous End Joining

[0374] Repair of a double-stranded break in genomic DNA can proceed along two different cellular pathways; homologous recombination (HR) or non-homologous end joining (NHEJ). Ku70 is a protein involved in NHEJ, which binds to the free DNA ends resulting from a double-stranded break in genomic DNA. To test whether lowering the intracellular concentration of a protein involved in NHEJ increases the frequency of HR, small interfering RNAs (siRNAs), prepared as described in Example 18, were used to inhibit expression of Ku70 mRNA, thereby lowering levels of Ku70 protein, in cells co-transfected with donor DNA and with plasmids encoding chimeric nucleases.

[0375] For these experiments, the T7 cell line (see Example 9 and Figure 27) was used. These cells contain a chromosomally-integrated defective eGFP gene, but have been observed to exhibit lower levels of targeted homologous recombination than the T18 cell line used in Examples 11-13.

[0376] T7 cells were transfected, as described in Example 11, with either 70 or 140 ng of one of two pools of dicer product targeting Ku70 (*see* Example 18). Protein blot analysis was performed on extracts derived from the transfected cells to determine whether the treatment of cells with siRNA resulted in a decrease in the levels of the Ku70 protein (*see* previous Example). Figure 39 shows that levels of the Ku70 protein were reduced in cells that had been treated with 70 ng of siRNA from pool E.

[0377] Separate cell samples in the same experiment were co-transfected with 70 or 140 ng of siRNA (pool D or pool E) along with 50 ng each of the 287-FokI and 296-FokI expression constructs (Example 7 and Table 12) and 500 ng of the 1.5 kbp GFP donor (Example 13), to determine whether lowering Ku70 levels increased the frequency of homologous recombination. The experimental protocol is described in Table 24. Restoration of eGFP activity, due to homologous recombination, was assayed by FACS analysis as described in Example 11.

Table 24

Expt. #	Donor[1]	ZFNs[2]	SiRNA[3]	% correction[4]
1	500 ng	-	-	0.05
2	-	50 ng each	-	0.01
3	500 ng	50 ng each	-	0.79
4	500 ng	50 ng each	70 ng pool D	0.68
5	500 ng	50 ng each	140 ng pool D	0.59
6	500 ng	50 ng each	70 ng pool E	1.25
7	500 ng	50 ng each	140 ng pool E	0.92

1. A plasmid containing a 1.5 kbp sequence encoding a functional eGFP protein which is homologous to the chromosomally integrated defective eGFP gene

2. Plasmids encoding the eGFP-targeted 287 and 296 zinc finger protein/*Fok*I fusion endonucleases

3. See Example 18

4. Percent of total fluorescence exhibiting high emission at 525 nm and low emission at 570 nm (region E of the FACS trace, see Example 11).

[0378] The percent correction of the defective eGFP gene in the transfected T7 cells (indicative of the frequency of targeted homologous recombination) is shown in the right-most column of Table 24. The highest frequency of targeted recombination is observed in Experiment 6, in which cells were transfected with donor DNA, plasmids encoding the two eGFP-targeted fusion nucleases and 70 ng of siRNA Pool E. Reference to Example 18 and Figure 39 indicates that 70 ng of Pool E siRNA significantly depressed Ku70 protein levels. Thus, methods that reduce cellular levels of proteins involved in NHEJ can be used as a means of facilitating homologous recombination.

Example 20: Zinc finger-*Fok*I fusion nucleases targeted to the human β-globin gene

[0379] A number of four-finger zinc finger DNA binding domains, targeted to the human β–globin gene, were designed and plasmids encoding each zinc finger domain, fused to a *Fok*I cleavage half-domain, were constructed. Each zinc finger domain contained four zinc fingers and recognized a 12 bp target site in the region of the human β-globin gene encoding the mutation responsible for Sickle Cell Anemia. The binding affinity of each of these proteins to its target sequence was assessed, and four proteins exhibiting strong binding (sca-r29b, sca-36a, sca-36b, and sca-36c) were used for construction of *Fok*I fusion endonucleases.

[0380] The target sites of the ZFP DNA binding domains, aligned with the sequence of the human β-globin gene, are shown below. The translational start codon (ATG) is in bold and underlined, as is the A-T substitution causing Sickle Cell Anemia.

```
sca-36a     GAAGTCTGCCGT  (SEQ ID NO:178)
sca-36b     GAAGTCtGCCGTT  (SEQ ID NO:179)
sca-36c     GAAGTCtGCCGTT  (SEQ ID NO:180)
CAAACAGACACCATGGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTG
GTTTGTCTGTGGTACCACGTAGACTGAGGACACCTCTTCAGACGGCAATGAC  (SEQ ID NO:181)
sca-r29b     ACGTAGaCTGAGG  (SEQ ID NO:182)
```

[0381] Amino acid sequences of the recognition regions of the zinc fingers in these four proteins are shown in Table 25. The complete amino acid sequences of these zinc finger domains are shown in Figure 40. The sca-36a domain recognizes a target site having 12 contiguous nucleotides (shown in upper case above), while the other three domain recognize a thirteen nucleotide sequence consisting of two six-nucleotide target sites (shown in upper case) separated by a single nucleotide (shown in lower case). Accordingly, the sca-r29b, sca-36b and sca-36c domains contain a non-canonical inter-finger linker having the amino acid sequence TGGGGSQKP (SEQ ID NO:183) between the second and the third of their four fingers.

Table 25

ZFP	F1	F2	F3	F4
sca-r29b	QSGDLTR (SEQ ID NO:184)	TSANLSR (SEQ ID NO:185)	DRSALSR (SEQ ID NO:186)	QSGHLSR (SEQ ID NO:187)
sca-36a	RSQTRKT (SEQ ID NO:188)	QKRNRTK (SEQ ID NO:189)	DRSALSR (SEQ ID NO:190)	QSGNLAR (SEQ ID NO:191)
sca-36b	TSGSLSR (SEQ ID NO:192)	DRSDLSR (SEQ ID NO:193)	DRSALSR (SEQ ID NO:194)	QSGNLAR (SEQ ID NO:195)
sca-36c	TSSSLSR (SEQ ID NO:196)	DRSDLSR (SEQ ID NO:197)	DRSALSR (SEQ ID NO:198)	QSGNLAR (SEQ ID NO:199)

Example 21: *In vitro* cleavage of a DNA target sequence by β-globin-targeted ZFP/*Fok*I fusion endonucleases

[0382] Fusion proteins containing a *Fok*I cleavage half-domain and one the four ZFP DNA binding domains described in the previous example were tested for their ability to cleave DNA *in vitro* with the predicted sequence specificity. These ZFP domains were cloned into the pcDNA3.1 expression vector via KpnI and BamHI sites and fused in-frame to the *Fok*I cleavage domain *via* a 4 amino acid ZC linker, as described above. A DNA fragment containing 700 bp of the human β-globin gene was cloned from genomic DNA obtained from K562 cells. The isolation and sequence of this fragment was described in Example 3, *supra*.

[0383] To produce fusion endonucleases (ZFNs) for the *in vitro* assay, circular plasmids encoding *Fok*I fusions to sca-r29b, sca-36a, sca-36b, and sca-36c protein were incubated in an *in vitro* transcription/translation system. *See* Example 4. A total of 2 ul of the TNT reaction (2 ul of a single reaction when a single protein was being assayed or 1 ul of each reaction when a pair of proteins was being assayed) was added to 13 ul of the cleavage buffer mix and 3 ul of labeled probe (~1 ng/ul). The probe was end-labeled with ^{32}P using polynucleotide kinase. This reaction was incubated for 1 hour at room temperature to allow binding of the ZFNs. Cleavage was stimulated by the addition of 8 ul of 8 mM $MgCl_2$, diluted in cleavage buffer, to a final concentration of approximately 2.5 mM. The cleavage reaction was incubated for 1 hour at 37°C and stopped by the addition of 11 ul of phenol/chloroform. The DNA was isolated by phenol/chloroform extraction and analyzed by gel electrophoresis, as described in

Example 4. As a control, 3 ul of probe was analyzed on the gel to mark the migration of uncut DNA (labeled "U" in figure 41).

[0384] The results are shown in Figure 41. Incubation of the target DNA with any single zinc finger/*Fok*I fusion resulted in no change in size of the template DNA. However, the combination of the sca-r29b nuclease with either of the sca-36b or sca-36c nucleases resulted in cleavage of the target DNA, as evidenced by the presence of two shorter DNA fragments (rightmost two lanes of Figure 41).

Example 22: ZFP/*Fok*I fusion endonucleases, targeted to the β-globin gene, tested in a chromosomal GFP reporter system

[0385] A DNA fragment containing the human β-globin gene sequence targeted by the ZFNs described in Example 20 was synthesized and cloned into a *Spe*I site in an eGFP reporter gene thereby, disrupting eGFP expression. The fragment contained the following sequence, in which the nucleotide responsible for the sickle cell mutation is in bold and underlined):

CTAGACACCATGGTGCATCTGACTCCTG**<u>T</u>**GGAGAAGTCTGCCGTTACTGCCCTAG (SEQ ID NO:200)

[0386] This disrupted eGFP gene containing inserted β–globin sequences was cloned into pcDNA4/TO (Invitrogen, Carlsbad, CA) using the *Hind*III and *Not*I sites, and the resulting vector was transfected into HEK293 TRex cells (Invitrogen). Individual stable clones were isolated and grown up, and the clones were tested for targeted homologous recombination by transfecting each of the sca-36 proteins (sca-36a, sca-36b, sca-36c) paired with sca-29b (*See* Example 20 and Table 25 for sequences and binding sites of these chimeric nucleases). Cells were transfected with 50 ng of plasmid encoding each of the ZFNs and with 500 ng of the 1.5-kb GFP Donor (Example 13). Five days after transfection, cells were tested for homologous recombination at the inserted defective eGFP locus. Initially, cells were examined by fluorescence microscopy for eGFP function. Cells exhibiting fluorescence were then analyzed quantitatively using a FACS assay for eGFP fluorescence, as described in Example 11.

[0387] The results showed that all cell lines transfected with sca-29b and sca-36a were negative for eGFP function, when assayed by fluorescence microscopy. Some of the lines transfected with sca-29b paired with either sca-36b or sca-36c were positive for eGFP expression, when assayed by fluorescence microscopy, and were therefore further analyzed by FACS analysis. The results of FACS analysis of two of these lines are shown in Table 26, and indicate that zinc finger nucleases targeted to β-globin sequences are capable of catalyzing sequence-specific double-stranded DNA cleavage to facilitate homologous recombination in living cells.

Table 26

Cell line	DNA transfected: sca-29b	sca-36a	sca-36b	sca-36c	% corr.[1]
#20	+	+			0
	+		+		0.08
	+			+	0.07
#40	+	+			0
	+		+		0.18
	+			+	0.12

1. Percent of total fluorescence exhibiting high emission at 525 nm and low emission at 570 nm (region E of the FACS trace, see Example 11).

Example 23: Effect of transcription level on targeted homologous recombination

[0388] Since transcription of a chromosomal DNA sequence involves alterations in its chromatin structure (generally to make the transcribed sequences more accessible), it is possible that an actively transcribed gene might be a more favorable substrate for targeted homologous recombination. This idea was tested using the T18 cell line (Example 9) which contains chromosomal sequences encoding a defective eGFP gene whose transcription is under the control of a doxycycline-inducible promoter.

[0389] Separate samples of T18 cells were transfected with plasmids encoding the eGFP-targeted 287 and 296 zinc finger/*Fok*I fusion proteins (Example 7) and a 1.5 kbp donor DNA molecule containing sequences that correct the defect in the chromosomal eGFP gene (Example 9). Five hours after transfection, transfected cells were treated with different concentrations of doxycycline, then eGFP mRNA levels were measured 48 hours after addition of doxycycline. eGFP fluorescence at 520 nm indicative of targeted recombination of the donor sequence into the chromosome to replace the inserted β–globin sequences) was measured by FACS at 4 days after transfection.

[0390] The results are shown in Figure 42. Increasing steady-state levels of eGFP mRNA normalized to GAPDH mRNA (equivalent, to a first approximation, to the rate of transcription of the defective chromosomal eGFP gene) are indicated by the bars. The number above each bar indicate the percent of cells exhibiting eGFP fluorescence. The results show that increasing transcription rate of the target gene is accompanied by higher frequencies of targeted recombination. This suggests that targeted activation of transcription (as disclosed, *e.g.* in co-owned U.S. Patents 6,534,261 and 6,607,882) can be used, in conjunction with targeted DNA cleavage, to stimulate targeted homologous recombination in cells.

Example 24: Generation of a cell line containing a mutation in the IL-2Rγ gene

[0391] K562 cells were transfected with plasmids encoding the 5-8GL0 and the 5-9DL0 zinc finger nucleases (ZFNs) (see Example 14; Table 17) and with a 1.5 kbp *Dra*I donor construct. The *Dra*I donor is comprised of a sequence with homology to the region encoding the 5th exon of the IL2Rγ gene, but inserts an extra base between the ZFN-binding sites to create a frameshift and generate a *Dra*I site.

[0392] 24 hours post-transfection, cells were treated with 0.2 μM vinblastine (final concentration) for 30 hours. Cells were washed three times with PBS and re-plated in medium. Cells were allowed to recover for 3 days and an aliquot of cells were removed to perform a PCR-based RFLP assay, similar to that described in Example 14, testing for the presence of a *Dra*I site. It was determined the gene correction frequency within the population was approximately 4%.

[0393] Cells were allowed to recover for an additional 2 days and 1600 individual cells were plated into 40x 96-well plates in 100 ul of medium.

[0394] The cells are grown for about 3 weeks, and cells homozygous for the *Dra*I mutant phenotype are isolated. The cells are tested for genome modification (by testing for the presence of a *Dra*I site in exon 5 of the IL-2Rγ gene) and for levels of IL-2Rγ mRNA (by real-time PCR) and protein (by Western blotting) to determine the effect of the mutation on gene expression. Cells are tested for function by FACS analysis.

[0395] Cells containing the *Dra*I frameshift mutation in the IL-2Rγ gene are transfected with plasmids encoding the 5-8GL0 and 5-9DL0 fusion proteins and a 1.5 kb *Bsr*BI donor construct (Example 14) to replace the *Dra*I frameshift mutation with a sequence encoding a functional protein. Levels of homologous recombination greater than 1% are obtained in these cells, as measured by assaying for the presence of a *Bsr*BI site as described in Example 14. Recovery of gene function is demonstrated by measuring mRNA and protein levels and by FACS analysis.

Example 25: ZFP/*Fok*I fusion endonucleases with different polarities

[0396] A vector encoding a ZFP/*Fok*I fusion, in which the ZFP domain was N-terminal to the *Fok*I domain, was constructed. The ZFP domain, denoted IL2-1, contained four zinc fingers, and was targeted to the sequence AACTCGGATAAT (SEQ ID NO:202), located in the third exon of the IL-2Rγ gene. The amino acid sequences of the recognition regions of the zinc fingers are given in Table 27.

Table 27: Zinc Finger Design of IL2-1 binding domain

Target sequence	F1 (AAT)	F2 (GAT)	F3 (TCG)	F4 (AAC)
AACTCGGATAAT (SEQ ID NO:203)	DRSTLIE (SEQ ID NO:204)	SSSNSLR (SEQ ID NO:205)	RSDDLSK (SEQ ID NO:206)	DNSNRIK (SEQ ID NO:207)

Note: The DNA target sequence is shown in the left-most column. The remaining columns show the amino acid sequences (in one-letter code) of residues -1 through +6 of each of the four zinc fingers, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein. The three-nucleotide subsite bound by each finger is shown in the top row adjacent to the finger designation.

[0397] Sequences encoding this zinc finger domain were joined to sequences encoding the cleavage half-domain of the *Fok*I restriction endonuclease (amino acids 384-579 according to Looney *et al.* (1989) *Gene* **80:**193-208) such that a four amino acid linker was present between the ZFP domain and the cleavage half-domain (*i.e.*, a four amino acid ZC linker). The *Fok*I cleavage half-domain was obtained by PCR amplification of genomic DNA isolated from the bacterial strain *Planomicrobium okeanokoites* (ATCC 33414) using the following primers:

5'-<u>GGATCC</u>CAACTAGTCAAAAGTGAAC (SEQ ID NO: 208)

5'-<u>CTCGAG</u>TTAAAAGTTTATCTCGCCG (SEQ ID NO: 209).

[0398] The PCR product was digested with *Bam*HI and *Xho*I (sites underlined in sequences shown above) and then ligated with a vector fragment prepared from the plasmid pcDNA-nls-ZFP1656-VP16-flag after *Bam*HI and *Xho*I digestion. The resulting construct, pcDNA-nls-ZFP1656-FokI, encodes a fusion protein containing, from N-terminus to C-terminus, a SV40 large T antigen-derived nuclear localization signal (NLS, Kalderon *et al.* (1984) *Cell* **39**:499-509), ZFP1656, and a *Fok*I cleavage half-domain, in a pcDNA3.1 (Invitrogen, Carlsbad, CA) vector backbone. This construct was digested with *Kpn*I and *Bam*HI to release the ZFP1656–encoding sequences, and a *Kpn*I/*Bam*HI fragment encoding the IL2-1 zinc finger binding domain was inserted by ligation. The resulting construct (pIL2-1C) encodes a fusion protein comprising, from N- to C-terminus, a nuclear localization signal, the four-finger IL2-1 zinc finger binding domain and a *Fok*I cleavage half-domain, with a four amino acid ZC linker.

[0399] A vector encoding a ZFP/*Fok*I fusion protein, in which the *Fok*I sequences were N-terminal to the ZFP sequences, was also constructed. The IL2-1 four-finger zinc finger domain was inserted, as a *Kpn*I/*Bam*HI fragment, into a vector encoding a fusion protein containing a NLS, the KOX-1 repression domain, EGFP and a FLAG epitope tag, that had been digested with *Kpn*I and *Bam*HI to release the EGFP-encoding sequences. This generated a vector containing sequences encoding, from N-terminus to C-terminus, a NLS (from the SV40 large T-Antigen), a KOX repression domain, the IL2-1 zinc finger domain and a FLAG epitope tag. This construct was then digested with *Eco*RI and *Kpn*I to release the NLS- and KOX-encoding sequences, and an *Eco*RI/*Kpn*I fragment (generated by PCR using, as template, a vector encoding *Fok*I) encoding amino acids 384-579 of the *Fok*I restriction enzyme and a NLS was inserted. The resulting construct, pIL2-1R encodes a fusion protein containing, from N-terminus to C-terminus, a *Fok*I cleavage half-domain, a NLS, and the four-finger IL2-1 ZFP binding domain. The ZC linker in this construct is 21 amino acids long and includes the seven amino acid nuclear localization sequence (PKKKRKV; SEQ ID NO: 210).

[0400] The 5-9D zinc finger domain binds the 12-nucleotide target sequence AAAGCGGCTCCG (SEQ ID NO:157) located in the fifth exon of the IL-2Rγ gene. *See* Example 14 (Table 17). Sequences encoding the 5-9D zinc finger domain were inserted into a vector to generate a *Fok*I/ZFP fusion, in which the *Fok*I sequences were N-terminal to the ZFP sequences. To make this construct, the pIL2-1R plasmid described in the previous paragraph was digested with *Kpn*I and *Bam*HI to release a fragment containing sequences encoding the IL2-1 zinc finger binding domain, and a

KpnI/BamHI fragment encoding the 5-9D zinc finger binding domain was inserted in its place. The resulting construct, p5-9DR, encodes a fusion protein containing, from N-terminus to C-terminus, a *FokI* cleavage half-domain, a NLS, and the four-finger 5-9D zinc finger binding domain. The ZC linker in this construct is 22 amino acids long and includes the seven amino acid nuclear localization sequence (PKKKRKV; SEQ ID NO: 210).

[0401] *See* co-owned U.S. Patents 6,453,242 and 6,534,261 for additional details of vector construction.

Example 26: Construction of synthetic substrates for DNA cleavage

[0402] The target sequences bound by the IL2-1 and 5-9D fusion proteins described above were introduced into double-stranded DNA fragments in a variety of orientations, to test the cleavage ability of zinc finger/*FokI* fusion proteins having an altered polarity in which the *FokI* domain is N-terminal to the ZFP domain. In template 1, the 5-9D target site is present in one strand and the IL2-1 target site is present on the complementary strand, with the 3'ends of the binding sites being proximal to each other and separated by six intervening nucleotide pairs. In template 2, the 5-9D and IL2-1 target sites are present on the same DNA strand, with the 3' end of the 5-9D binding site separated by six nucleotide pairs from the 5' end of the IL2-1 binding site.

[0403] DNA fragments of approximately 442 base pairs, containing the sequences described above, were obtained as amplification products of plasmids into which the templates had been cloned. The IL2-1 and 5-9D target sites were located within these fragments such that double-stranded DNA cleavage between the two target sites would generate DNA fragments of approximately 278 and 164 base pairs. Amplification products were radioactively labeled by transfer of orthophosphate from γ-32P-ATP using T4 polynucleotide kinase.

Example 27: Targeted DNA Cleavage with zinc finger/FokI fusions having altered polarity

[0404] The IL2-1C, IL2-1R and 5-9DR fusion proteins were obtained by incubating plasmids encoding these proteins in a TNT coupled reticulocyte lysate (Promega, Madison, WI). Cleavage reactions were conducted in 23 μl of a mixture containing 1 μl of TNT reaction for each fusion protein, 1 μl labeled digestion substrate and 20 μl cleavage buffer. Cleavage buffer was prepared by adding 1 μl of 1M dithiothreitol and 50 μl of bovine serum albumin (10 mg/ml) to 1 ml of 20 mM Tris-Cl, pH 8.5, 75 mM NaCl, 10 μM $ZnCl_2$, 5% (v/v) glycerol. Cleavage reactions were incubated at 37°C for 2 hours, then shaken with 13 μl phenol/chloroform/isoamyl alcohol (25:24:1). After centrifugation, 10 μl of the aqueous phase was analyzed on a 10% polyacrylamide gel. Radioactivity in the gel was detected using a Phosphorimager (Molecular Dynamics) and quantitated using ImageQuant software (Molecular Dynamics).

[0405] Figure 44 shows the results obtained using two chimeric nucleases having a NH_2-*Fok*I domain-zinc finger domain-COOH polarity to cleave a substrate in which the binding sites for the two chimeric nucleases are located on opposite strands and the 3'ends of the binding sites are proximal to each other and separated by six nucleotide pairs. Incubation of the substrate with either of the IL2-1R or 5-9DR nucleases alone does not result in cleavage of the substrate (compare lanes 2 and 3 with lane 1), while incubation of both nucleases results in almost complete cleavage of the DNA substrate at the intended target site (lane 4).

[0406] Figure 45 shows the ability of a first chimeric nuclease having a NH_2-zinc finger domain-*Fok*I domain-COOH polarity, and a second chimeric nuclease having a NH_2-*Fok*I domain-zinc finger domain-COOH polarity, to cleave a substrate in which the binding sites for the two chimeric nucleases are located on the same strand, and the 3' end of the first binding site is proximal to the 5' end of the second binding site and separated from it by six nucleotide pairs. Only the combination of the 5-9DR and the IL2-1C nucleases (*i.e.* each nuclease having a different polarity) was successful in cleaving the substrate having both target sites on the same strand (compare lane 6 with lanes 1-5).

Example 28: Chimeric nucleases with different ZC linker lengths

[0407] Two sets of fusion proteins with different ZC linker lengths, in which the *Fok*I domain is amino terminal to the ZFP domain, were designed. The *Fok*I domain is amino acids 384-579 according to Looney *et al.* (1989) *Gene* **80**:193-208. The ZFP domain was selected from the IL1-2 (Table 27), 5-8G (Table 17) and 5-9D (Table 17) domains. The first set had the structure NH_2-NLS-*Fok*I-ZFP-Flag-COOH. In this set, proteins having ZC linker lengths of 13, 14, 18, 19, 28 and 29 amino acids were designed. The second set had the structure NH_2-*Fok*I-NLS-ZFP-Flag-COOH and proteins with ZC linkers of 21, 22, 23, 24, 28, 29, 38 and 39 amino acids were designed. Note that, in the second set, the NLS is part of the ZC linker. Plasmids encoding these fusion proteins are also constructed.

[0408] Model DNA sequences were designed to test the cleavage activity of these fusion proteins and to determine optimal ZC linker lengths as a function of distance between the target sites for the two fusion proteins. The following sequences were designed:

1. 5-9D target site and IL2-1 target site on opposite strands
2. 5-9D target site and IL2-1 target site on same strand
3. 5-9D target site and 5-8G target site on opposite strands
4. 5-9D target site and 5-8G target site on same strand

[0409] For each of these four pairs of target sites, sequences are constructed in which the separation between the two target sites is 4, 5, 6 or 7 base pairs.

[0410] These sequences are introduced into labeled substrates as described in Example 26 and are used to test the various fusion proteins described in this example for their ability to cleave DNA, according to the methods described in Example 27.

Example 29: Construction of a stable cell line containing an integrated defective eGFP reporter gene

[0411] An eGFP (enhanced green fluorescent protein) coding sequence containing a frameshift mutation and a fragment of exon 5 of the IL-2Rγ gene, operatively liked to a tetracycline-regulated CMV promoter, was constructed as follows. A silent mutation was inserted into the eGFP coding sequences in the pEGFP-NI vector (BD BioSciences) to create a novel SpeI site. Subsequently a one-nucleotide deletion (creating a frameshift mutation) was introduced downstream of the new SpeI site. The following sequence from exon 5 of the IL-2Rγ gene, containing target sites for the 5-8G and 5-9D zinc finger/*Fok*I fusion proteins (described in Example 14, Table 17, *supra*), was inserted into the newly-introduced SpeI site:

CTAGCTACACGTTTCGTGTTCGGAGCCGCTTTAACCCACTCTGTGGA AGTGCTCCTAG (SEQ ID NO:214)

[0412] The resulting plasmid contained sequences encoding mutant eGFP containing a fragment of DNA sequence from exon 5 of the IL2Rγ gene. This plasmid was digested with *Hind*III and *Not*I, releasing a fragment containing the mutated eGFP sequence (including the inserted IL-2Rγ exon 5 sequences). This fragment was inserted into the *Hind*III and *Not*I sites of the pcDNA4/TO vector (Invitrogen), resulting in a construct in which expression of eGFP sequence is controlled by a 2X tet-operator-regulated CMV promoter. A schematic diagram of this plasmid is shown in Figure 46.

[0413] This construct was used to transform HEK293 TRex cells (Invitrogen), and a stable cell line containing an integrated copy of this construct was isolated. In this cell line, the eGFP coding sequences are transcribed upon addition of doxycycline, but because of the frameshift mutation and the IL-2Rγ insertion, no functional protein is expressed.

Example 30: Targeted homology-dependent integration of a puromycin resistance marker into a chromosomal eGFP gene

[0414] An experiment was conducted to test for integration of a puromycin resistance marker into the mutant chromosomal eGFP gene described in the preceding Example.

[0415] A promoterless donor was constructed that contained sequences encoding puromycin resistance (denoted "puro sequences"), flanked by sequences homologous to the eGFP cDNA construct, as follows. Sequences were PCR amplified from the pTRE2pur-HA vector (BD Biosciences) to generate a puro sequence with flanking SpeI sites and a consensus Kozak sequence upstream of the ATG initiation codon. Amplification primers were:

puro-5': ACTAGTGCCGCCACCATGACCGAGTACAAGCCCA (SEQ ID NO:215)
puro-3': ACTAGTCAGGCACCGGGCTT (SEQ ID NO:216)

This PCR fragment was cloned into the pEGFP-N1 vector containing a modified eGFP gene that encoded a novel SpeI restriction site and a frameshift mutation that prevented functional expression of the gene (see Example 29). This eGFP/Puromycin gene was cloned into the pcDNA4/TO vector, via HindIII and NotI sites, to create the vector pcDNA4/TO/GFPpuro, which also served as the positive control in experiments to obtain Puromycin resistant cells by targeted integration. In order to create a promoterless donor, the pcDNA4/TO/GFPpuro vector was PCR amplified with the following primers:

GFP-Bam	CGAATTCTGCAGTCGAC	(SEQ ID NO:217)
pcDNA42571	TGCATACTTCTGCCTGC	(SEQ ID NO:218)

[0416] The resulting amplification product was Topo cloned into the pCR4-TOPO vector and its sequence was confirmed. This created a donor with 413 bp of sequence homologous to the chromosomal eGFP construct upstream of the puro sequences and 1285 bp of sequence homologous to the chromosomal eGFP construct downstream of the puro sequences.

[0417] To test for targeted integration of puro sequences, the cell line described in Example 29 was subjected to targeted DNA cleavage by zinc finger/FokI fusion proteins in the presence of the donor construct described in this example, transfected cells were selected for puromycin resistance and their chromosomal DNA was analyzed. The two zinc finger/*Fok*I fusion proteins (ZFNs), designed to cleave within the exon 5 sequences inserted into the eGFP gene (5-8G and 5-9D) have been described in Example 14, Table 17, *supra*. Puromycin resistance can arise from either homology-dependent or homology-independent integration of donor sequences at the cleavage site located within the IL-2Rγ sequences inserted into the eGFP coding sequences. Homology-dependent integration of the donor construct will result in replacement of IL-2Rγ sequences by the puro sequences.

[0418] HEK 293 cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 2 mM L-glutamine and maintained at 37°C in an atmosphere of 5% CO_2. To test for targeted integration of the puro sequences, cells were transfected with 50 ng of each ZFN-encoding plasmid and 500 ng of donor plasmid. In negative control experiments, cells were transfected either with 50 ng of each ZFN-encoding plasmid or 500 ng of donor plasmid. As a positive control, 500 ng of the pcDNA4/TO/GFPpuro vector was transfected into HEK293 cells. Cells were transfected using LipofectAMINE 2000 Reagent (Invitrogen) in Opti-MEM I reduced serum medium. Puromycin resistance was assayed by the addition of doxycycline to 2 ng/ml (to activate transcription of the integrated sequences) and puromycin to 2ug/ml (final concentration) in the growth medium.

[0419] Puromycin resistant colonies were obtained only from cells that had been transfected with both ZFN-encoding plasmids and the donor plasmid. Twenty-four clonal populations were isolated, subjected to >6-weeks of selection, then analyzed by PCR for a targeted integration event. The following PCR primers were used to detect whether a targeted integration event occurred:

CMVPuro-5'	TTTGACCTCCATAGAAGACA	(SEQ ID NO:219)
CMVPuro-3'	GCGCACCGTGGGCTTGTACT	(SEQ ID NO:220)

[0420] One of the primers is complementary to the exogenous puro sequences and the other is complementary to sequences in the CMV promoter present in the integrated reporter construct. Twenty-one out of 24 colonies yielded amplification products whose sizes were consistent with targeted integration of puro sequences. These fragments were cloned and their nucleotide sequences were determined. Sequence analysis indicated that eight out of the 24 clones had undergone homology-directed integration of puro sequences into the chromosomal eGFP construct, while 13 had undergone homology-independent integration of donor DNA into chromosomal sequences, accompanied by partial duplication of the puro sequences.

Example 31: Codon Optimization of zinc finger/*Fok*I fusion proteins targeted to exon 5 of the IL-2Rγ gene

[0421] Fusion proteins containing the 5-8G and 5-9D zinc finger binding domains (Table 17) joined to a *Fok*I cleavage half-domain by a 4 amino acid ZC linker (L0) have been described *supra*. *See, e.g.,* Example 14 and Example 24. Polynucleotides encoding these two fusion proteins were designed so that the codons were optimized for expression in mammalian cells. The codon-optimized nucleotide sequences encoding these two fusion proteins are as follows:

```
5-8G L0 FokI
aattcgctagcgccaccatggcccccaagaagaagaggaaagtgggaatccacggggtacccgccgcta
tggccgagaggcccttccagtgtcggatctgcatgcggaacttcagccggagcgacaacctgagcgtgc
acatccgcacccacacaggcgagaagccttttgcctgtgacatttgtgggaggaaatttgcccgcaacg
cccaccgcatcaaccacaccaagatccacaccggatctcagaagccctttcagtgcagaatctgcatga
gaaacttctcccggtccgacaccctgagcgaacacatcaggacacacaccggcgagaaacccttcgcct
gcgacatctgtggccgcaagtttgccgccagaagcacccgcacaaatcacacaaagattcacctgcggg
gatcccagctggtgaagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgc
cccacgagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatgaagg
tgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagcagaaagcctgacg
```

gcgccatctatacagtgggcagccccatcgattacggcgtgatcgtggacacaaaggcctacagcggcg
gctacaatctgcctatcggccaggccgacgagatgcagagatacgtggaggagaaccagacccggaata
agcacatcaaccccaacgagtggtggaaggtgtaccctagcagcgtgaccgagttcaagttcctgttcg
tgagcggccacttcaagggcaactacaaggcccagctgaccaggctgaaccacatcaccaactgcaatg
gcgccgtgctgagcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactgg
aggaggtgcggcgcaagttcaacaacggcgagatcaacttctgataac (SEQ ID NO:221)

5-9D L0 FokI
aattcgctagcgccaccatggccccccaagaagaagaggaaagtgggaatccacggggtacccgccgcta
tggccgagaggcccttccagtgtcggatctgcatgcggaacttcagcaggagcgacaccctgagcgaac
acatccgcacccacacaggcgagaagccttttgcctgtgacatttgtgggaggaaatttgccgccagaa
gcacccgcacaacccacaccaagatccacaccggatctcagaagccctttcagtgcagaatctgcatga
gaaacttctcccggtccgacagcctgagcaagcacattaggacccacaccggggagaaacccttcgcct
gcgacatctgtggccgcaaatttgcccagcgcagcaacctgaaagtgcacacaaagattcacctgcggg
gatcccagctggtgaagagcgagctggaggagaagaagtccgagctgcggcacaagctgaagtacgtgc
cccacgagtacatcgagctgatcgagatcgccaggaacagcacccaggaccgcatcctggagatgaagg
tgatggagttcttcatgaaggtgtacggctacaggggaaagcacctgggcggaagcagaaagcctgacg
gcgccatctatacagtgggcagccccatcgattacggcgtgatcgtggacacaaaggcctacagcggcg
gctacaatctgcctatcggccaggccgacgagatgcagagatacgtggaggagaaccagacccggaata
agcacatcaaccccaacgagtggtggaaggtgtaccctagcagcgtgaccgagttcaagttcctgttcg
tgagcggccacttcaagggcaactacaaggcccagctgaccaggctgaaccacatcaccaactgcaatg
gcgccgtgctgagcgtggaggagctgctgatcggcggcgagatgatcaaagccggcaccctgacactgg
aggaggtgcggcgcaagttcaacaacggcgagatcaacttctgataac (SEQ ID NO:222)

Example 32: Growth and transfection of K-562 cells for targeted homologous integration

[0422] Human K-562 erythroleukemia cells (ATCC) were cultured at 37°C in DMEM supplemented with 10% fetal bovine serum, penicillin and streptomycin, and transfected (Nucleofector; Amaxa) with 2.5 µg of an expression vector encoding two zinc finger nucleases (ZFN) designed to introduce a double-strand break at a position surrounding the codon for arginine 226 in the endogenous IL2Rγ gene. The two nucleases (5-8G and 5-9D) have been described in Example 14, Table 17, *supra*. Sequences encoding the nucleases were separated by sequences encoding a 2A peptide. *See, e.g.*, Szymczak *et al.* (2004) *Nature Biotechnol.* **22**:589-594. At the same time, the cells were transfected with either 25 or 50 µg of a donor DNA plasmid carrying a 1.5 kb DNA stretch of IL2Rγ chromosomal DNA sequence centered on exon 5 (Urnov *et al.* (2005) *Nature* **435**:646-651), interrupted by the DNA sequence to be inserted (see examples 33-36 below). Seventy two hours after transfection, genomic DNA was isolated (DNEasy; Qiagen) and cell genotype at the IL2Rγ locus was determined by PCR of the exon 5-containing stretch of the X chromosome, using primers that anneal outside of the 1.5 kb region of donor homology and generate a 1.6 kilobase pair amplification product from the wild-type IL2Rγ sequence (Urnov *et al.*, *supra*). PCR products were analyzed by gel electrophoresis and, where indicated, by restriction digestion. Control samples included: (1) cells transfected with a GFP-encoding expression vector, (2) cells transfected solely with the expression vector encoding the ZFNs, and (3) cells transfected solely with the donor DNA molecule.

Example 33: Targeted homology-dependent integration of a 12-nucleotide exogenous sequence into the endogenous IL-2Rγ gene

[0423] Cell growth and transfection were conducted as described in Example 32. The donor DNA molecule was engineered to contain a 12 nucleotide pair sequence tag containing a novel diagnostic recognition site for the restriction enzyme *Stu*I. Cellular DNA was isolated and used as a template for amplification as described in Example 32, then digested with *Stu*I. As shown in Figure 47, all control samples carried chromosomes yielding amplification products that were resistant to cleavage by the restriction enzyme. In contrast, 15% of all amplification products in the cell sample transfected both with the donor DNA molecule and the ZFN expression construct were sensitive to the restriction enzyme, indicating integration of the donor DNA. Direct nucleotide sequence determination of the chromosome-derived PCR product confirmed that integration was homology-dependent.

Example 34: Targeted homology-dependent integration of exogenous open reading frames into the endogenous IL-2Rγ gene

[0424] Cell growth and transfection were conducted as described in Example 32. In this experiment, two different donor DNA molecules were used. Donor DNA molecule #1 was engineered to contain the entire 720 bp ORF of enhanced green fluorescent protein (eGFP) flanked by sequences homologous to the chromosomal IL2Rγ locus (see Example 32). Donor DNA molecule #2 contained a 924 bp sequence consisting of the entire eGFP ORF followed by a polyadenylation signal; this sequence was flanked by IL2Rγ-homologous sequences (see Example 32). Following transfection, cellular DNA was isolated and used as a template for amplification as described in Example 32. As shown in Figure 48, all control samples carried chromosomes yielding PCR products of wild-type size (~1.6 kb). In contrast, 3-6% of all chromosomes in the cell samples transfected both with the ORF-carrying donors and the ZFN expression construct yielded amplification products that were larger than the wild-type-chromosome-derived PCR product, and the difference in size was consistent with the notion that ZFN-driven targeted integration of the eGFP ORFs had occurred. Direct nucleotide sequence determination of the chromosome-derived PCR products confirmed this observation and also indicated that integration was homology-dependent.

Example 35: Targeted homology-dependent integration of an exogenous "therapeutic half-gene" into the endogenous IL-2Rγ gene

[0425] Cell growth and transfection were conducted as described in Example 32. The donor DNA molecule consisted of a 720 nucleotide-pair partial IL2Rγ cDNA containing the downstream portion of exon 5, and complete copies of exons 6, 7 and 8, (including a translation termination codon and a polyadenylation signal within exon 8). These cDNA sequences were flanked, on one side, by sequences homologous to the upstream portion of exon 5 and the adjoining portion of intron 4 and, on the other side, by sequences homologous to the downstream portion of exon 5 and the adjoining portion of intron 5 (see Figure 49). Because two copies of the downstream portion of exon 5 are present in the donor construct, and to ensure that recombination occurred in the copy adjacent to exon 8, several silent sequence changes were introduced into the copy adjacent to exon 6. These changes did not alter the coding potential of the exogenous exon 5 sequences, but introduced sufficient non-homology with chromosomal sequences to prevent use of these sequences for the initial homing event in the repair of the break. Thus, integration of the donor construct at the site targeted by the nucleases can be used to correct any IL2Rγ mutation in exons 6, 7 or 8 and in the downstream portion of exon 5 contained in the donor construct.

[0426] Following transfection, cellular DNA was isolated and used as a template for amplification as described in Example 32. As shown in Figure 50, a control sample in which cells were transfected with a GFP-encoding plasmid contained chromosomes yielding PCR products of wild-type size only (~1.6 kb). In contrast, 6% of all chromosomes in the cell samples transfected both with the therapeutic half-gene-carrying donor and the ZFN expression construct were larger than the wild-type-chromosome-derived PCR product, and the difference in size was consistent with the notion that ZFN-driven targeted integration of the "therapeutic half-gene" had occurred. Direct nucleotide sequence determination of the larger PCR product confirmed that homology-dependent integration of the donor construct had occurred.

Example 36: Targeted homology-dependent integration of an exogenous 7.7 kilobase pair expression construct into the endogenous IL-2Rγ gene

[0427] Cell growth and transfection were conducted as described in Example 32. A donor DNA molecule was constructed that contained a 7.7 kbp antibody expression construct flanked by sequences homologous to IL2Rγ exon 5 and adjacent sequences (See Example 32). In this experiment, two topological forms of the donor were used: a plasmid donor, in which the vector backbone abuts an insert with two homology arms interrupted by the expression construct ("circular"); and a linear donor, which contains two homology arms interrupted by the expression construct ("linear").

[0428] DNA was isolated from transfected cells as described in Example 32. The DNA was analyzed by PCR, using two primer pairs designed to detect the junction between integrated exogenous sequences and endogenous IL-2Rγ exon 5 sequences. Thus, for each primer pair, one of the primers was complementary to endogenous exon 5 sequence, and the other primer was complementary to the expression construct (see upper portion of Figure 51). As shown in the lower portion of Figure 51, no PCR product was observed in control samples transfected only with donor DNA. In contrast, PCR products of the expected size were observed in cell samples transfected with donor DNA (either linear or circular) and the ZFN-encoding plasmid. Critically, primer sets specific for both ends of the expression construct yielded identical results, consistent with the notion that ZFN-driven targeted integration of the exogenous 7.7 kb sequence has occurred. Nucleotide sequence determination of the amplification products confirmed that homology-dependent integration had occurred.

Example 37: Targeted, homology-independent integration of exogenous sequences at an endogenous chromosomal locus

[0429] A pair of zinc finger/*Fok*I fusion proteins were constructed to bind to two target sites, separated by six nucleotide pairs, in the Chinese hamster dihydrofolate reductase (DHFR) gene and cleave the gene between the two target sites. The nucleotide sequences of the target sites, and the amino acid sequences of the recognition regions of the fusion proteins, are shown in Table 28.

Table 28: Zinc Finger Designs for the CHO DHFR gene

Target sequence	F1	F2	F3	F4
GGAAGGTCTCCG (SEQ ID NO:223)	RSDTLSE (SEQ ID NO:224)	NNRDRTK (SEQ ID NO:225)	RSDHLSA (SEQ ID NO:226)	QSGHLSR (SEQ ID NO:227)
AATGCTCAGGTA (SEQ ID NO:228)	QSGALAR (SEQ ID NO:229)	RSDNLRE (SEQ ID NO:230)	QSSDLSR (SEQ ID NO:231)	TSSNRKT (SEQ ID NO:232)

Note: The DNA target sequence is shown in the left-most column. The remaining columns show the amino acid sequences (in one-letter code) of residues -1 through +6 of each of the four zinc fingers, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein.

[0430] Chinese hamster ovary (CHO) cells were cultured in adherent medium (DMEM + 10% FBS supplemented with 2 mM L-glutamine plus non-essential amino acids) at 37°C. $3x10^5$ cells were grown to 70% confluence in 12-well plates and transiently transfected (using Lipofectamine 2000®) with 100 ng each of the two fusion protein-encoding plasmids. 24 hours after transfection, 20 μM vinblastine was added to the growth medium. Medium was replaced 24 hours after addition of vinblastine. 24 hours after replacement of medium, cellular DNA was purified (Qiagen) and DHFR gene sequences surrounding the target sites were amplified by PCR. Primers were designed such that DNA from cells containing a wild-type DHFR gene were expected to yield a 383 nucleotide pair amplification product. Unexpectedly, two amplification products were obtained, one of the expected size and another approximately 150 nucleotide pairs larger.

[0431] To determine if mutations had been induced at the cleavage site, the amplification product was analyzed using a Cel-1 assay, in which the amplification product is denatured and renatured, followed by treatment with the mismatch-specific Cel-1 nuclease. *See*, for example, Oleykowski *et al.* (1998) *Nucleic Acids res.* **26:**4597-4602; Qui *et al.* (2004) *BioTechniques* **36:**702-707; Yeung *et al.* (2005) *BioTechniques* **38:**749-758. The results of the Cel-1 assay (Figure 52) showed that, in addition to small mismatches in the reannealed products resulting from non-homologous end-joining at the cleavage site (indicated by the presence of two low molecular weight bands in rightmost lane of Figure 52), a larger insertion had also occurred (indicated by the presence of a high molecular weight band, identified as "Mutant," in lanes 3 and 5 of Figure 52). This corroborated the observation of a larger amplification product described above.

[0432] The nucleotide sequences of the two amplification products described above were determined, to characterize the nature of the insertion, and are shown in Figure 53. The sequence shown on the top line (SEQ ID NO:233) is the wild-type DHFR sequence, while the sequence shown on the bottom line (SEQ ID NO:234) consists of the DHFR sequence containing, at the cleavage site, an insertion of 157 base pairs and a deletion of a single nucleotide pair. Further analysis revealed that the inserted 157 base pairs correspond to a portion of the vector plasmid encoding the zinc finger/*Fok*I fusion proteins. Moreover, when uptake of fluorescent methotrexate was assayed, cells containing this mutation showed 53% mean methotrexate uptake, compared to wild-type CHO cells, consistent with loss of function of one copy of the DHFR gene.

[0433] Thus, targeted, homology-independent integration of exogenous vector sequences occurred at the site of targeted cleavage in the DHFR gene, resulting in the generation of a heterozygous DHFR$^-$ mutant cell line.

Example 38: Multiple mutations in the *Fok*I dimerization domain for targeted cleavage and homology-directed repair of an endogenous gene

[0434] Additional sequence alterations were introduced into the mutagenized *Fok*I cleavage half-domain described in Example 5, in which residue 490 was converted from glutamic acid to lysine (E490K), to provide further improvement in its cleavage specificity. In one embodiment (mutant X2), amino acid 538 was converted from isoleucine to lysine (I538K). In further embodiments, amino acid 486 of the X2 mutant was converted from glutamine to glutamic acid (Q486E) to generate the X3A mutant, or from glutamine to isoleucine (Q486I) to generate the X3B mutant. The amino acid sequences of the E490K, X2, X3A and X3B mutants, compared to the amino acid sequence of the wild-type *Fok*I cleavage half-domain, are presented in Figure 54.

[0435] Plasmids were constructed in which sequences encoding these mutant cleavage half-domains were fused to sequences encoding the 5-8G and 5-9D zinc finger domains (see Example 14). Various combinations of these mutants were then assayed for their ability to stimulate homology-directed repair of a double-stranded break in exon 5 of the IL-2Rγ gene, in the presence of a donor DNA sequence containing a *Bsr*BI site. The assay system and procedures described in Example 15 were used, except that cells were not treated with vinblastine and 20 Units of *Bsr*BI was used for digestion.

[0436] After a 48 hour exposure of the gel, the Phosphorimager screen was read and the intensity of the RFLP-derived and wild-type bands were quantified using ImageQuant software (MolecularDynamics). The intensity of the RFLP-derived band, as percentage of the total radioactivity for the wild-type and RFLP bands, is given in Table 29. The results indicates that the X3 *Fok*I mutant functions significantly better when paired with the Q486E mutant than when paired with a second copy of itself.

Table 29: Homology-directed alteration of an endogenous gene using zinc finger/*Fok*I fusion proteins containing mutations in the *Fok*I dimerization interface*

Sample	5-8	5-9	% GC
1	WT (1 ug)	WT (1 ug)	2.6
2	WT (2.5 ug)	WT (2.5 ug)	<1
3	WT (5 ug)	WT (5 ug)	1.5
4	WT (7.5 ug)	WT (7.5 ug)	<1
5	X3 (1 ug)	Q486E (1 ug)	4.1
6	X3 (2.5 ug)	Q486E (2.5 ug)	4.3
7	X3 (5 ug)	Q486E (5 ug)	8.6
8	X3 (7.5 ug)	Q486E (7.5 ug)	3.6
9	X3 (5 ug)	X3 (5 ug)	0
10	Q486E (5 ug)	Q486E (5 ug)	2.3

* K562 cells were transfected with plasmids encoding two zinc finger/*Fok*I fusion proteins and with a plasmid containing a donor DNA sequence homologous to exon 5 of the IL-2Rγ gene but containing a sequence change resulting in the presence of a *Bsr*BI site. The second and third columns identify the nature of the *Fok*I cleavage half-domain in the 5-8 and 5-9 zinc finger fusion proteins, as follows: WT (wild-type *Fok*I cleavage half-domain); Q486E mutant cleavage half-domain (containing a single amino acid change compared to wild-type, described in Example 5); X3 mutant cleavage half-domain (containing three amino acid changes compared to wild-type, shown in Figure 54). "%GC" refers to fraction of total amplification product that is cleaved by *Bsr*BI, as measured by radioactivity in *Bsr*BI digestion products.

Example 39: Zinc finger/*Fok*I fusion proteins with multiple mutations in the *Fok*I dimerization domain tested with a chromosomal GFP reporter gene assay

[0437] The cell line described in Example 29, containing a chromosomally integrated mutant eGFP coding sequence operatively linked to a tetracycline-regulated

CMV promoter, was used in experiments to test different combinations of zinc finger/*Fok*I fusion proteins (ZFNs) containing amino acid sequence alterations in the dimerization interface of the *Fok*I cleavage half-domain. *See* Example 38 and Figure 54. The exogenous donor DNA construct, previously described in Example 13 and Figure 32, contained a 1527 nucleotide pair insert homologous to wild-type eGFP coding sequences. It was constructed by amplification of eGFP sequences using the following primers:

GFPnostart	GGCGAGGAGCTGTTCAC (SEQ ID NO:235)
pcDNA42571	TGCATACTTCTGCCTGC (SEQ ID NO:236)

The amplification product was topo cloned into the pCR4-TOPO vector to generate the donor construct, denoted pCR4-TOPO_GFPDonor_1.5KB. Targeted, homology-directed integration of this donor sequence will result in replacement of the mutant chromosomal eGFP sequences with wild-type eGFP sequences and doxycycline-inducible expression of functional eGFP.

[0438] Cells containing the chromosomally integrated mutant eGFP sequences (described above) were grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen), supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 2 mM L-glutamine and maintained at 37°C in an atmosphere of 5% CO_2. Cells were transfected using LipofectAMINE 2000 Reagent (Invitrogen) in Opti-MEM I reduced serum medium. Cells were transfected with plasmids encoding the ZFNs only (5 ng each), donor plasmid only (500 ng), or plasmids encoding the ZFNs (5 ng each) + donor plasmid (500 ng). Expression of the chromosomal eGFP coding sequences was activated by the addition of 2ng/ml doxycycline (final concentration) to the growth medium 5 hours post-transfection. The cells were harvested 3 days post-transfection and assayed by flow cytometry for eGFP expression. The results, shown in Table 30, indicate the fusion proteins containing mutations in the dimerization interface function more effectively to promote homology-directed repair in the presence of a different cleavage half-domain, suggesting they are less prone to homodimerization.

Table 30: Homology-directed alteration of a mutant chromosomal eGFP gene using zinc finger/*Fok*I fusion proteins containing mutations in the *Fok*I dimerization interface*

	5-8:					
5-9 ↓	**WT**	**Q486E**	**E490K**	**X2**	**X3A**	**X3B**
WT	0.66	0.38	0.61	0.40	0.70	0.18
Q486E	0.26	0.14	0.54	0.53	0.50	0.23
E490K	0.58	0.42	0.30	0.01	0.02	0.03
X2	0.14	0.55	0.07	0.01	0.01	0.01
X3A	0.43	0.43	0.02	0.01	0.03	0.01
X3B	0.19	0.33	0.06	0.02	0.03	0.02

* Cells were transfected with a donor construct and two ZFN expression constructs: one expressing a 5-8 zinc finger binding domain fused to a *Fok*I cleavage half-domain and the other expressing a 5-9 zinc finger binding domain fused to a *Fok*I cleavage half-domain. The nature of the cleavage half-domain fused to the 5-8 zinc finger binding domain is given across the top row; the nature of the cleavage half-domain fused to the 5-9 zinc finger binding domain is given down the leftmost column. Numbers indicate percentage of cells exhibiting eGFP fluorescence for each pair of ZFNs tested.

Example 40: Targeted homology-dependent integration of a 41-nucleotide exogenous sequence into the endogenous CCR-5 gene

[0439] Growth and transfection of K562 cells were conducted as described in Example 32. Cells were transfected with 2.5 μg of a construct encoding two zinc finger/*Fok*I fusion proteins (separated by a 2A peptide sequence) in which the zinc finger domains (7568 and 7296) were designed to bind target sites in the human CCR-5 gene, and 50 μg of donor construct (see below). The target sites for the zinc finger domains (boxed) are separated by 5 nucleotide pairs, as shown below.

7568

5′ CTGGTCATCCTCATCCTGAT AAACTGCAAAAG GCT

GAC CAGTAGGAGTAG GACTATTTGACGTTTTCCGA 5′

7296

(SEQ ID NO:237)

[0440] The nucleotide sequences of the target sites, and the amino acid sequences of the recognition regions of the zinc finger domains, are shown in Table 31.

Table 31: Zinc Finger Designs for the human CCR-5 gene

Target sequence	F1	F2	F3	F4
GATGAGGATGAC (SEQ ID NO:238)	DRSNLSR (SEQ ID NO:239)	TSANLSR (SEQ ID NO:240)	RSDNLAR (SEQ ID NO:241)	TSANLSR (SEQ ID NO:242)
AAACTGCAAAAG (SEQ ID NO:243)	RSDHLSE (SEQ ID NO:244)	QNANRIT (SEQ ID NO:245)	RSDVLSE (SEQ ID NO:246)	QRNHRTT (SEQ ID NO:247)

Note: The DNA target sequence is shown in the left-most column. The remaining columns show the amino acid sequences (in one-letter code) of residues -1 through +6 of each of the four zinc fingers, with respect to the start of the alpha-helical portion of each zinc finger. Finger F1 is closest to the amino terminus of the protein.

[0441] The donor DNA molecule comprised a ~2 kilobase-pair portion of the human CCR-5 gene engineered to contain a 41 nucleotide pair sequence tag containing a novel diagnostic recognition site for the restriction enzyme *Bgl*I. The donor molecule was constructed by mutagenizing the CCR-5 gene fragment to create a *Xba*I site, and introducing the 41-nucleotide tag into that *Xba*I site. As a result, the 41-nucleotide tag was flanked by approximately 0.5 kilobase pairs of CCR-5 sequence on one side and approximately 1.5 kilobase pairs of CCR-5 sequence on the other. This sequence is shown below, with the 41-nucleotide pair tag shown in upper case and the *Bgl*I site underlined.

gttgtcaaagcttcattcactccatggtgctatagagcacaagattttatttggtgagatggtgctttcatgaattcccccaacag
agccaagctctccatctagtggacagggaagctagcagcaaaccttcccttcactacaaaacttcattgcttggccaaaaaga
gagttaattcaatgtagacatctatgtaggcaattaaaaacctattgatgtataaaacagtttgcattcatggagggcaactaaat
acattctaggactttataaaagatcactttttatttatgcacagggtggaacaagatggattatcaagtgtcaagtccaatctatga
catcaattattatacatcggagccctgccaaaaaatcaatgtgaagcaaatcgcagcccgcctcctgcctccgctctactcact
ggtgttcatctttggttttgtgggcaacatgctggtcatcctcatctagaTCAGTGAGTAT<u>GCCCTGATGGC</u>
GTCTGGACTGGATGCCTCGtctagataaactgcaaaaggctgaagagcatgactgacatctacctgctcaa
cctggccatctctgacctgtttttccttcttactgtccccttctgggctcactatgctgccgcccagtgggactttggaaatacaat
gtgtcaactcttgacagggctctattttataggcttcttctctggaatcttcttcatcatcctcctgacaatcgataggtacctggct
gtcgtccatgctgtgtttgctttaaaagccaggacggtcacctttggggtggtgacaagtgtgatcacttgggtggtggctgtg
tttgcgtctctcccaggaatcatctttaccagatctcaaaaagaaggtcttcattacacctgcagctctcattttccatacagtcag
tatcaattctggaagaatttccagacattaaagatagtcatcttggggctggtcctgccgctgcttgtcatggtcatctgctactc
gggaatcctaaaaactctgcttcggtgtcgaaatgagaagaagaggcacagggctgtgaggcttatcttcaccatcatgattg
tttattttctcttctgggctccctacaacattgtccttctcctgaacaccttccaggaattctttggcctgaataattgcagtagctct
aacaggttggaccaagctatgcaggtgacagagactcttgggatgacgcactgctgcatcaaccccatcatctatgcctttgt
cggggagaagttcagaaactacctcttagtcttcttccaaaagcacattgccaaacgcttctgcaaatgctgttctattttccagc
aagaggctcccgagcgagcaagctcagtttacacccgatccactggggagcaggaaatatctgtgggcttgtgacacgga
ctcaagtgggctggtgacccagtcagagttgtgcacatggcttagttttcatacacagcctgggctgggggtggggtgggag
aggtctttttaaaaggaagttactgttatagagggtctaagattcatccatttatttggcatctgtttaaagtagattagatcttttaa
gcccatcaattatagaaagccaaatcaaaatatgttgatgaaaaatagcaacctttttatctccccttcacatgcatcaagttattg
acaaactctcccttcactccgaaagttccttatgtatatttaaaagaaagcctcagagaattgctgattcttgagtttagtgatctg

aacagaaataccaaaattatttcagaaatgtacaactttttacctagtacaaggcaacatataggttgtaaatgtgtttaaaacag gtctttgtcttgctatggggagaaaagacatgaatatgattagtaaagaaatgacacttttcatgtgtgatttc (SEQ ID NO:248)

[0442] Six days after transfection, cellular DNA was isolated and used as a template for amplification as described in Example 32, then digested with *Bgl*I. In DNA from cells transfected with both the donor construct and the construct encoding two zinc finger/*Fok*I fusion proteins, approximately 1% of the amplification products were cleaved by *Bgl*I, indicative of targeted insertion of the sequence tag into the CCR-5 gene. DNA from untransfected cells, cells that were transfected only with the donor construct, and cells that were transfected only with the construct encoding two zinc finger/*Fok*I fusion proteins did not yield amplification products that were cleaved by *Bgl*I. It is significant that the targeted insertion of this 41-nucleotide sequence tag generates a frameshift mutation in the CCR-5 gene, thereby inactivating gene function, including its function as a receptor for HIV.

CLAIMS

1. A protein comprising an engineered zinc finger protein DNA-binding domain, wherein the DNA-binding domain comprises four zinc finger recognition regions ordered F1 to F4 from N-terminus to C-terminus, and

(i) wherein F1, F2, F3, and F4 comprise the following amino acid sequences:

F1: QSGALAR (SEQ ID NO:229)

F2: RSDNLRE (SEQ ID NO:230)

F3: QSSDLSR (SEQ ID NO:231)

F4: TSSNRKT (SEQ ID NO:232) or

(ii) wherein F1, F2, F3, and F4 comprise the following amino acid sequences:

F1: RSDTLSE (SEQ ID NO:224)

F2: NNRDRTK (SEQ ID NO:225)

F3: RSDHLSA (SEQ ID NO:226)

F4: QSGHLSR (SEQ ID NO:227).

2. The protein according to claim 1, further comprising a *FokI* cleavage domain or a *FokI* cleavage half-domain.

3. The protein of claim 2, wherein the cleavage half-domain is a wild-type *FokI* cleavage half-domain or an engineered *FokI* cleavage half-domain comprising an amino acid mutation that is E490K, E490R, or Q486E.

4. A polynucleotide encoding the protein of claim 2 or 3.

5. An isolated cell comprising the protein of any one of claims 1 to 3 or the polynucleotide of claim 4.

6. A method for inactivating an endogenous dhfr gene in a cell *in vitro*, the method comprising:

(a) introducing, into the cell, a polynucleotide according to claim 4, such that the first polynucleotide expresses the protein in the cell and wherein the protein binds to the target site and cleaves the dihydrofolate reductase (dhfr) gene.

7. A method of producing a recombinant protein of interest in a host cell *in vitro*, the host cell comprising an endogenous dihydrofolate reductase (dhfr) gene, the method comprising the steps of:

(a) inactivating the endogenous dhfr gene of the host cell by the method of claim 6;

(b) introducing an expression vector comprising transgene, the transgene comprising a dhfr gene and a sequence encoding a protein of interest into the host cell; and

(c) selecting cells in which the transgene is stably integrated into and expressed by the host cell, thereby producing the recombinant protein.

8. The method of claim 6 or 7, wherein the host cell is a CHO cell.

9. A dihydrofolate reductase (dhfr)-deficient cell line produced by:

inactivating endogenous dhfr genes in a population of cells *in vitro* by introducing, into the cells, a polynucleotide according to claim 4, such that the protein binds to the target site and cleaves the dhfr gene; and

culturing the cells such that a dhfr-deficient cell line is produced.

10. The dhfr negative cell line of claim 9, wherein the cell line is a CHO cell line.

FIGURE 1

```
                          MetGlyPheLeuLysLeuIle   (SEQ ID
      NO:2)
CTGCCGCCGGCGCCGCGGCCGTCATGGGGTTCCTGAAACTGATT
GACGGCGGCCGCGGCGCCGGCAGTACCCCAAGGACTTTGACTAA   (SEQ ID
      NO:1)
    |    |    |    |    |    |    |    |
    5   10   15   20   25   30   35   40
```

FIGURE 2

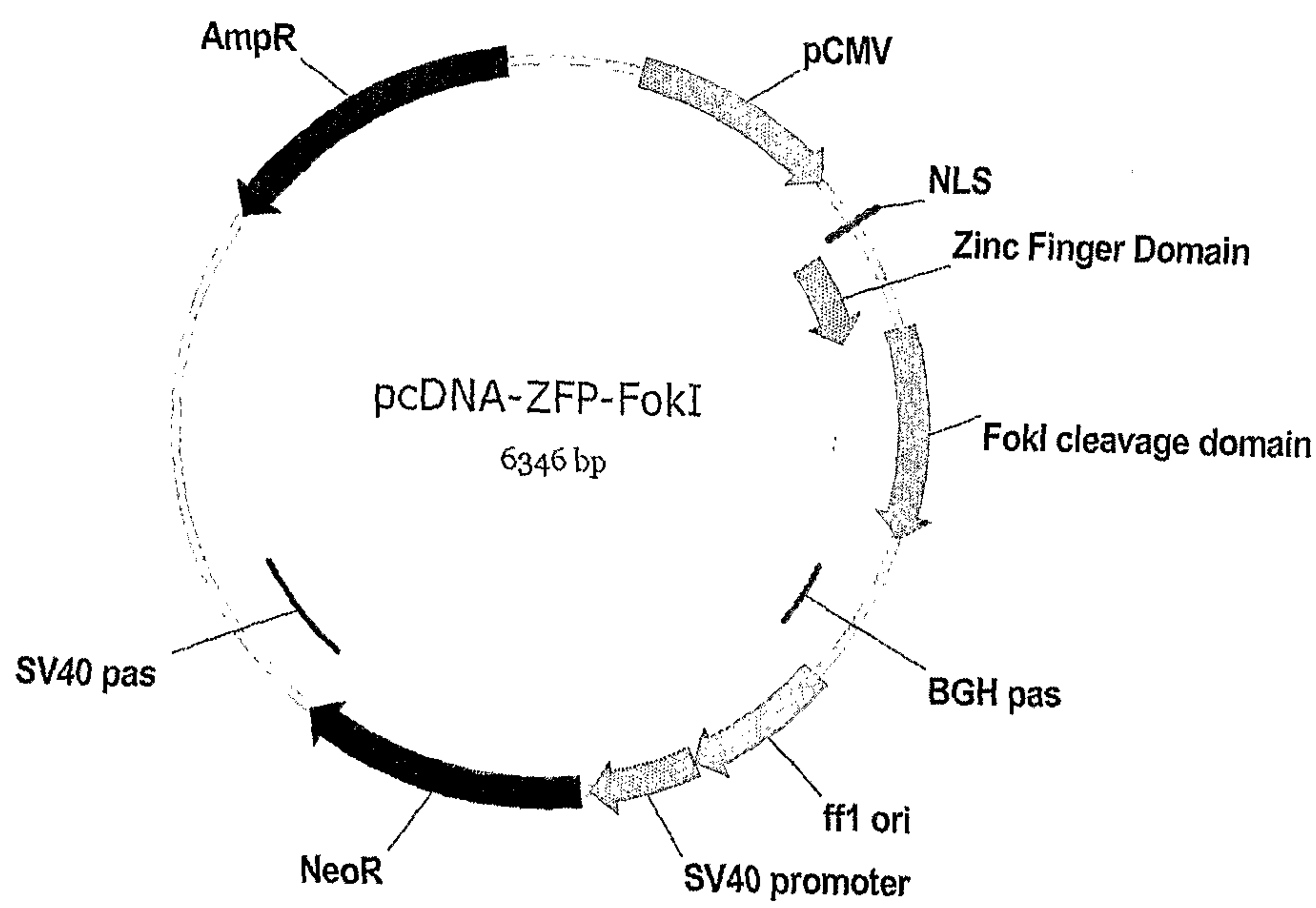
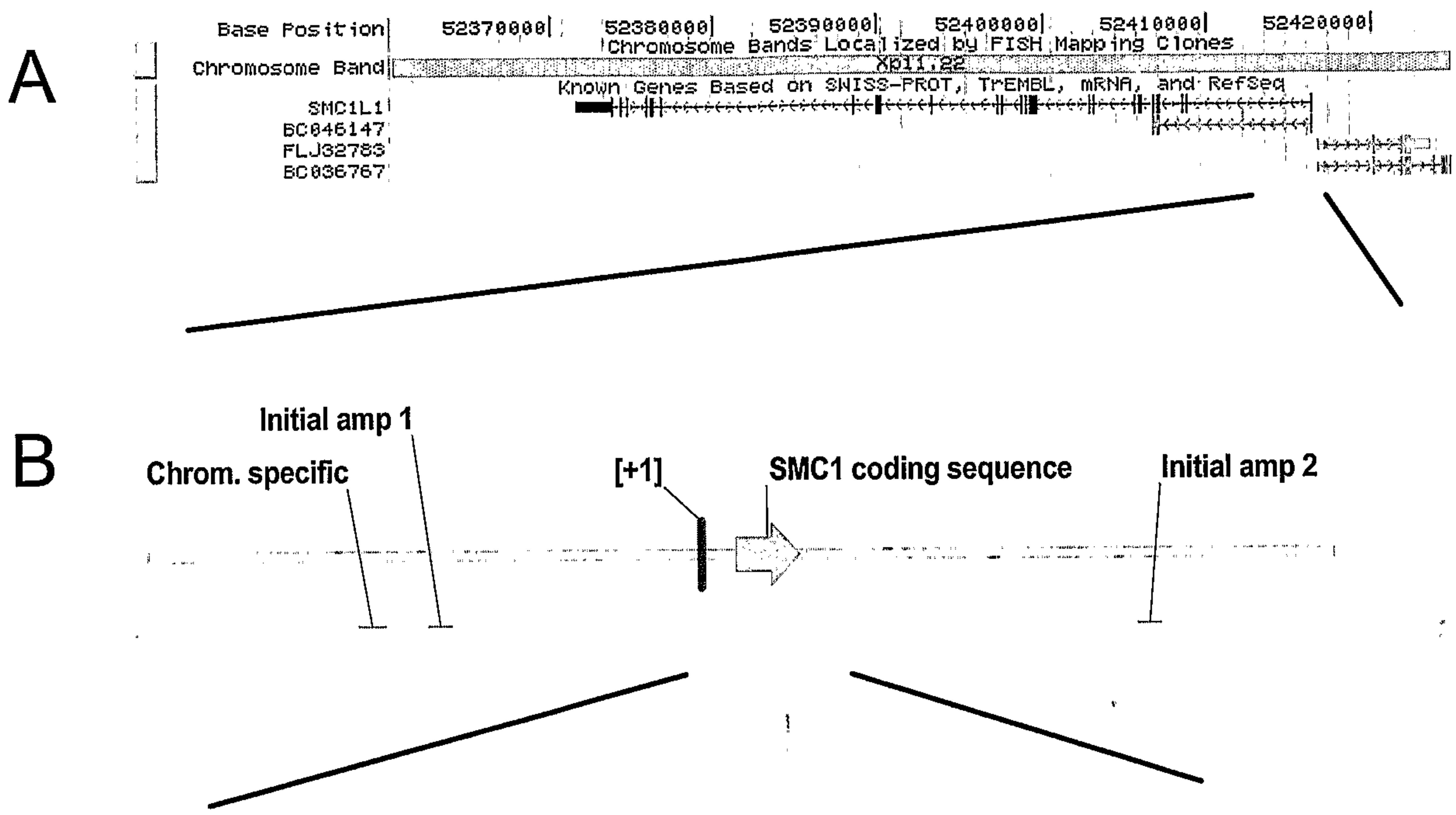

AmpR
pCMV
NLS
Zinc Finger Domain
pcDNA-ZFP-FokI
6346 bp
FokI cleavage domain
SV40 pas
BGH pas
ff1 ori
NeoR
SV40 promoter

FIGURE 3



C

```
                        SMC1 ZFP #1
                        ~~~~~~~~~~~~~
                        MetGlyPheLeuLysLeuIleGlu
CTGCCGCCGGCGCCGCGGCCGTCATGGGGTTCCTGAAACTGATTGAG
GACGGCGGCCGCGGCGCCGGCAGTACCCCAAGGACTTTGACTAACTC
   ~~~~~~~~~~~~~
   SMC1 ZFP #2
```

Chromosomal target DNA

D

```
CTGCCGCCGGCGCCGCGGCCGTCATAAGAAGCTTCCTGAAACTGATTGAG
GACGGCGGCCGCGGCGCCGGCAGTATTCTTCGAAGGACTTTGACTAACTC
```

donor DNA (fragment)

FIGURE 4

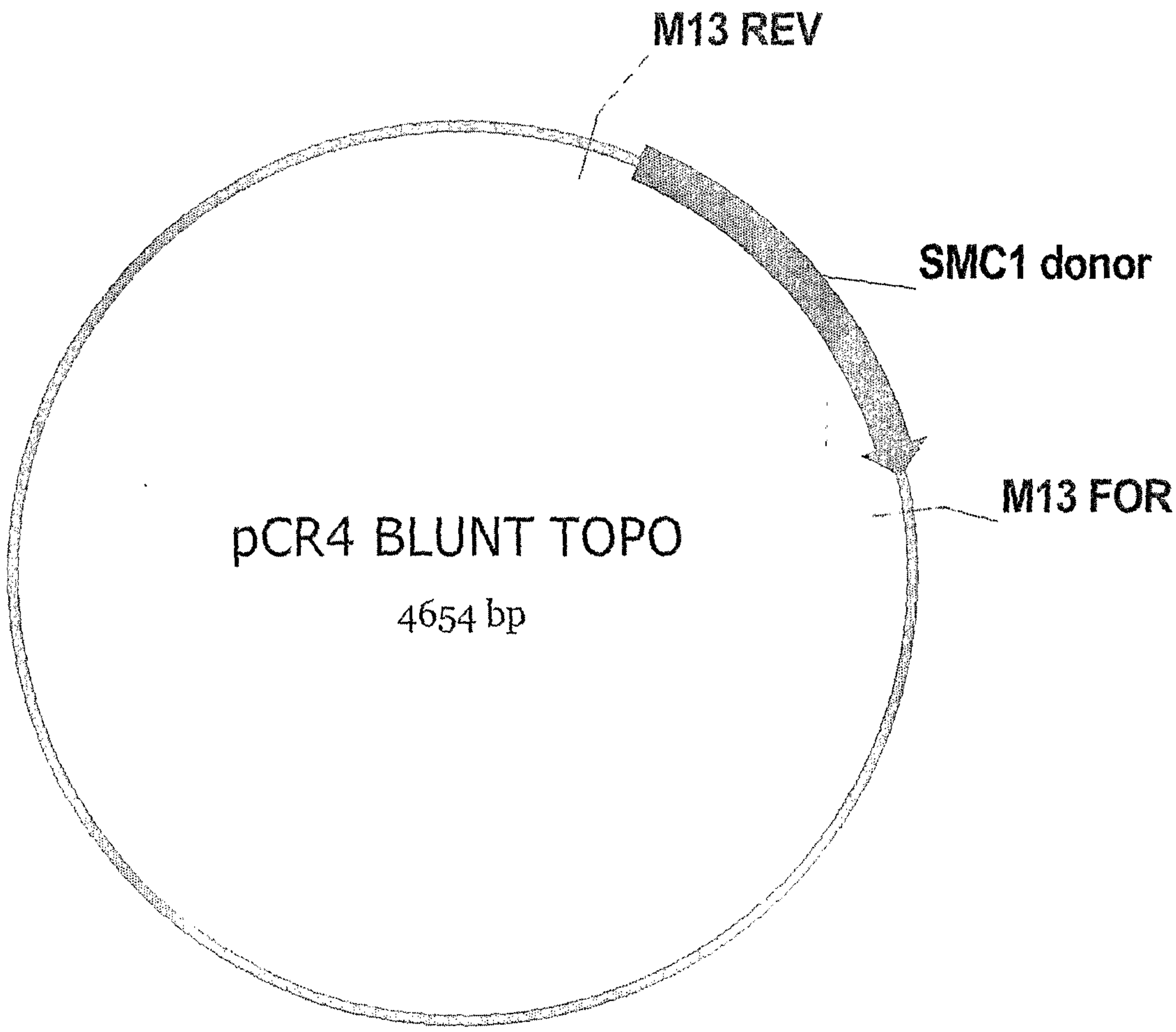


FIGURE 5

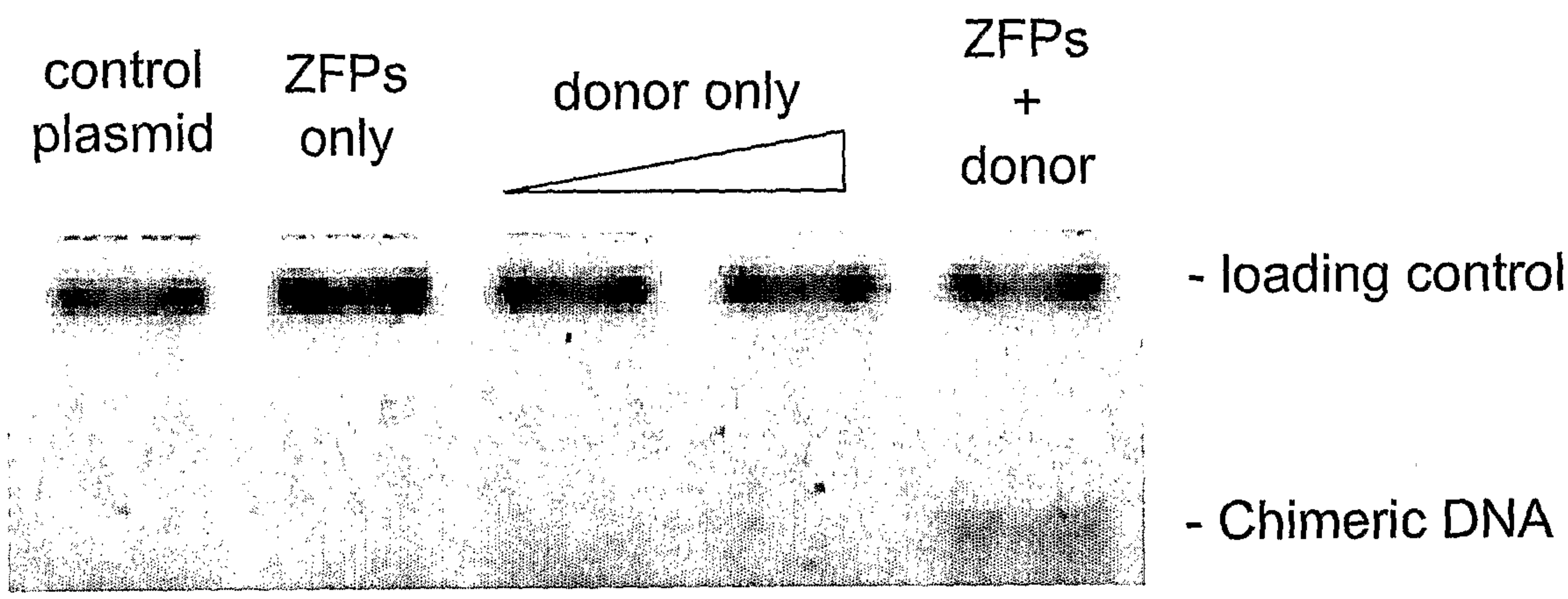

control plasmid
ZFPs only
donor only
ZFPs + donor
- loading control
- Chimeric DNA

FIGURE 6

TAGTCCTGCAGGTTTAAACGAATTCGCCCTTCTCAGCAAGCGTGAGCTCA
50
GGTCTCCCCCGCCTCCTTGAACCTCAAGAACTGCTCTGACTCCGCCCAGC
100
AACAACTCCTCCGGGGATCTGGTCCGCAGGAGCAAGTGTTTGTTGTTGCC
150
ATGCAACAAGAAAAGGGGGCGGAGGCACCACGCCAGTCGTCAGCTCGCTC
200
CTCGTATACGCAACATCAGTCCCCGCCCCTGGTCCCACTCCTGCCGGAAG
250
GCGAAGATCCCGTTAGGCCTGGACGTATTCTCGCGACATTTGCCGGTCGC
300
CCGGCTTGCACTGCGGCGTTTCCCGCGCGGGCTACCTCAGTTCTCGGGCG
350
TACGGCGCGGCCTGTCCTACTGCTGCCGGCGCCGCGGCCGTCATaagaag
400
cTTCCTGAAACTGATTGAAGGGCGAATTCGCGGCCGCTAAATTCAATTCG
450
CCCTATAGTGAGT

(SEQ ID NO: 6)

FIGURE 7

```
                    TyrLysAsnSerAspAsnAspLysVal        (SEQ ID
NO:8)
CTTCCAACCTTTCTCCTCTAGGTACAAGAACTCGGATAATGATAAAGTCC
GAAGGTTGGAAAGAGGAGATCCATGTTCTTGAGCCTATTACTATTTCAGG     (SEQ ID
NO:7)
      |    |    |    |    |    |    |    |    |    |
      5   10   15   20   25   30   35   40   45   50
```

FIGURE 8

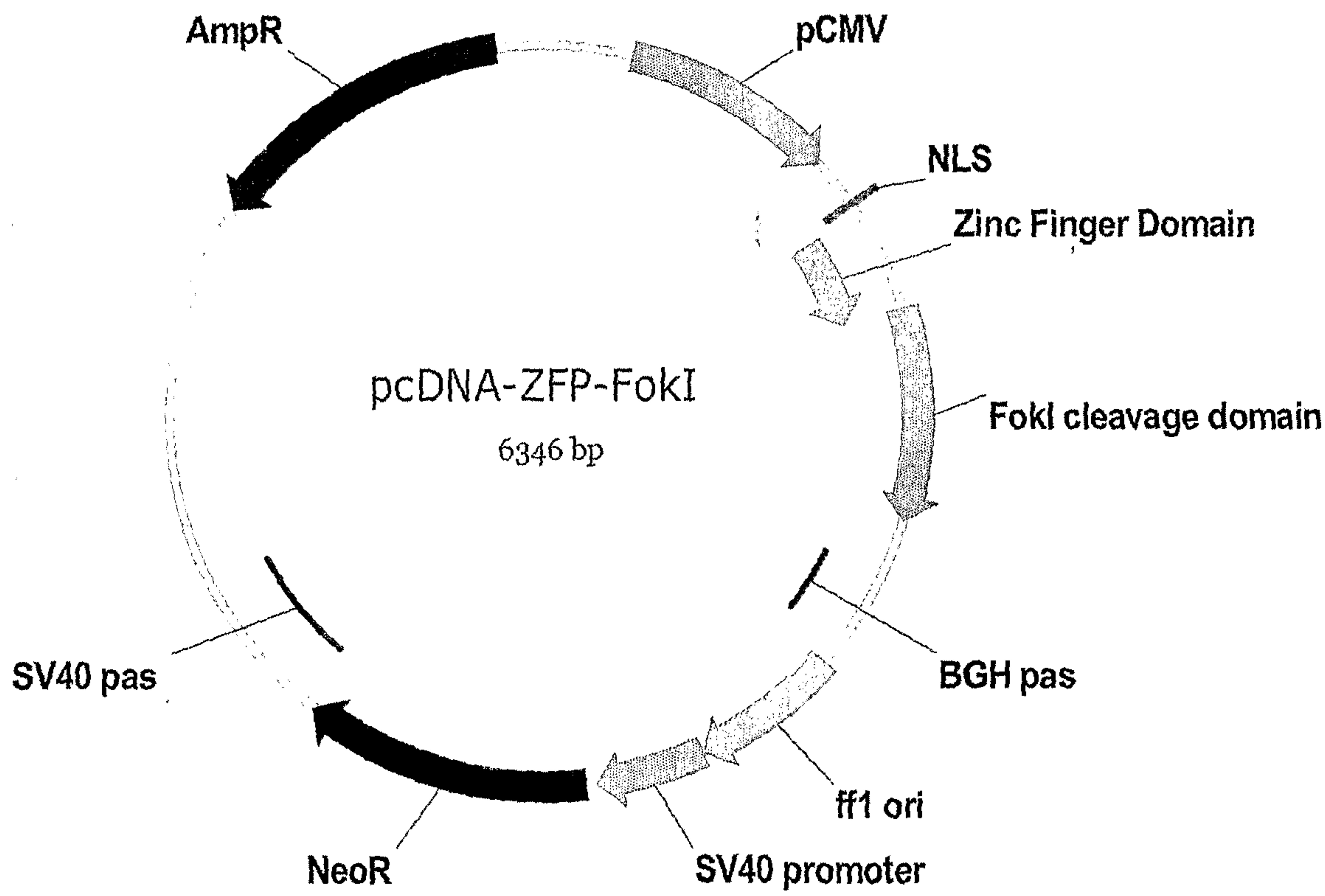
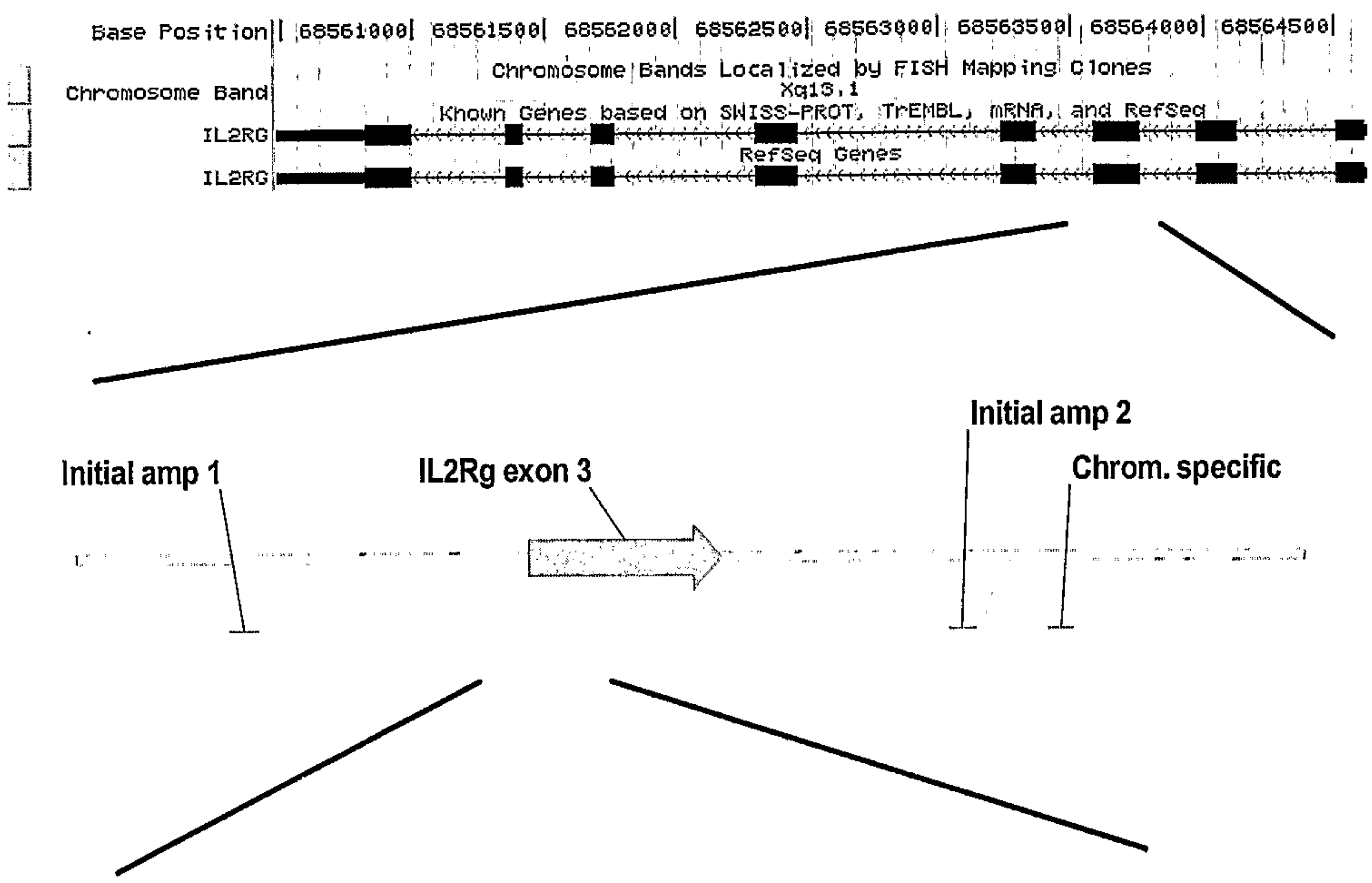

AmpR
pCMV
NLS
Zinc Finger Domain
pcDNA-ZFP-FokI
6346 bp
FokI cleavage domain
BGH pas
SV40 pas
ff1 ori
SV40 promoter
NeoR

FIGURE 9

A

B



C

```
                                               IL2Rγ ZFP #1
                                               ~~~~~~~~~~~~
                                          TyrLysAsnSerAspAsnAspLysVal
GTTCCTCTTCCTTCCAACCTTTCTCCTCTAGGTACAAGAACTCGGATAATGATAAAGTC
CAAGGAGAAGGAAGGTTGGAAAGAGGAGATCCATGTTCTTGAGCCTATTACTATTTCAG
                         ~~~~~~~~~~~~~
                         IL2Rγ ZFP #2
```

Chromosomal target DNA

D

```
GTTCCTCTTCCTTCCAACCTTTCTCCTCTAGGTAAAAGAATTCCGACAACGATAAAGTC
CAAGGAGAAGGAAGGTTGGAAAGAGGAGATCCATTTTCTTAAGGCTGTTGCTATTTCAG
```

Donor DNA (fragment)

FIGURE 10

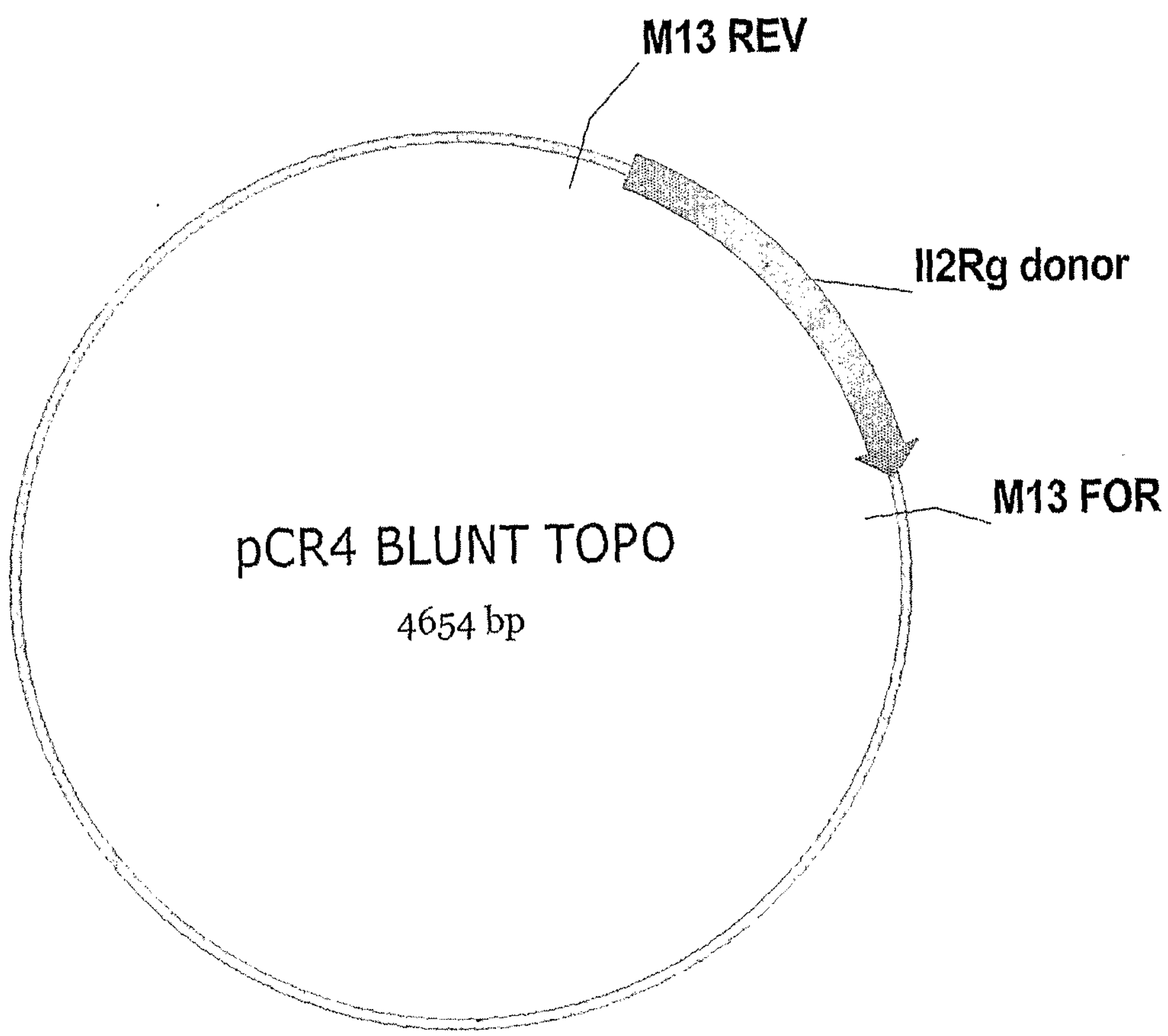


FIGURE 11

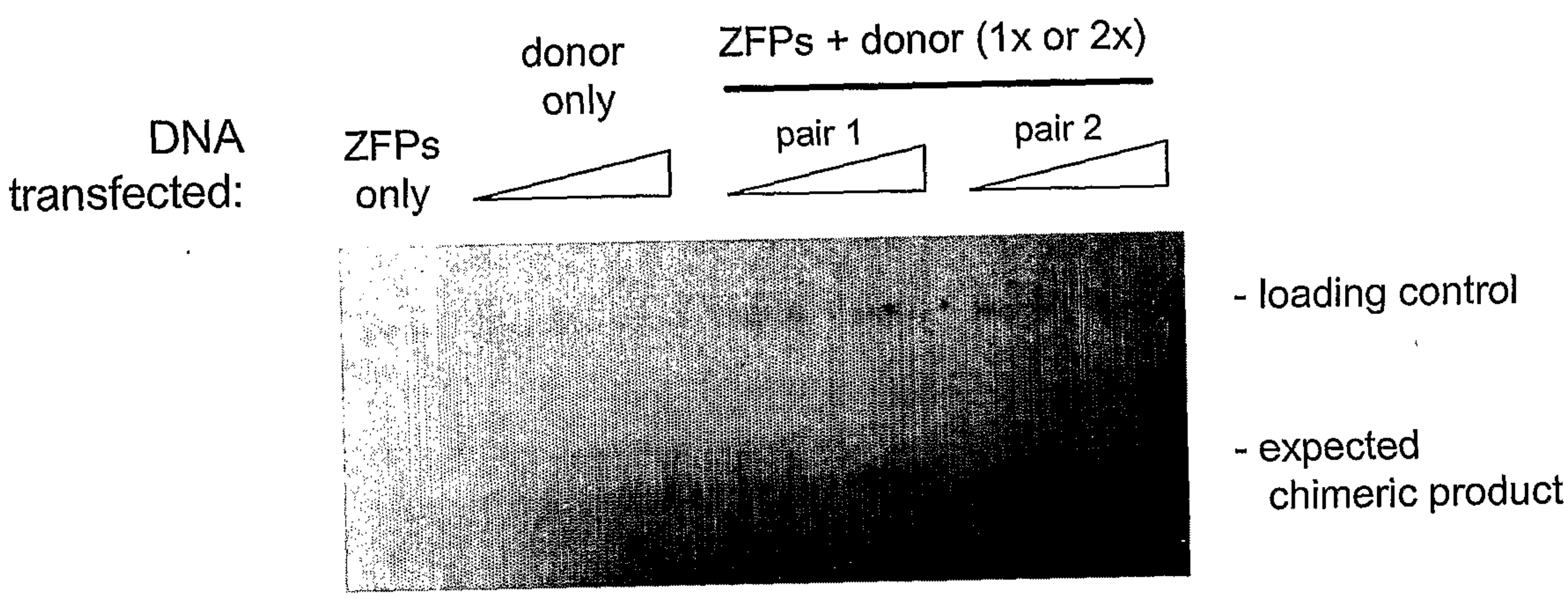

DNA transfected:
ZFPs only
donor only
ZFPs + donor (1x or 2x)
pair 1
pair 2
- loading control
- expected chimeric product

FIGURE 12

TAGTCCTGCAGGTTTAAACGAATTCGCCCTTTCCTCTAGGTAaAAGAAtT
50
CcGAcAAcGATAAAGTCCAGAAGTGCAGCCACTATCTATTCCCTGAAGAA
100
ATCACTTCTGGCTGTCAGTTGCAAAAAAAGGAGATCCACCTCTACCAAAC
150
ATTTGTTGTTCAGCTCCAGGACCCACGGGAACCCAGGAGACAGGCCACAC
200
AGATGCTAAAACTGCAGAATCTGGGTAATTTGGAAAGAAAGGGTCAAGAG
250
ACCAGGGATACTGTGGGACATTGGAGTCTACAGAGTAGTGTTCTTTTATC
300
ATAAGGGTACATGGGCAGAAAAGAGGAGGTAGGGGATCATGATGGGAAGG
350
GAGGAGGTATTAGGGGCACTACCTTCAGGATCCTGACTTGTCTAGGCCAG
400
GGGAATGACCACATATGCACACATATCTCCAGTGATCCCCTGGGCTCCAG
450
AGAACCTAACACTTCACAAACTGAGTGAATCCCAGCTAGAACTGAACTGG
500
AACAACAGATTCTTGAACCACTGTTTGGAGCACTTGGTGCAGTACCGGAC
550
TAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGT
600
ATTACAATTCACTGGCCGTCGTTT

(SEQ ID NO:12)

FIGURE 13

TACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGT
50
TTGAAGTCCAACTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGG
100
CTGTCATCACTTAGACCTCACCCTGTGGAGCCACACCCTAGGGTTGGCCA
150
ATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGGCTGGGCATAAAAGT
200
CAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACTGTGTTC
250
ACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTG**A**GGAGAAG
300
TCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGG
350
TGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGA
400
CCAATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGAT
450
AGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGC
500
TGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATCTG
550
TCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAA
600
GAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCA
650
AGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTG
700

(SEQ ID NO:13)

FIGURE 14

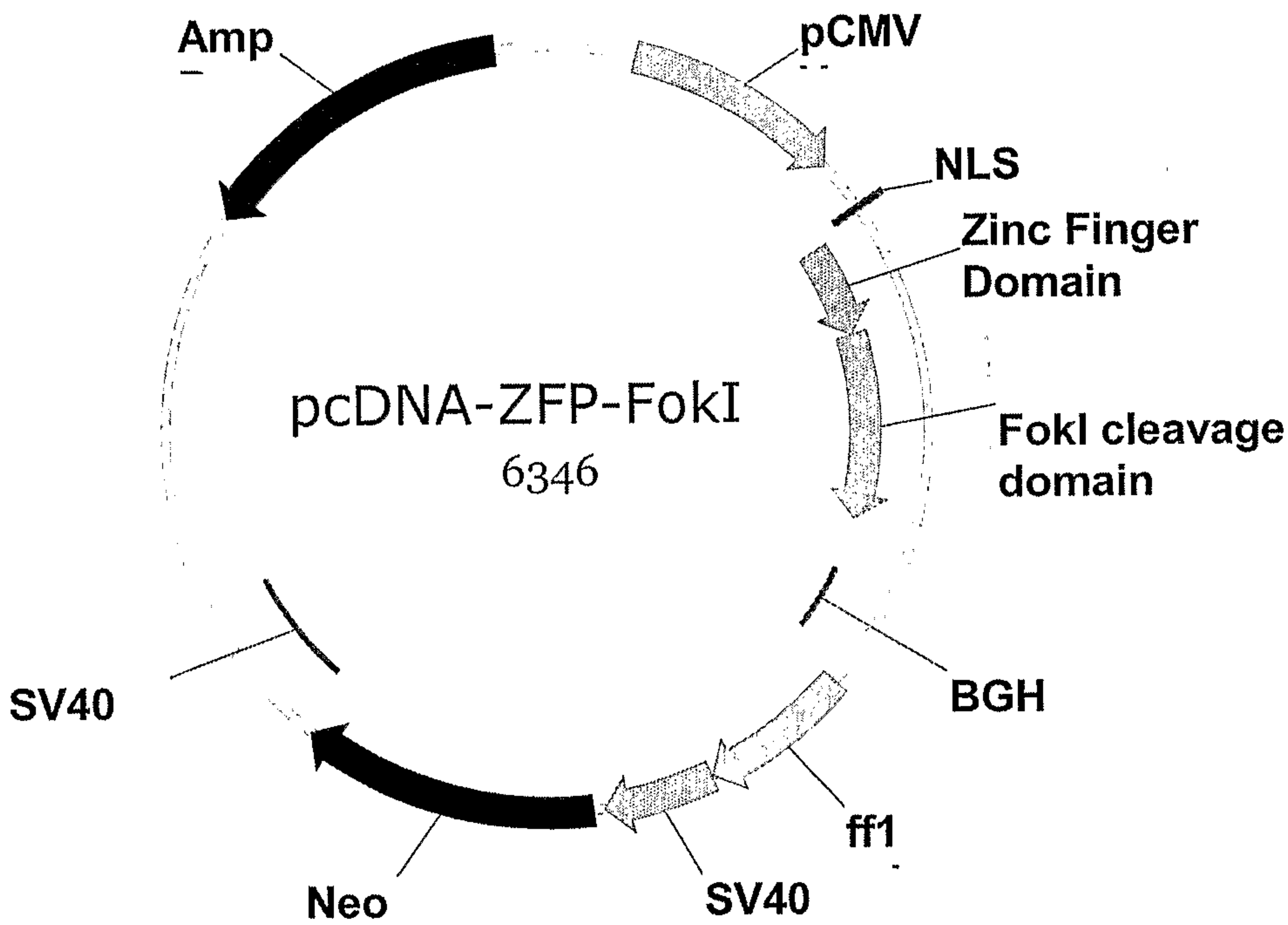
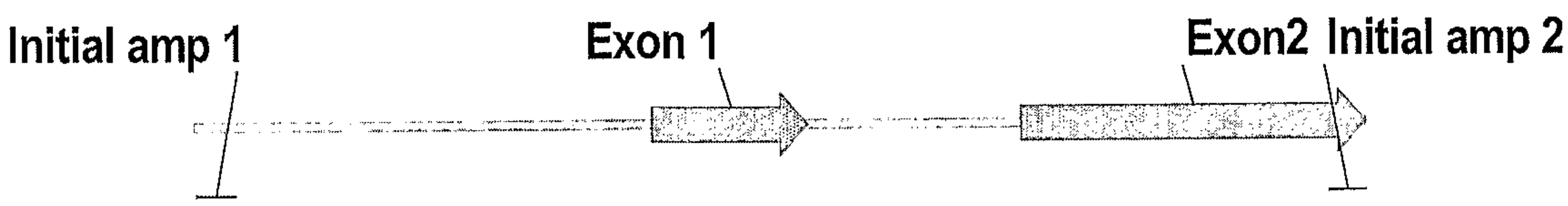

Amp
pCMV
NLS
Zinc Finger Domain
pcDNA-ZFP-FokI
6346
FokI cleavage domain
BGH
ff1
SV40
Neo
SV40

FIGURE 15



Beta globin human genomic TARGET
700 bp

FIGURE 16

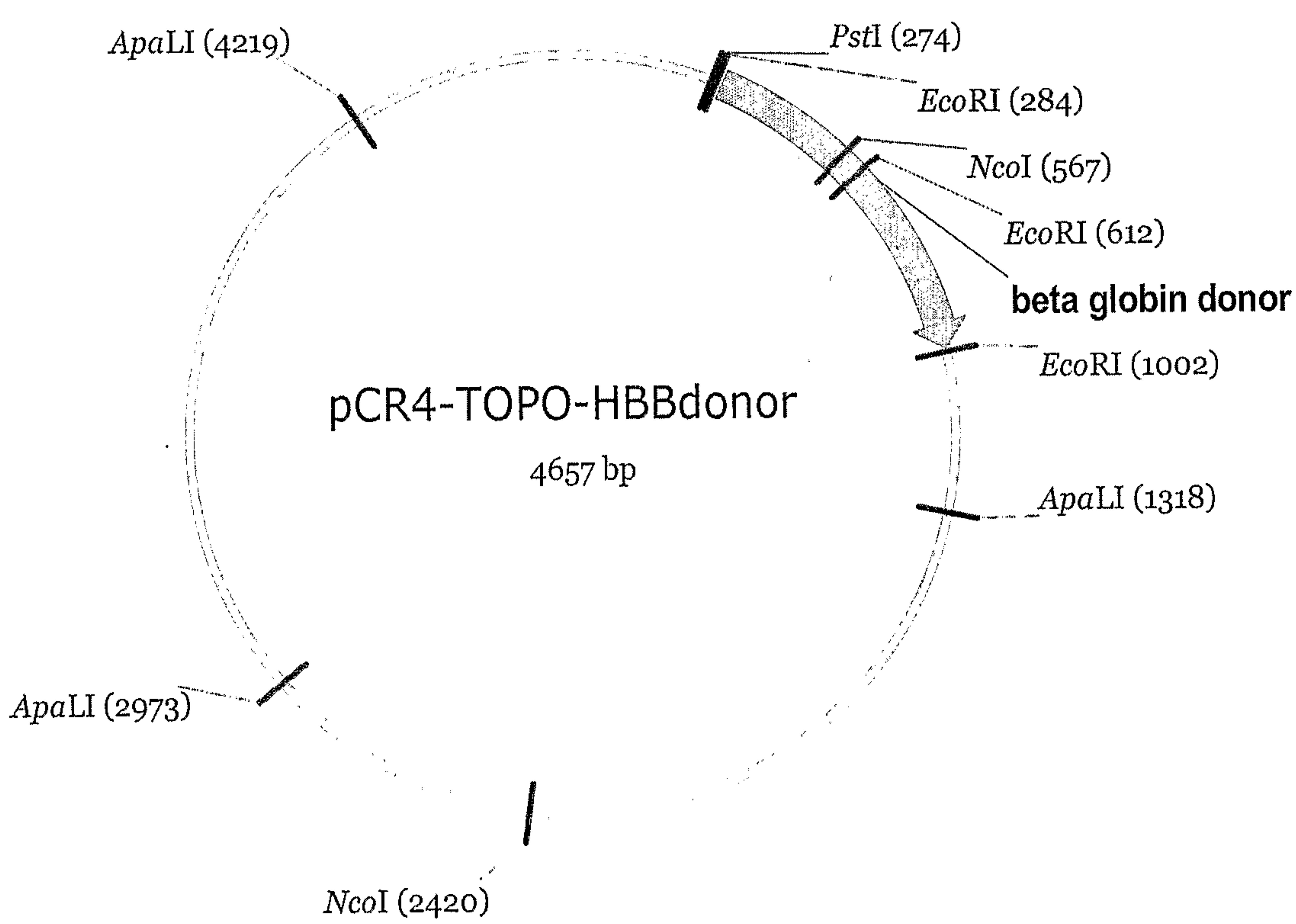
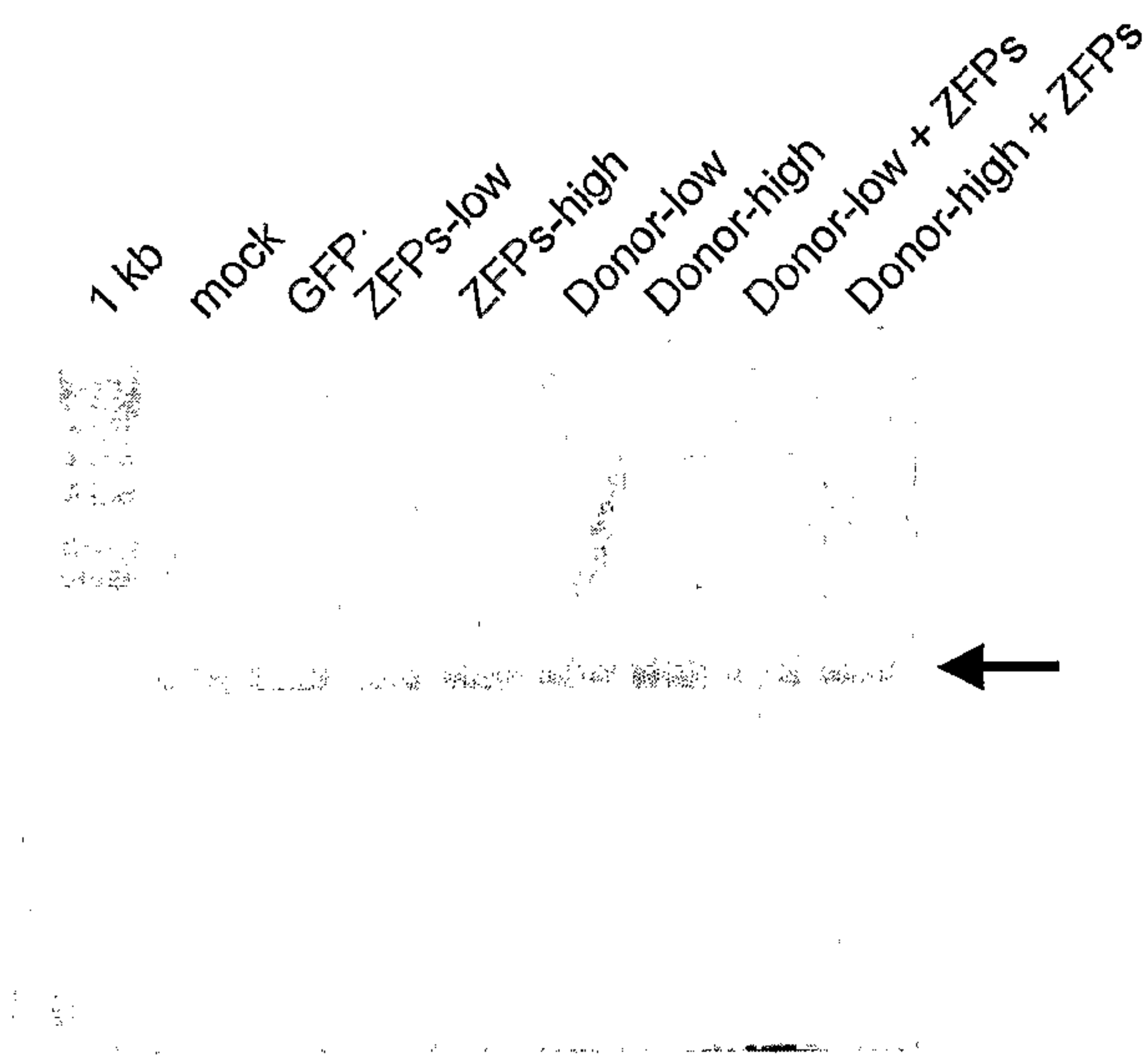
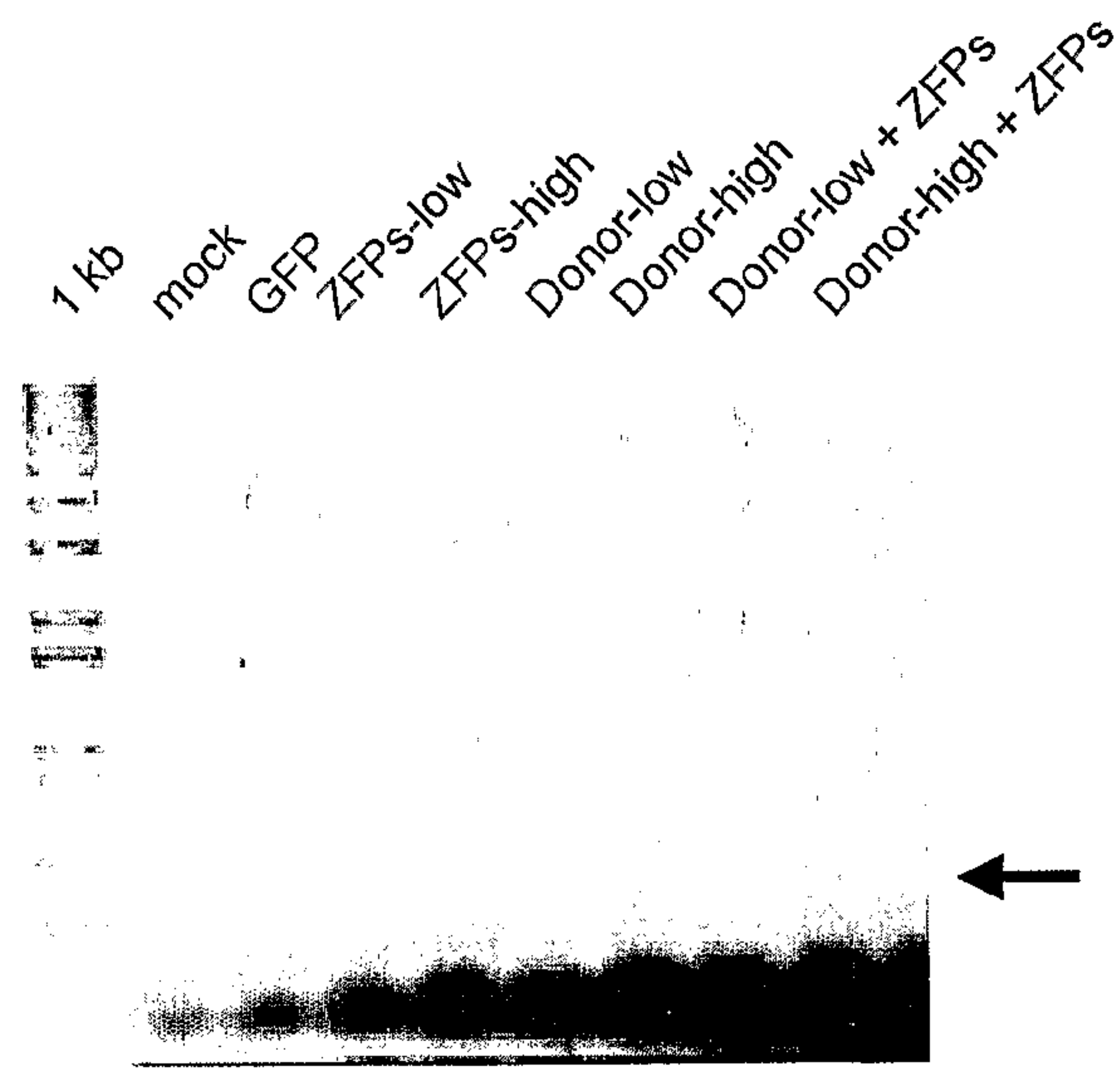

PstI (274)
EcoRI (284)
NcoI (567)
EcoRI (612)
beta globin donor
EcoRI (1002)
ApaLI (1318)
NcoI (2420)
ApaLI (2973)
ApaLI (4219)
pCR4-TOPO-HBBdonor
4657 bp

FIGURE 17



Loading Control



GC-specific Product

FIGURE 18

TGCTTACCAAGCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGA
50
GGATGTTTTTAGTAGCAATTTGTACTGATGGTATGGGGCCAAGAGATATA
100
TCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACTCCTAAGCCAGTGCCA
150
GAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCCTGTGG
200
AGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGG
250
AGCCAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATT
300
TGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGT
350
GCATCTGACTCCTGAGGAGAAGTCTGgCGTTAgTGCCCgaattccgAtcG
400
TcAACcac

(SEQ ID NO:14)

FIGURE 19

```
                                            5-8
CACGTTTCGTGTTCGGAGCCGCTTTAACCCACTCTGTGGAAG
GTGCAAAGCACAAGCCTCGGCGAAATTGGGTGAGACACCTTC
   5-10
```

(SEQ ID NO:15)

FIGURE 20

```
  1 MAPKKKRKVG IHGVPAAMAE RPFQCRICMR NFSRSDNLSE HIRTHTGEKP FACDICGRKF
 61 ARNAHRINHT KIHTGSQKPF QCRICMRNFS RSDTLSEHIR THTGEKPFAC DICGRKFAAR
121 STRTTHTKIH LRQKDAARGS QLVKSELEEK KSELRHKLKY VPHEYIELIE IARNSTQDRI
181 LEMKVMEFFM KVYGYRGKHL GGSRKPDGAI YTVGSPIDYG VIVDTKAYSG GYNLPIGQAD
241 EMQRYVEENQ TRNKHINPNE WWKVYPSSVT EFKFLFVSGH FKGNYKAQLT RLNHITNCNG
301 AVLSVEELLI GGEMIKAGTL TLEEVRRKFN NGEINF
```

(SEQ ID NO:16)

FIGURE 21

```
  1 MAPKKKRKVG IHGVPAAMAE RPFQCRICMR NFSRSDSLSR HIRTHTGEKP FACDICGRKF
 61 ADSSNRKTHT KIHTGGGGSQ KPFQCRICMR NFSRSDSLSV HIRTHTGEKP FACDICGRKF
121 ADRSNRITHT KIHLRQKDAA RGSQLVKSEL EEKKSELRHK LKYVPHEYIE LIEIARNSTQ
181 DRILEMKVME FFMKVYGYRG KHLGGSRKPD GAIYTVGSPI DYGVIVDTKA YSGGYNLPIG
241 QADEMQRYVE ENQTRNKHIN PNEWWKVYPS SVTEFKFLFV SGHFKGNYKA QLTRLNHITN
301 CNGAVLSVEE LLIGGEMIKA GTLTLEEVRR KFNNGEINF
```

(SEQ ID NO:17)

FIGURE 22

CGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTG
AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGC
ACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC
AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATG
GCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAG
GACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC
CCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGAC
CCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATC

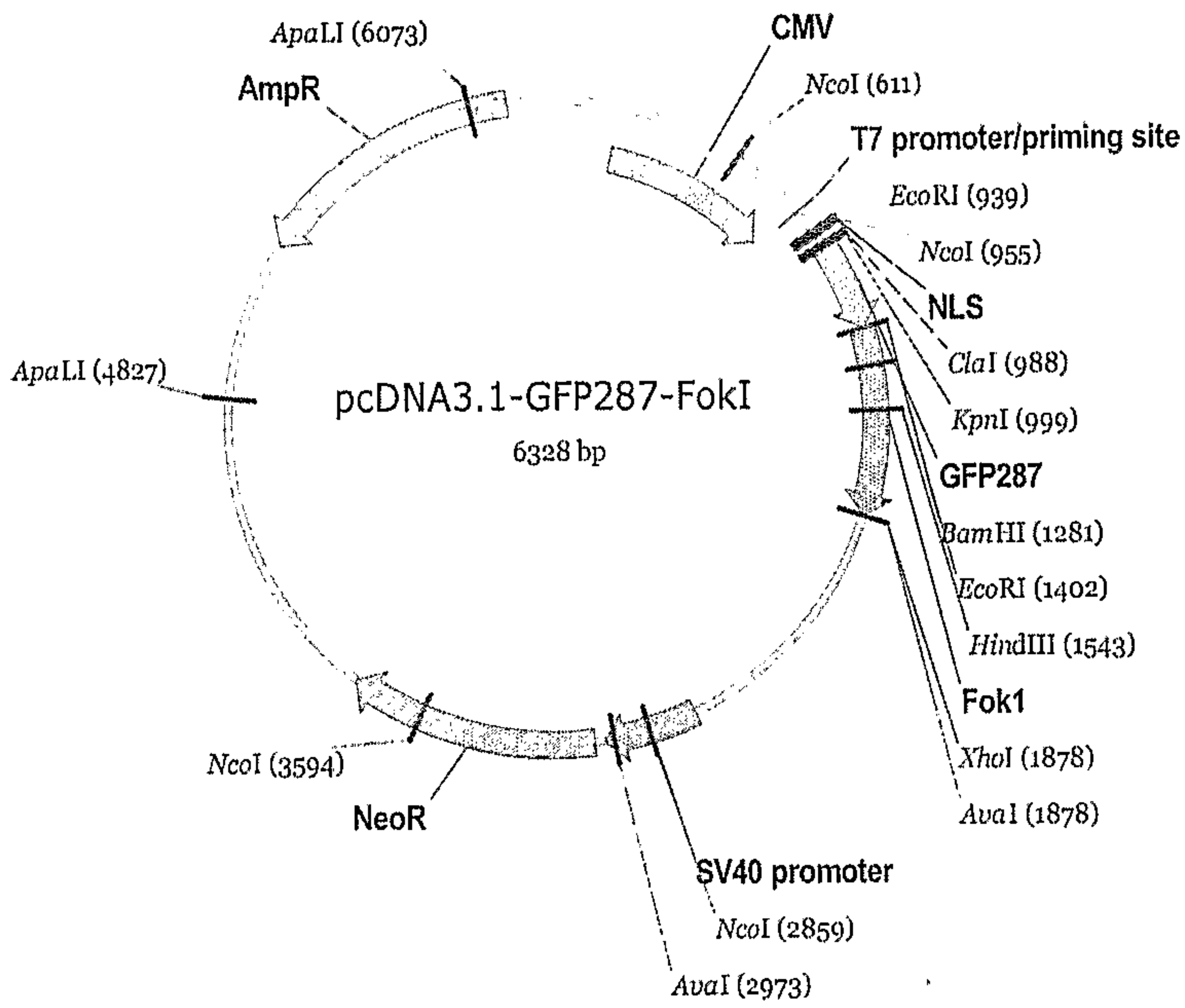
(SEQ ID NO:18)

FIGURE 23

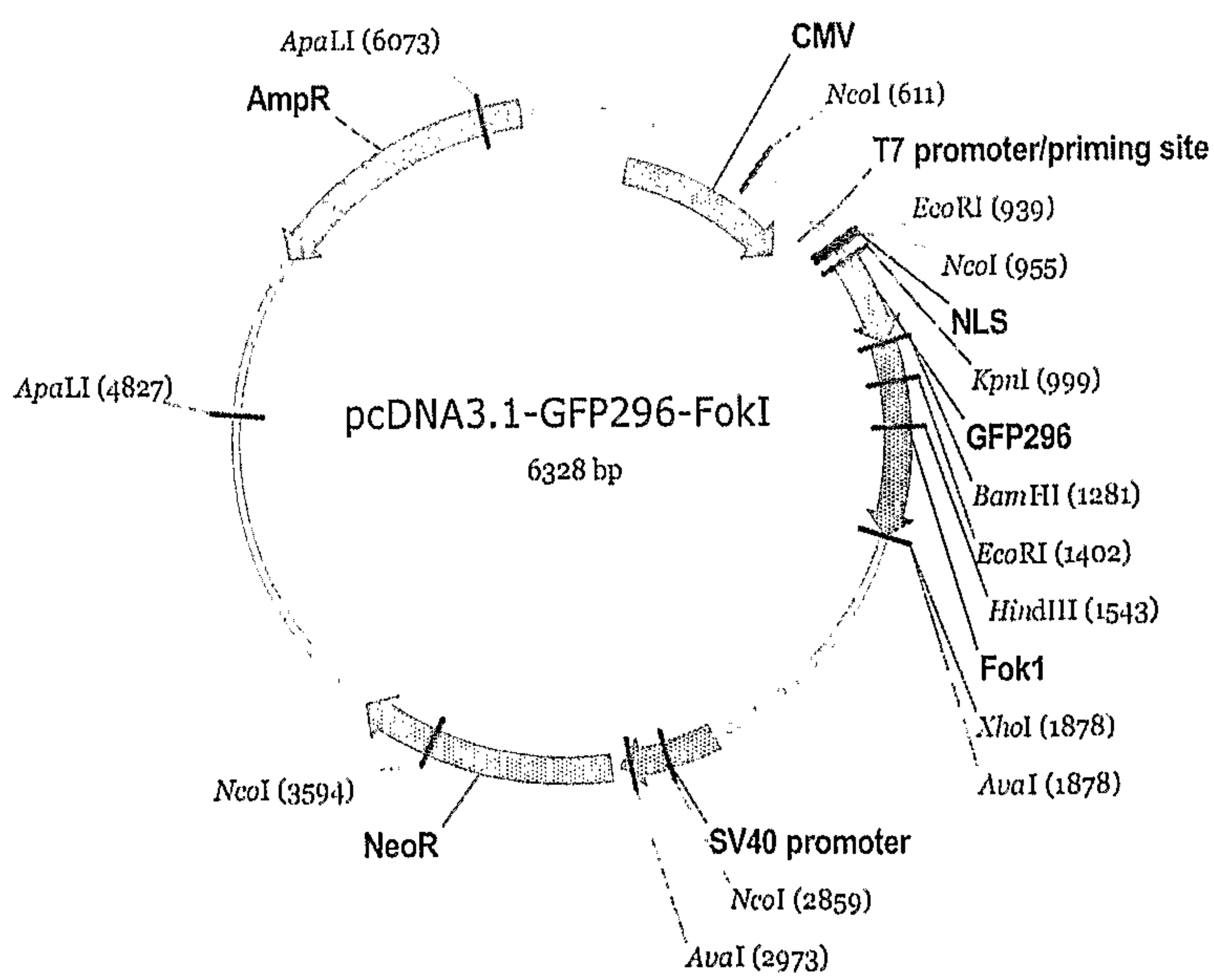
CGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATGGTG
AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCTAACAC
GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
CATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG
CGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAA
CATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGC
CGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCC
CGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCAC
TCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATC

(SEQ ID NO:19)

FIGURE 24



ApaLI (6073)
AmpR
CMV
NcoI (611)
T7 promoter/priming site
EcoRI (939)
NcoI (955)
NLS
ClaI (988)
KpnI (999)
GFP287
BamHI (1281)
EcoRI (1402)
HindIII (1543)
Fok1
XhoI (1878)
AvaI (1878)
ApaLI (4827)
pcDNA3.1-GFP287-FokI
6328 bp
NcoI (3594)
NeoR
SV40 promoter
NcoI (2859)
AvaI (2973)



ApaLI (6073)
AmpR
CMV
NcoI (611)
T7 promoter/priming site
EcoRI (939)
NcoI (955)
NLS
KpnI (999)
GFP296
BamHI (1281)
EcoRI (1402)
HindIII (1543)
Fok1
XhoI (1878)
AvaI (1878)
ApaLI (4827)
pcDNA3.1-GFP296-FokI
6328 bp
NcoI (3594)
NeoR
SV40 promoter
NcoI (2859)
AvaI (2973)

FIGURE 25

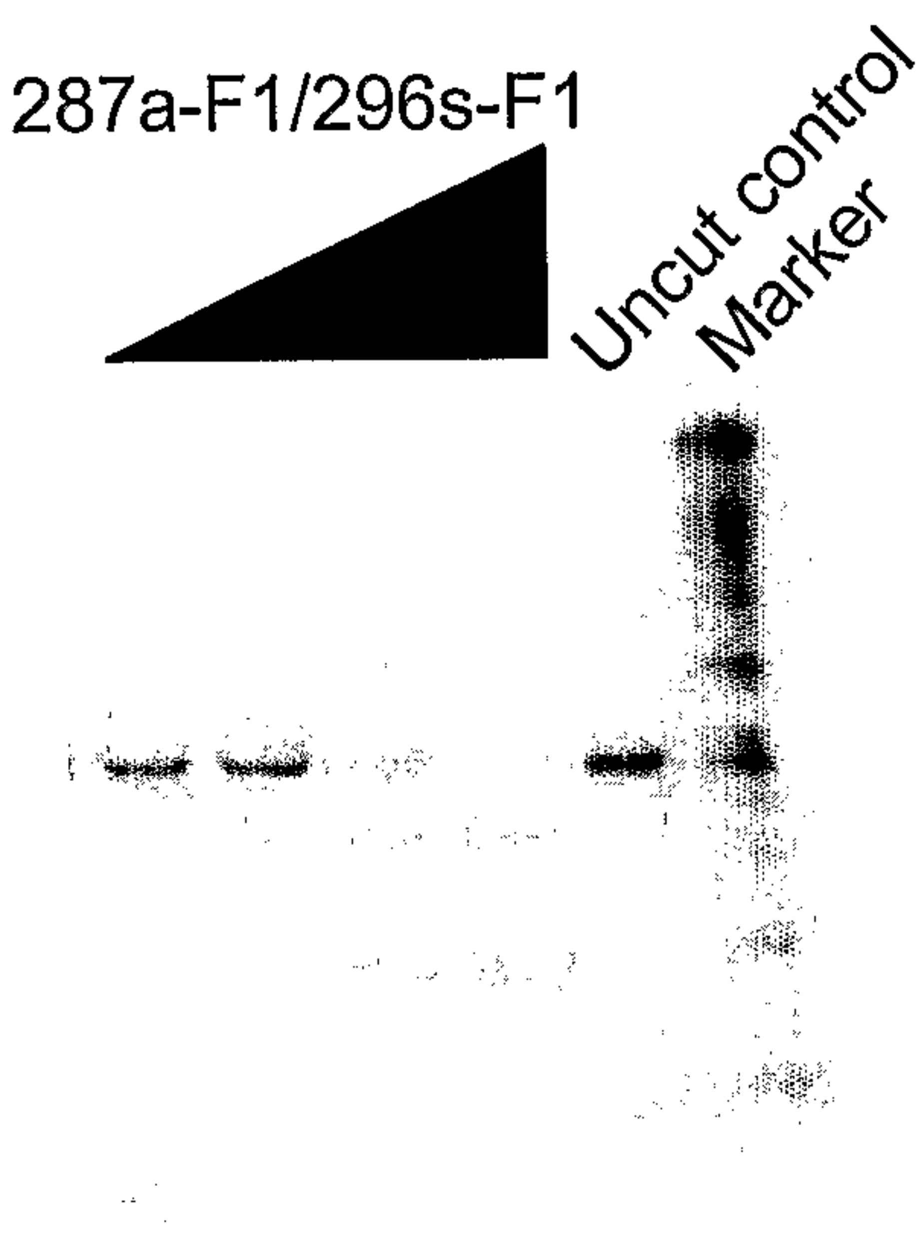


FIGURE 26

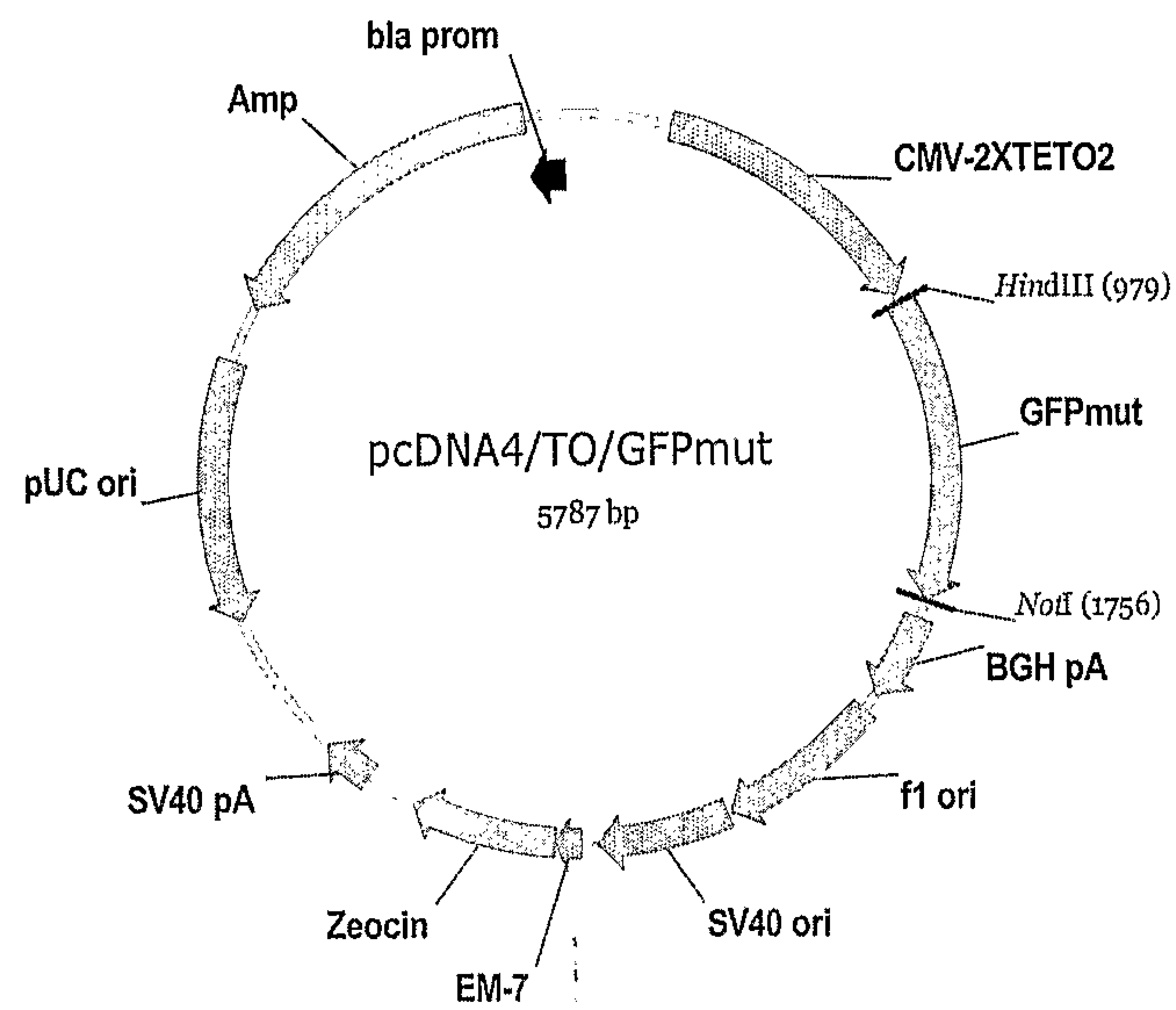

bla prom
Amp
CMV-2XTETO2
HindIII (979)
GFPmut
pcDNA4/TO/GFPmut
5787 bp
pUC ori
NotI (1756)
BGH pA
f1 ori
SV40 pA
Zeocin
EM-7
SV40 ori

FIGURE 27

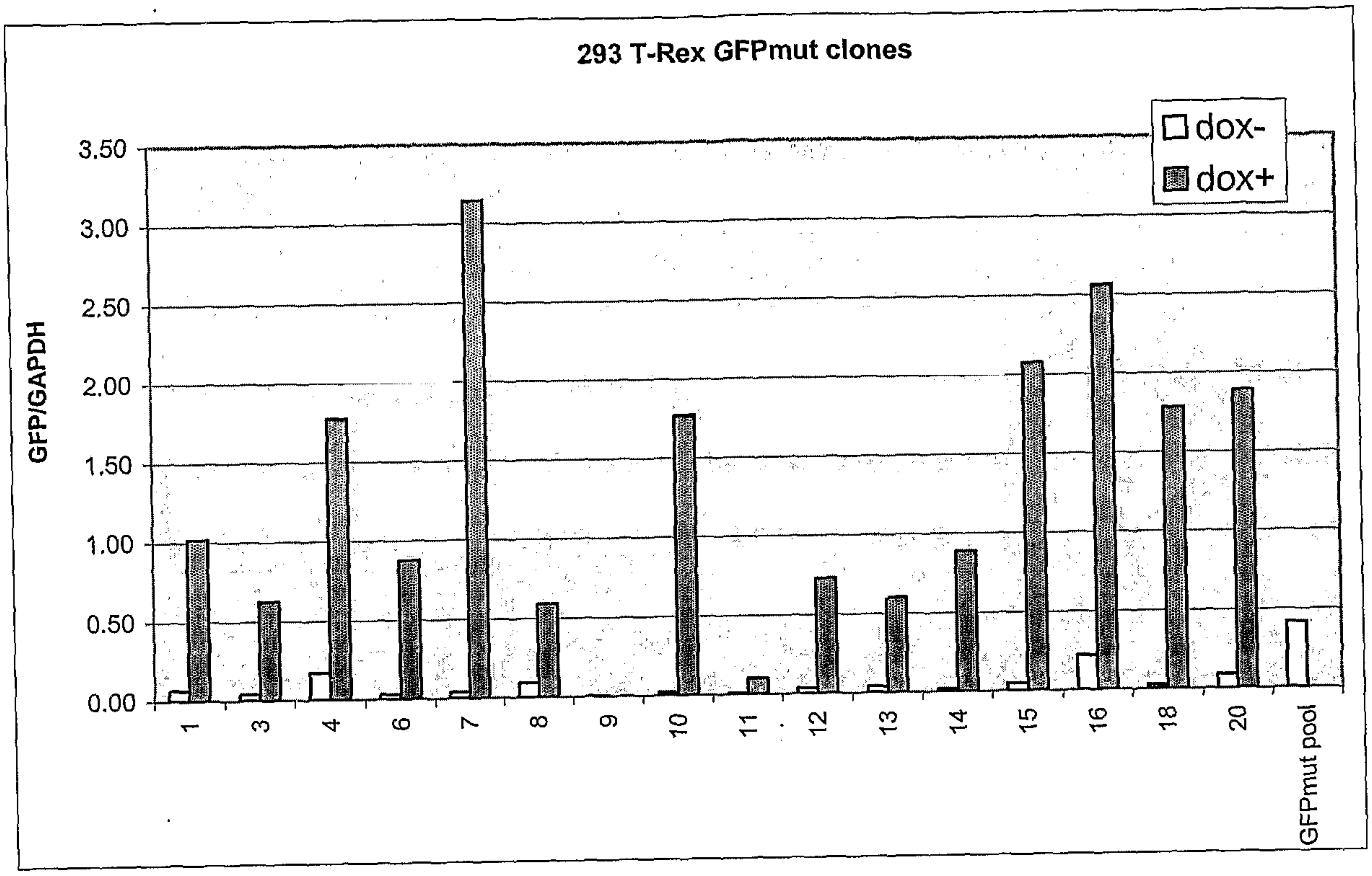

293 T-Rex GFPmut clones
dox-
dox+
GFP/GAPDH
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1
3
4
6
7
8
9
10
11
12
13
14
15
16
18
20
GFPmut pool

FIGURE 28

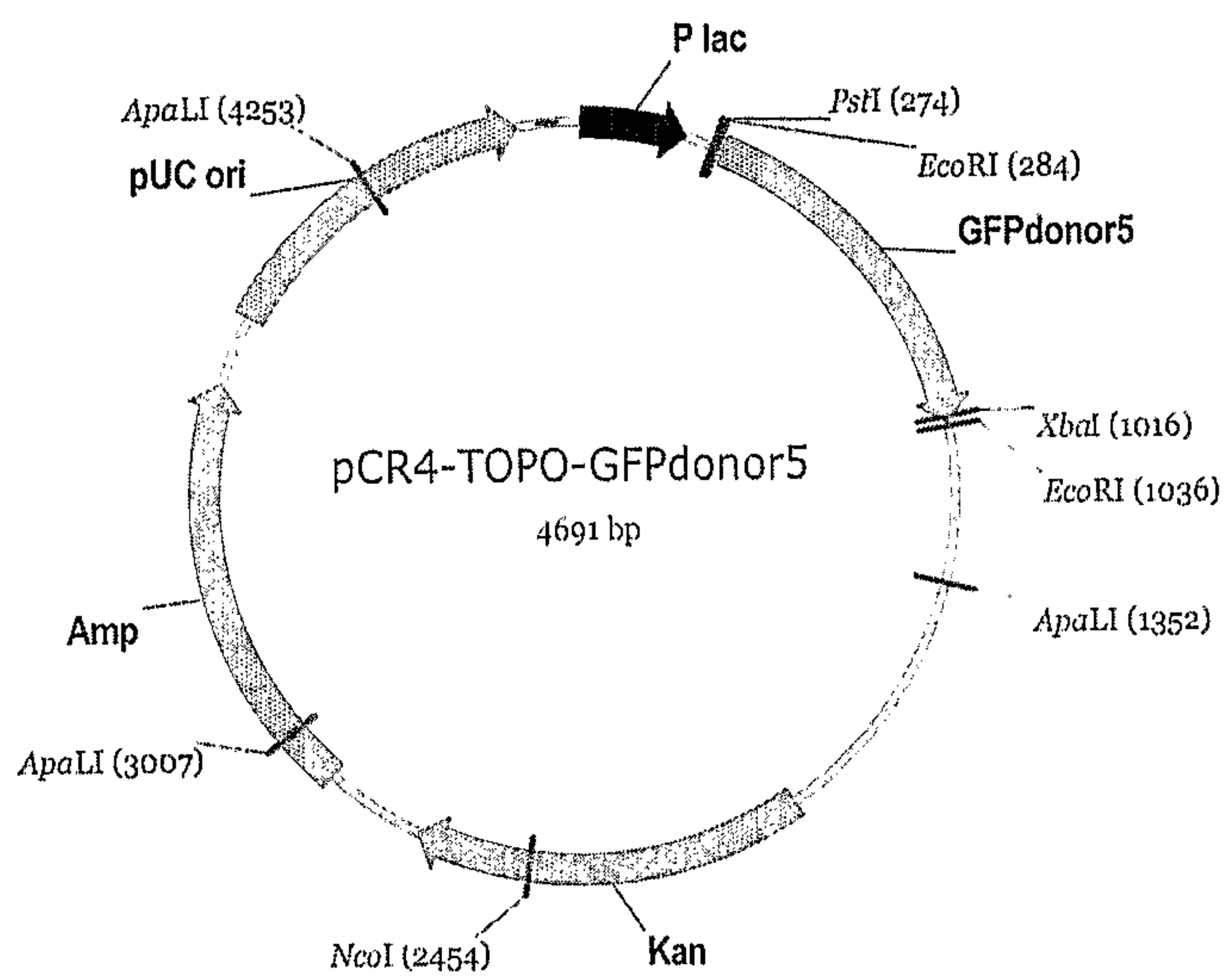

P lac
PstI (274)
EcoRI (284)
GFPdonor5
XbaI (1016)
EcoRI (1036)
ApaLI (1352)
Kan
NcoI (2454)
ApaLI (3007)
Amp
pUC ori
ApaLI (4253)
pCR4-TOPO-GFPdonor5
4691 bp

FIGURE 29

GGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC
GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGAC
AAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATC

(SEQ ID NO:20)

FIGURE 30

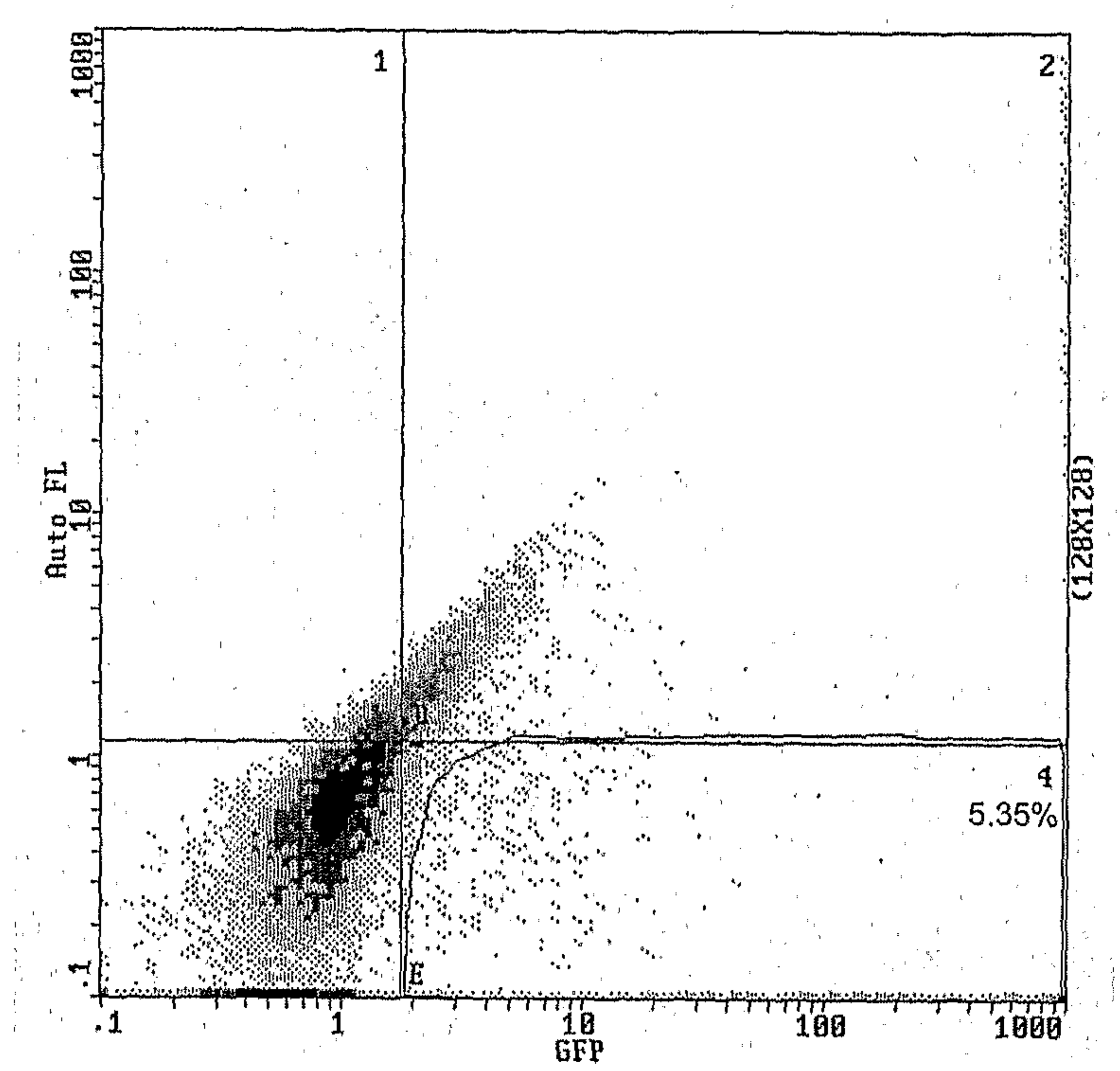


FIGURE 31

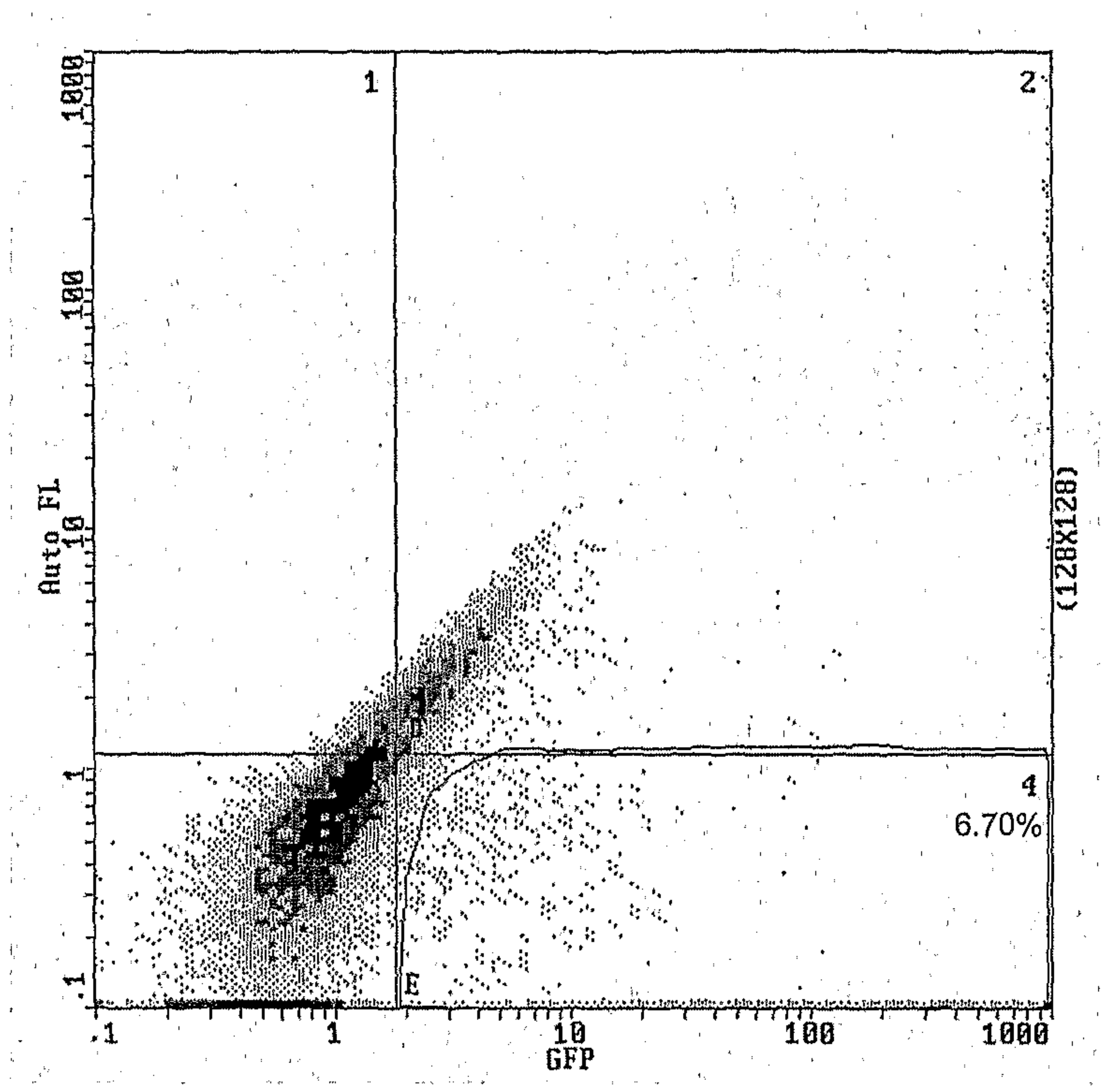


FIGURE 32

GGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAG
CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC
GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGAC
AAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGC
AGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTG
CTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAAC
GAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAAAGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAA
CCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCT
CCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAA
ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG
TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATC
CCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCG
TGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCT
TTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAG
GGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATG
GTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGT
CCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAA
ATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTT
AGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCA

(SEQ ID NO:21)

FIGURE 33

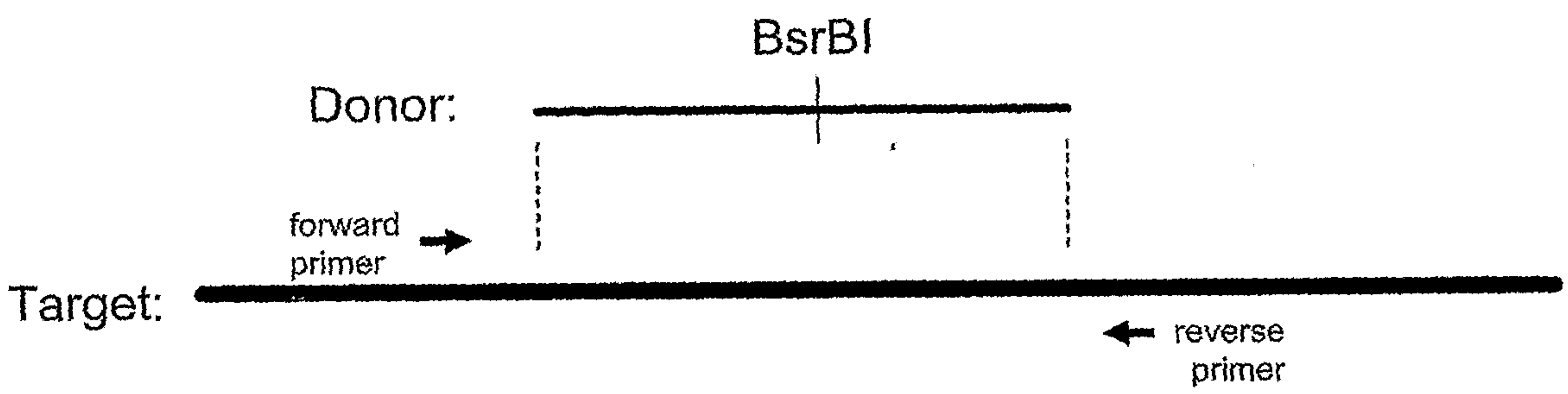


FIGURE 34

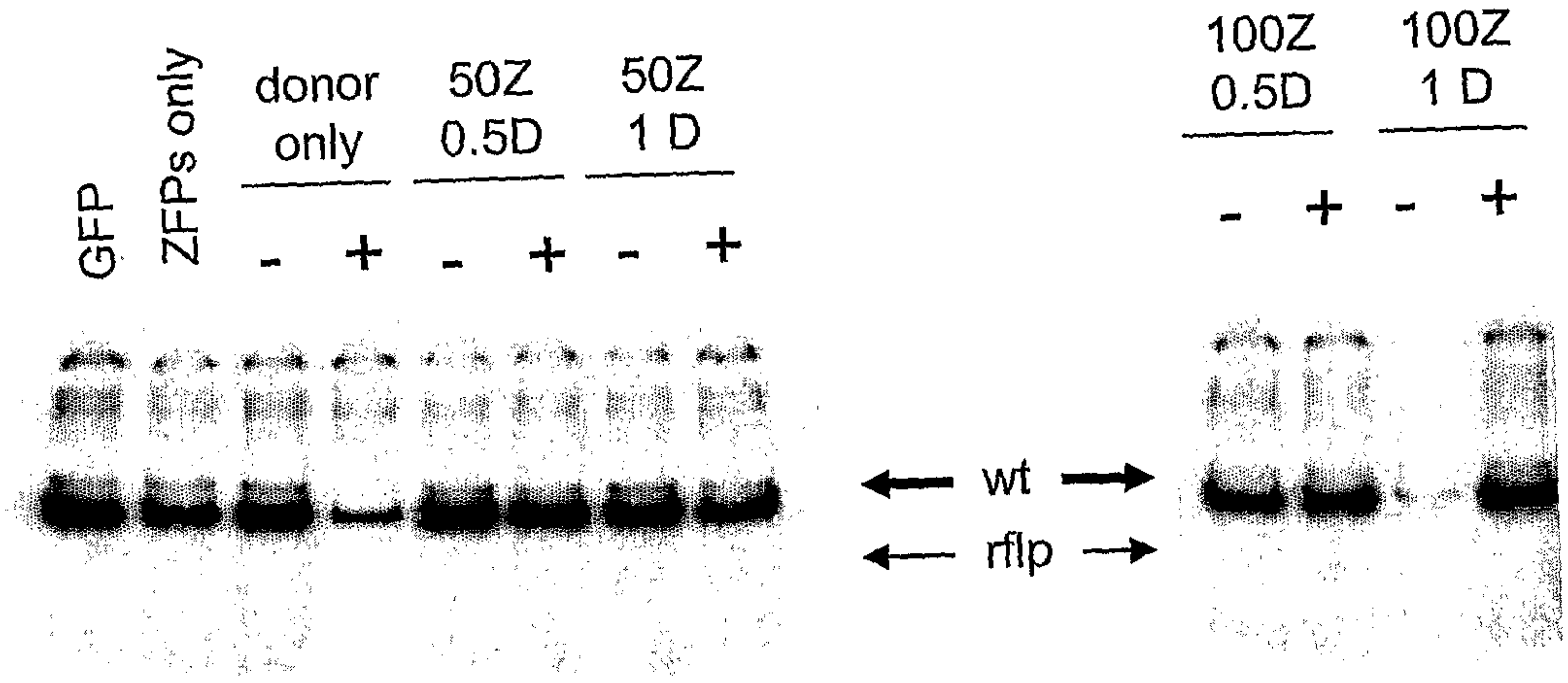

GFP
ZFPs only
donor only
50Z 0.5D
50Z 1 D
100Z 0.5D
100Z 1 D
- + - + - +
- + - +
wt
rflp

FIGURE 35

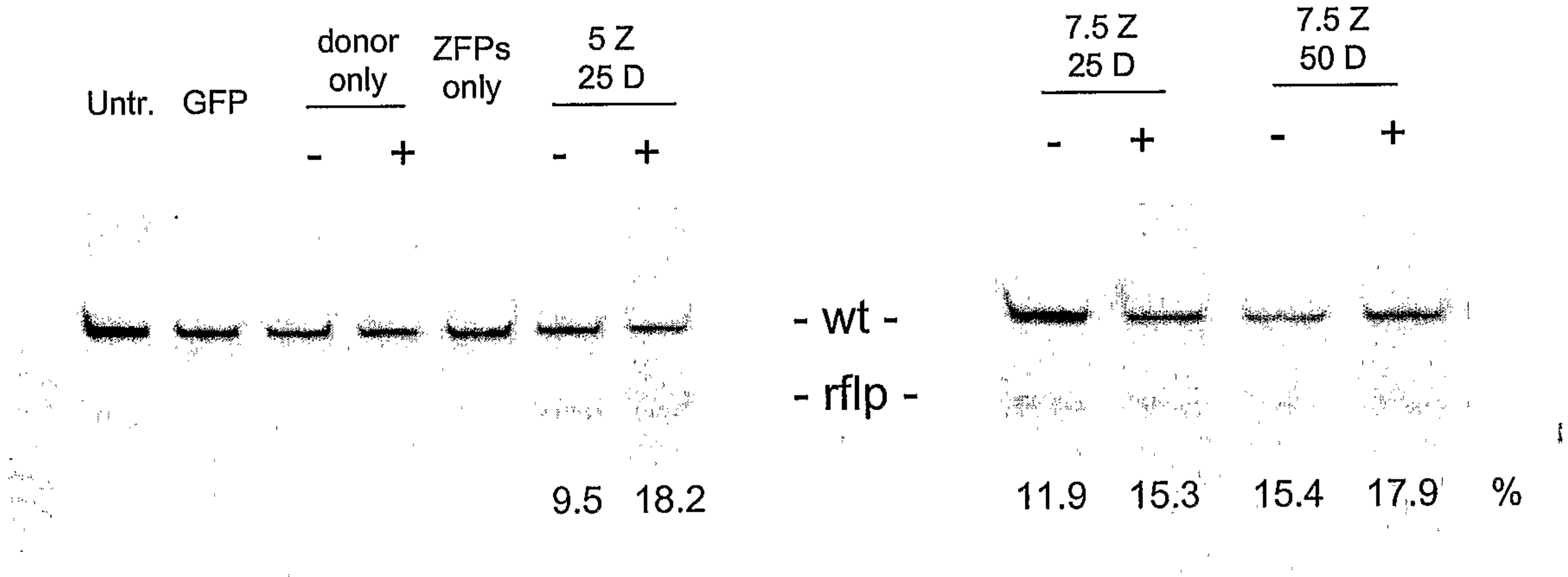

Untr.
GFP
donor only
- +
ZFPs only
5 Z 25 D
- +
7.5 Z 25 D
- +
7.5 Z 50 D
- +
- wt -
- rflp -
9.5
18.2
11.9
15.3
15.4
17.9
%

FIGURE 36

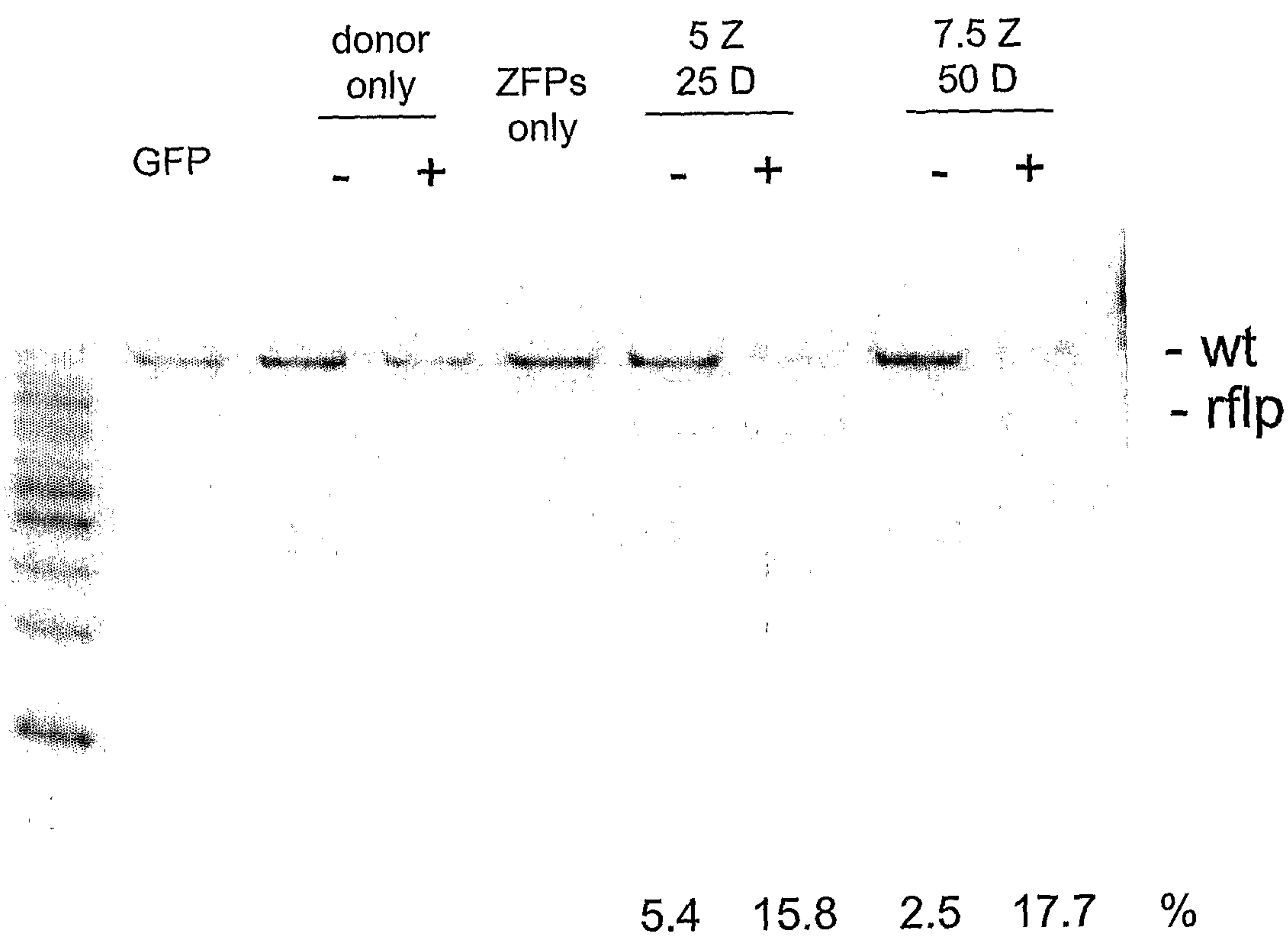


FIGURE 37

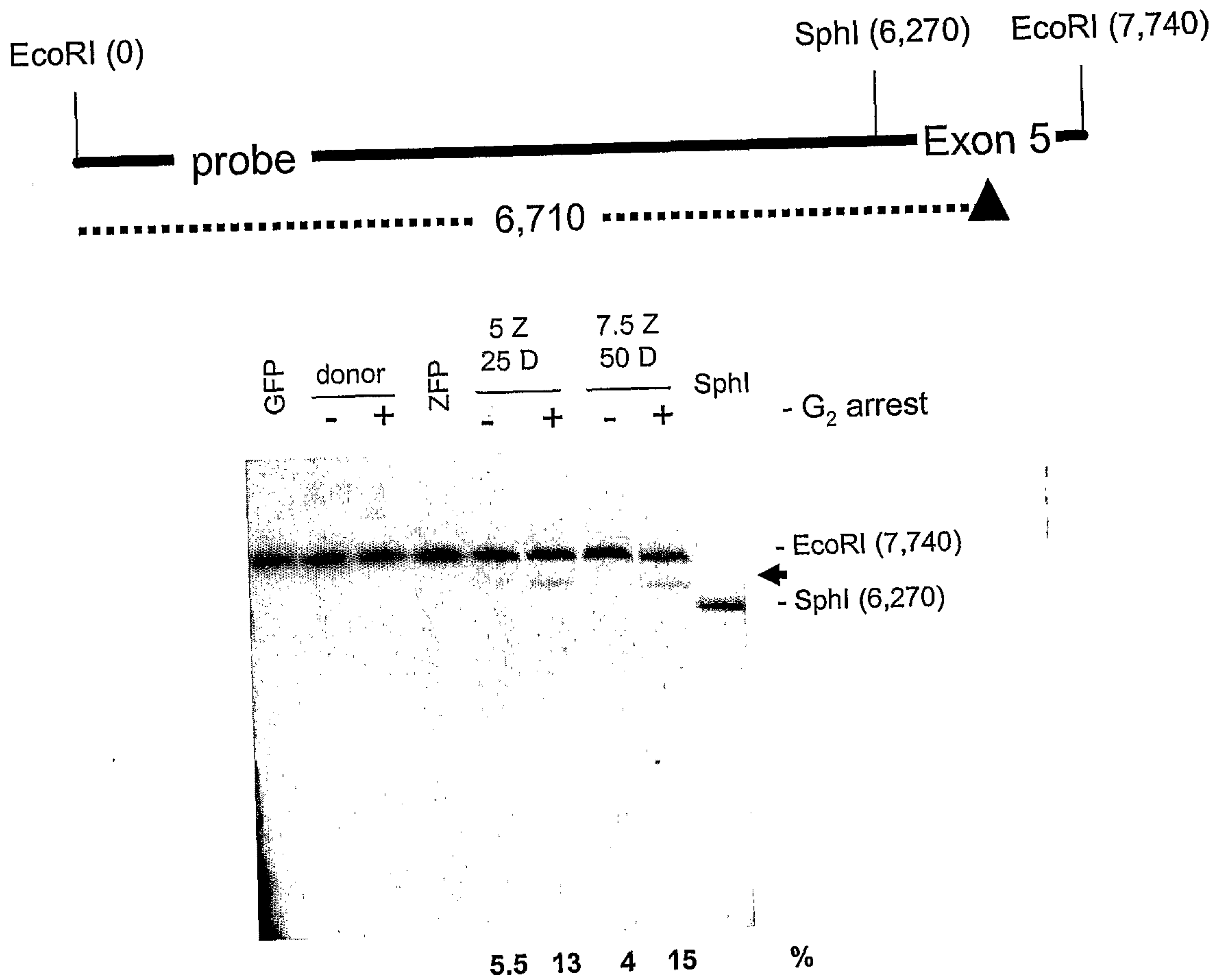

EcoRI (0)
SphI (6,270)
EcoRI (7,740)
probe
Exon 5
6,710
GFP
donor
- +
ZFP
5 Z
25 D
- +
7.5 Z
50 D
- +
SphI
- G_2 arrest
- EcoRI (7,740)
- SphI (6,270)
5.5 13 4 15 %

FIGURE 38

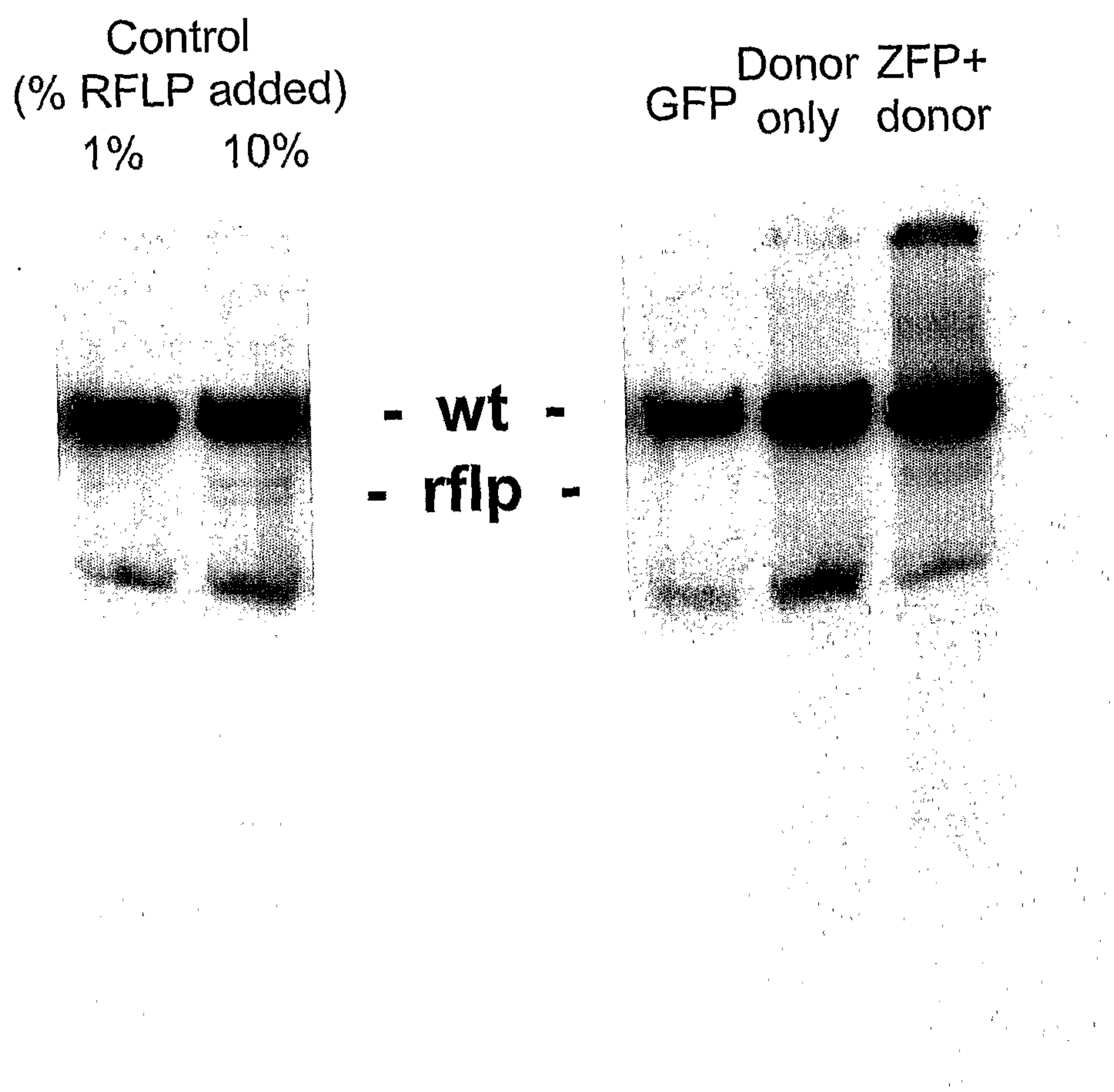


FIGURE 39

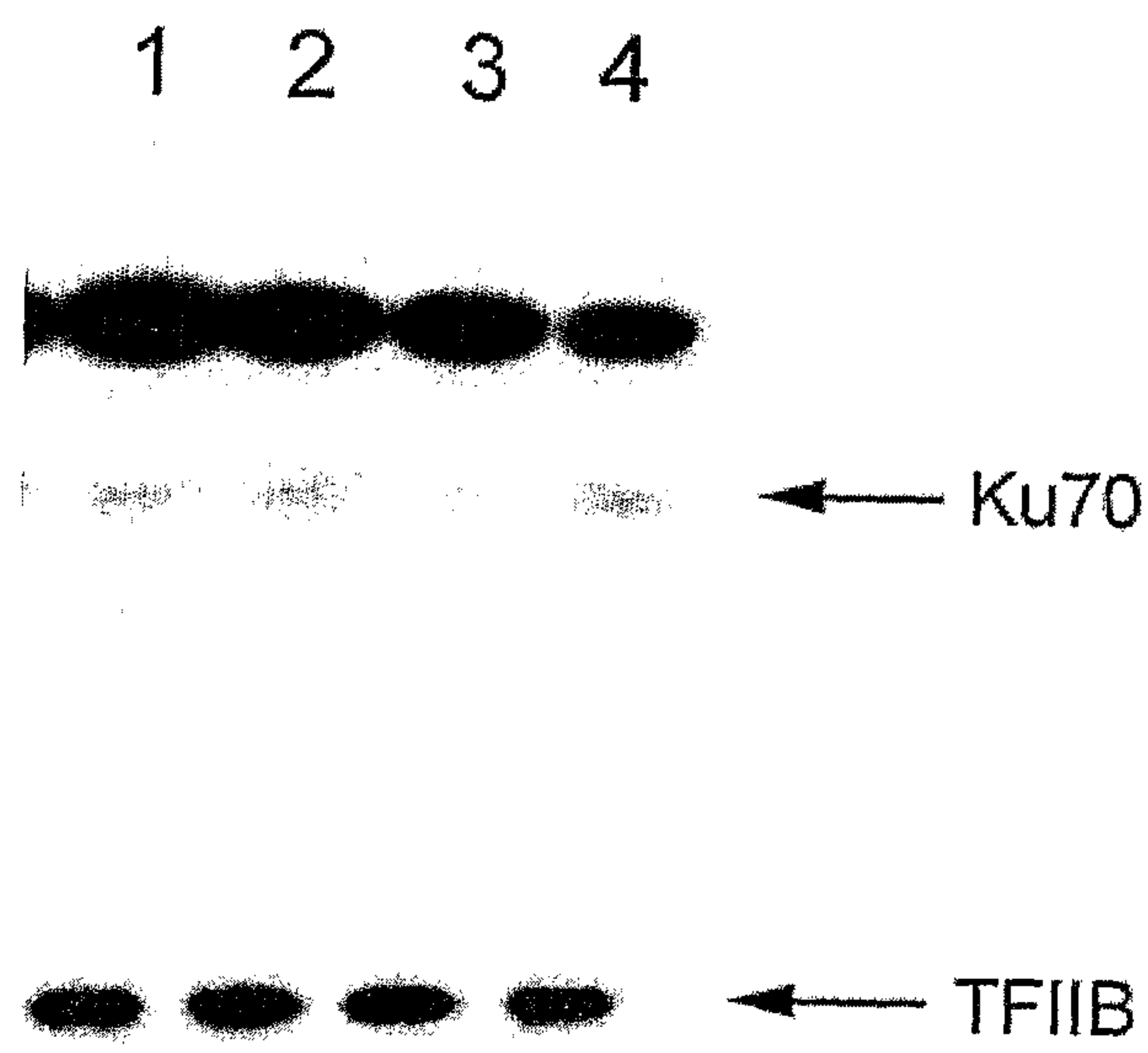


FIGURE 40

sca-29b:

MAERPFQCRICMRNFS<u>QSGDLTR</u>HIRTHTGEKPFACDICGRKFA<u>TSANLSR</u>
HTKIHTGGGGSQKPFQCRICMRNFS<u>DRSALSR</u>HIRTHTGEKPFACDICGRKF
A<u>QSGHLSR</u>HTKIH (SEQ ID NO:22)

sca-36a:

MAERPFQCRICMRNFS<u>RSQTRKT</u>HIRTHTGEKPFACDICGRKFA<u>QKRNRTK</u>
HTKIHTGSQKPFQCRICMRNFS<u>DRSALSR</u>HIRTHTGEKPFACDICGRKFA<u>QS</u>
<u>GNLAR</u>HTKIH (SEQ ID NO:23)

sca-36b:

MAERPFQCRICMRNFS<u>TSGSLSR</u>HIRTHTGEKPFACDICGRKFA<u>DRSDLSR</u>
HTKIHTGGGGSQKPFQCRICMRNFS<u>DRSALSR</u>HIRTHTGEKPFACDICGRKF
A<u>QSGNLAR</u>HTKIH (SEQ ID NO:24)

sca-36c:

MAERPFQCRICMRNFS<u>TSSSLSR</u>HIRTHTGEKPFACDICGRKFA<u>DRSDLSR</u>H
TKIHTGGGGSQKPFQCRICMRNFS<u>DRSALSR</u>HIRTHTGEKPFACDICGRKFA
<u>QSGNLAR</u>HTKIH (SEQ ID NO:25)

FIGURE 41

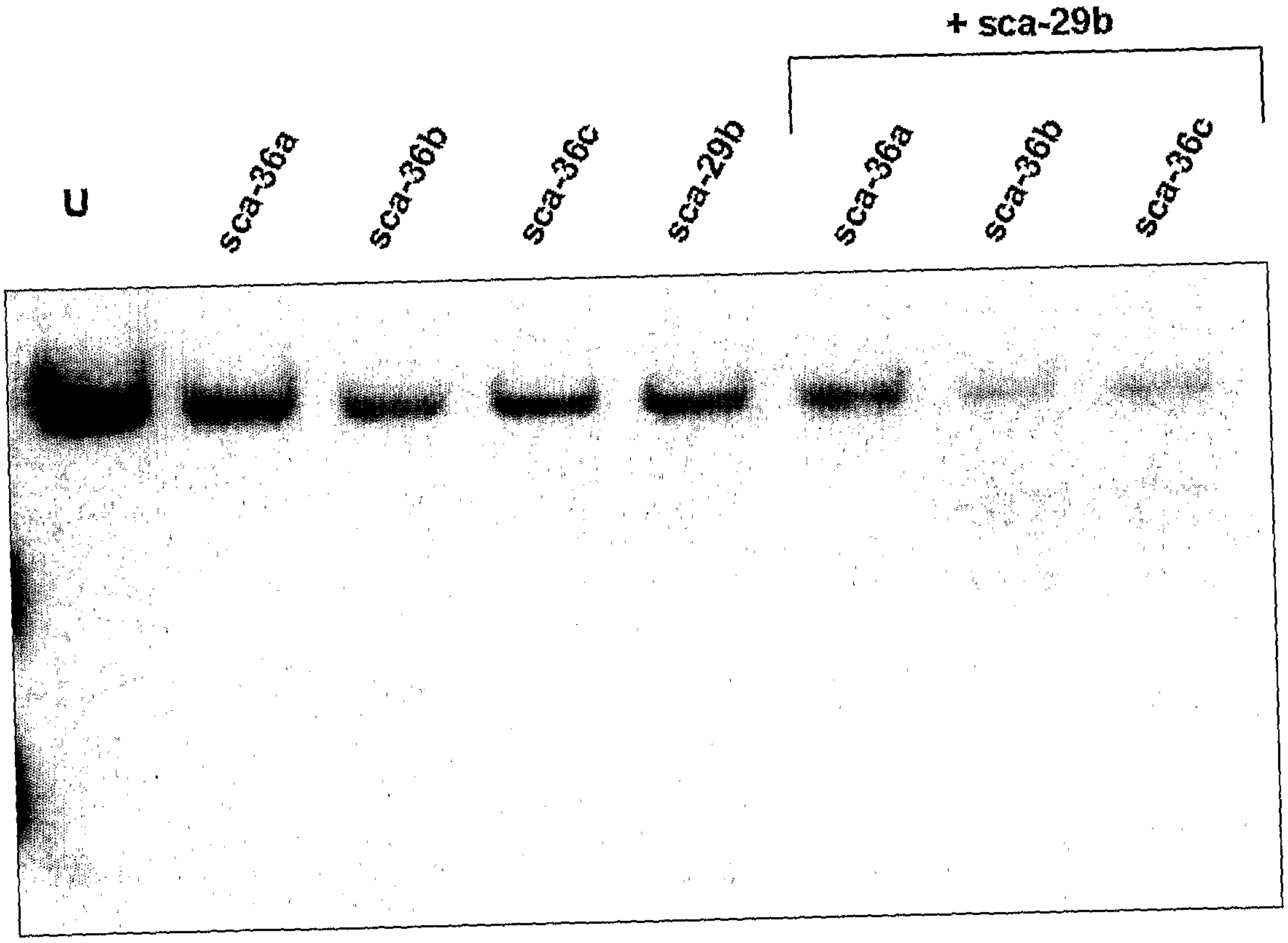


FIGURE 42

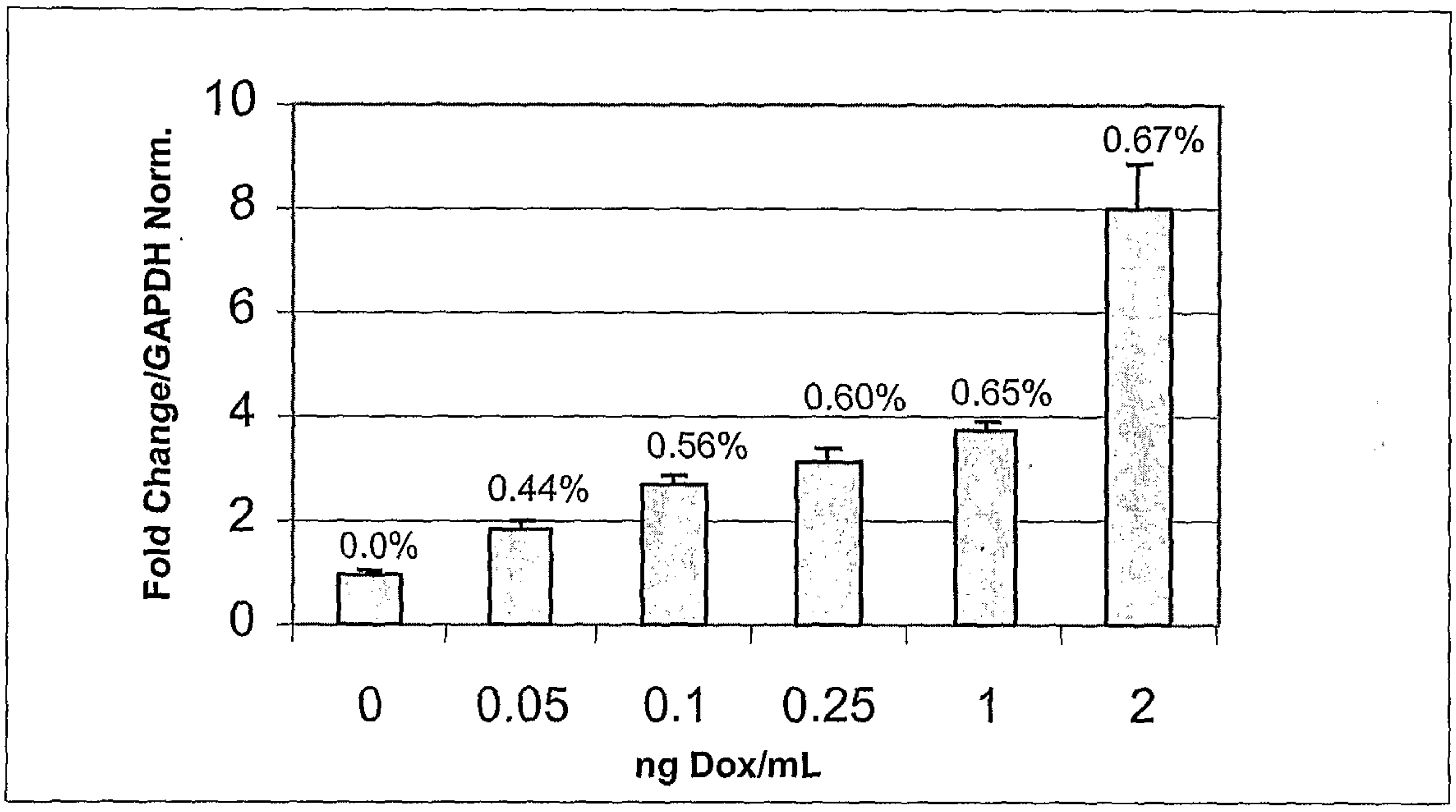

10
8
6
4
2
0
Fold Change/GAPDH Norm.
0.0%
0.44%
0.56%
0.60%
0.65%
0.67%
0
0.05
0.1
0.25
1
2
ng Dox/mL

FIGURE 43

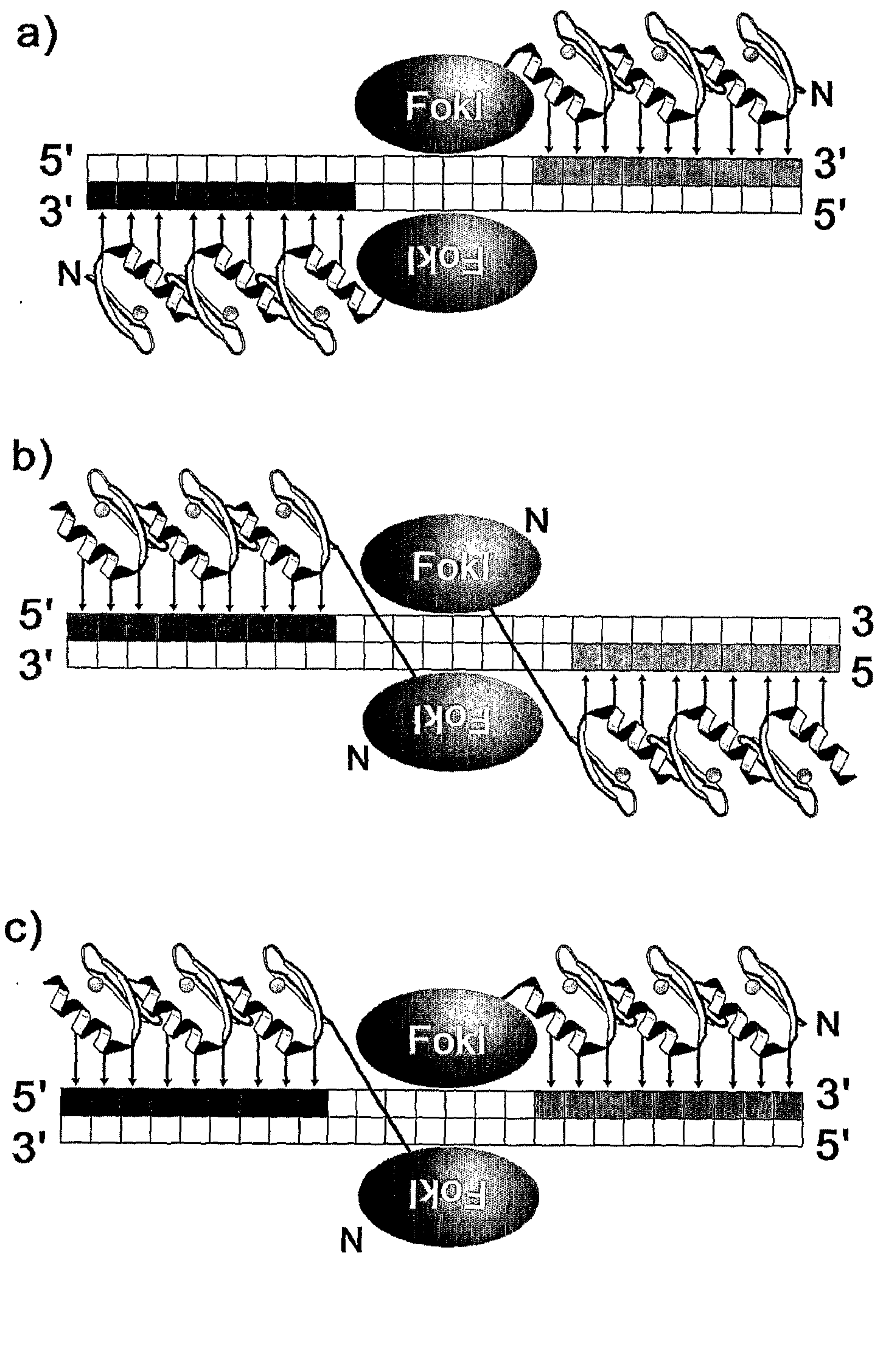
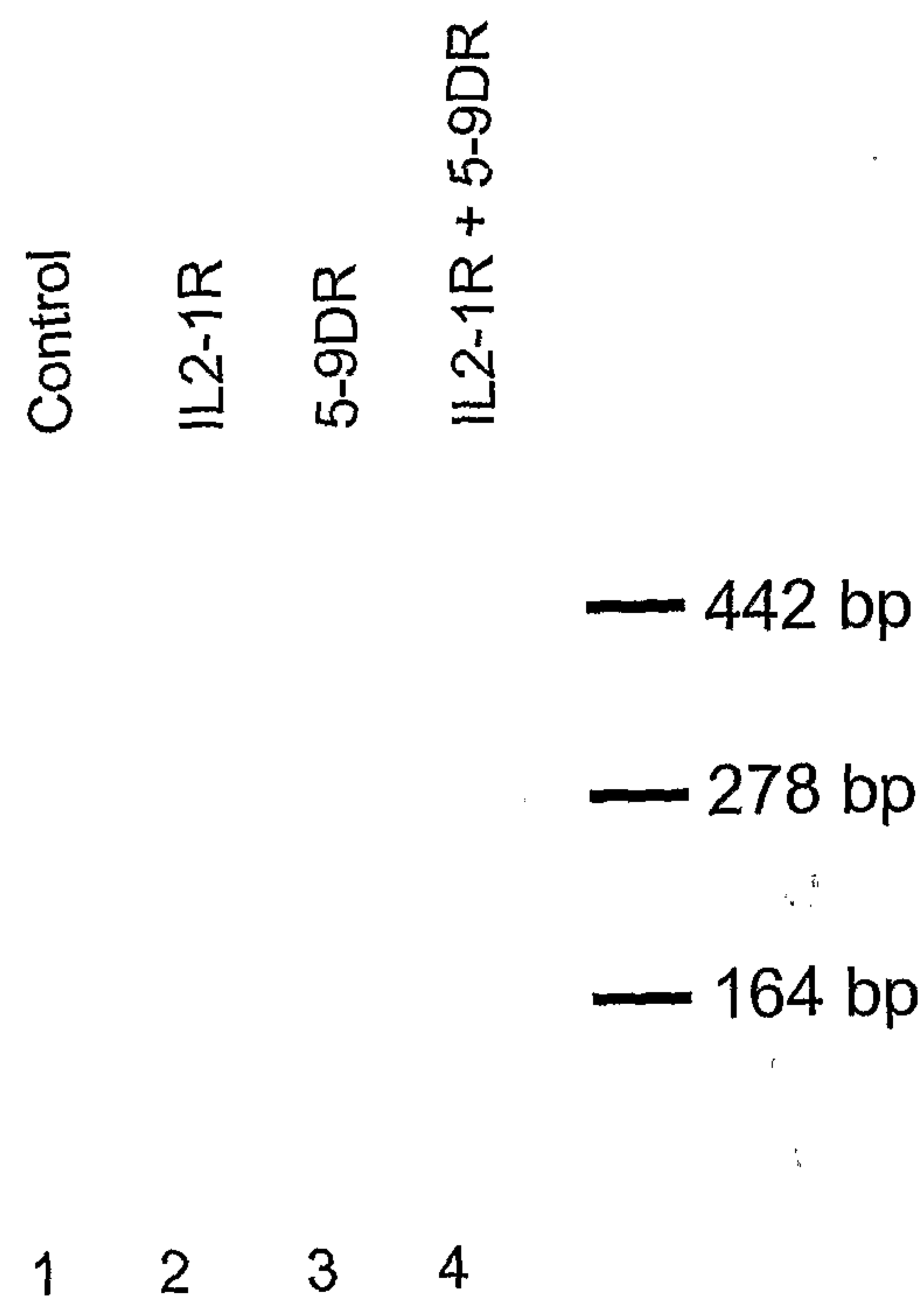


FIGURE 44



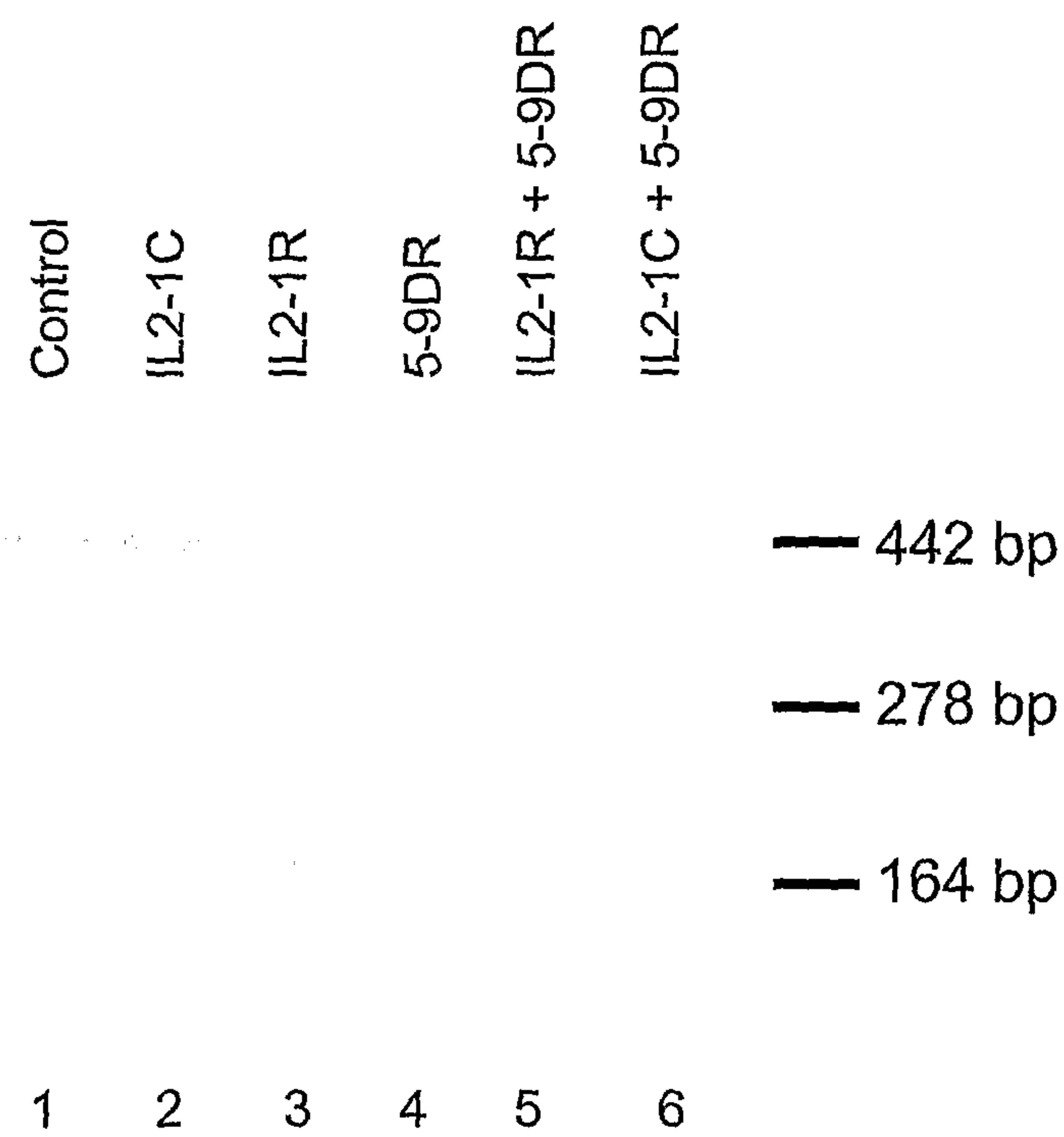
59D

```
-tAAAGCGGCTCCGaaccacATTATCCGAGTTc-
-aTTTCGCCGAGGCttggtgTAATAGGCTCAAg-
```

IL2-1

(SEQ ID NO: 211)

FIGURE 45



59D **IL2-1**

-t**AAAGCGGCTCCG**aaccac**AACTCGGATAAT**g-
-aTTTCGCCGAGGCttggtgTTGAGCCTATTAc-

(SEQ ID NO: 212)

FIGURE 46

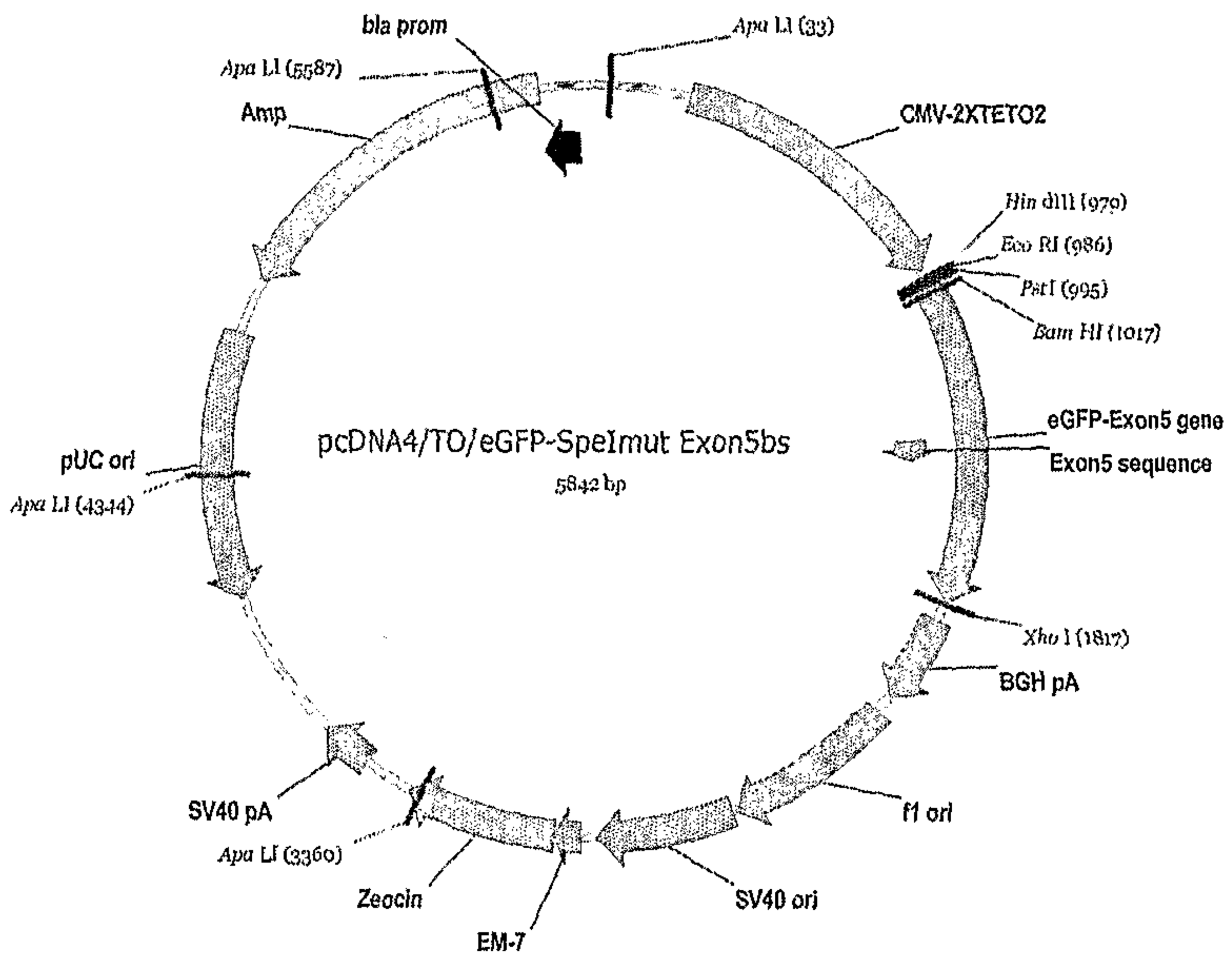

bla prom
Apa LI (5587)
Amp
Apa LI (33)
CMV-2XTETO2
Hin dIII (979)
Eco RI (986)
Pst I (995)
Bam HI (1017)
eGFP-Exon5 gene
Exon5 sequence
pcDNA4/TO/eGFP-SpeImut Exon5bs
5842 bp
pUC ori
Apa LI (4344)
Xho I (1817)
BGH pA
f1 ori
SV40 pA
Apa LI (3360)
Zeocin
EM-7
SV40 ori

FIGURE 47

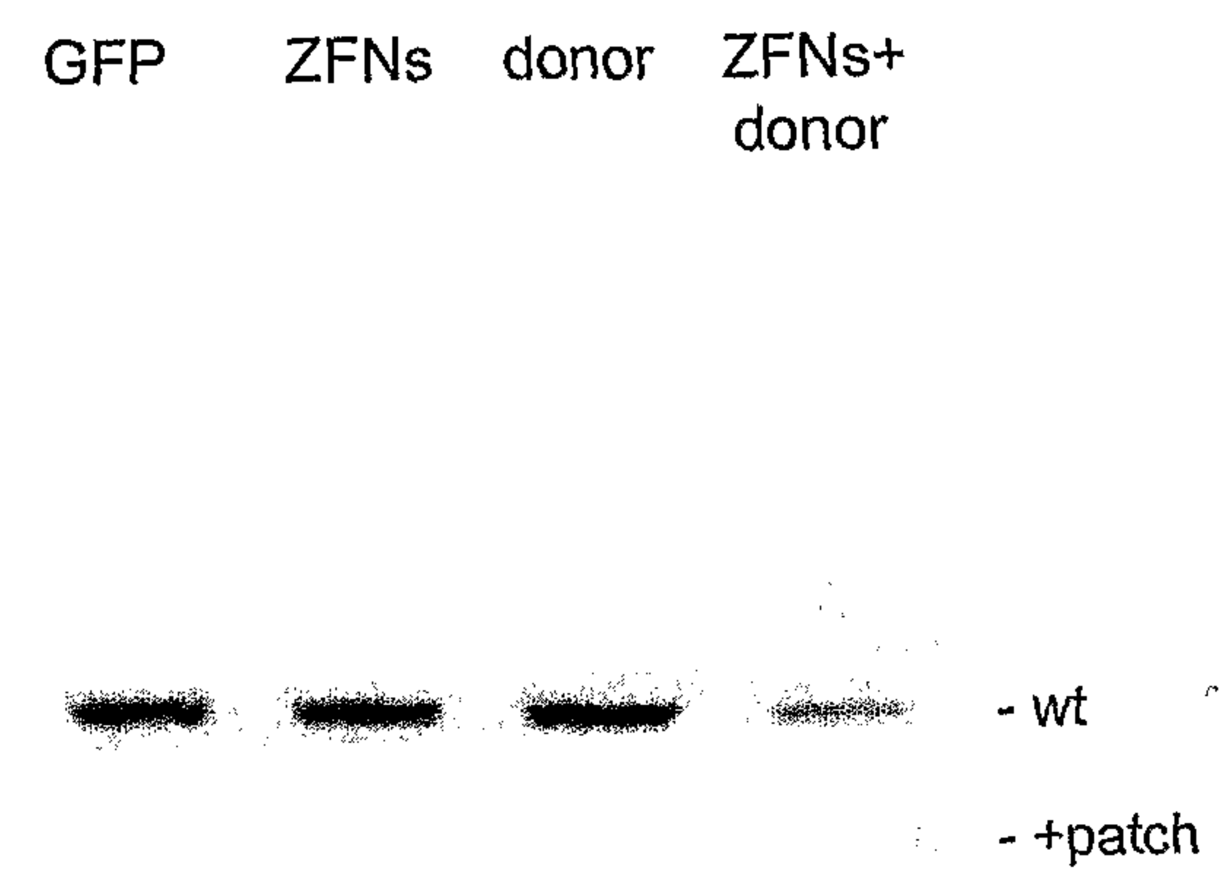


FIGURE 48

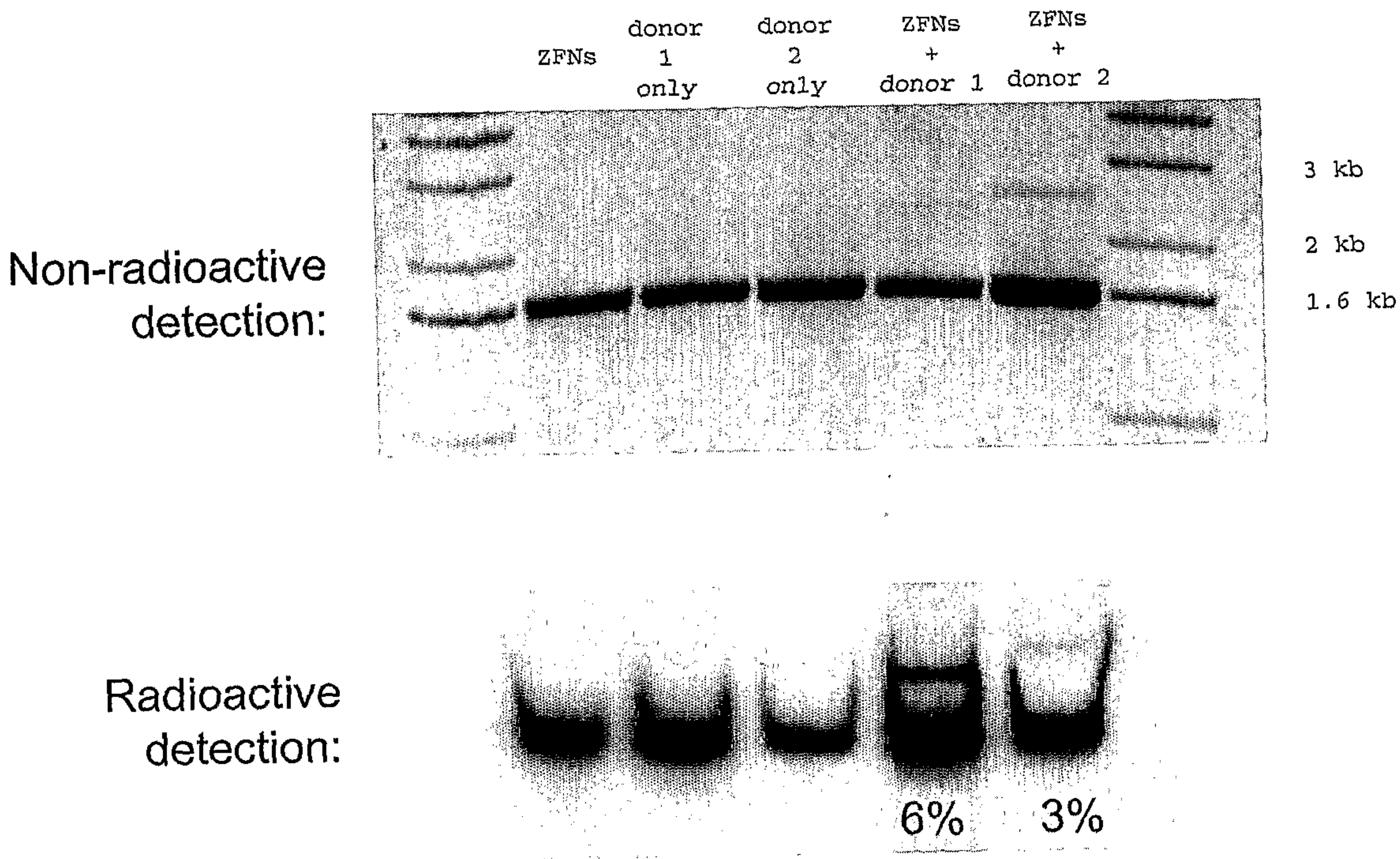

ZFNs
donor 1 only
donor 2 only
ZFNs + donor 1
ZFNs + donor 2
3 kb
2 kb
1.6 kb
Non-radioactive detection:
Radioactive detection:
6%
3%

FIGURE 49

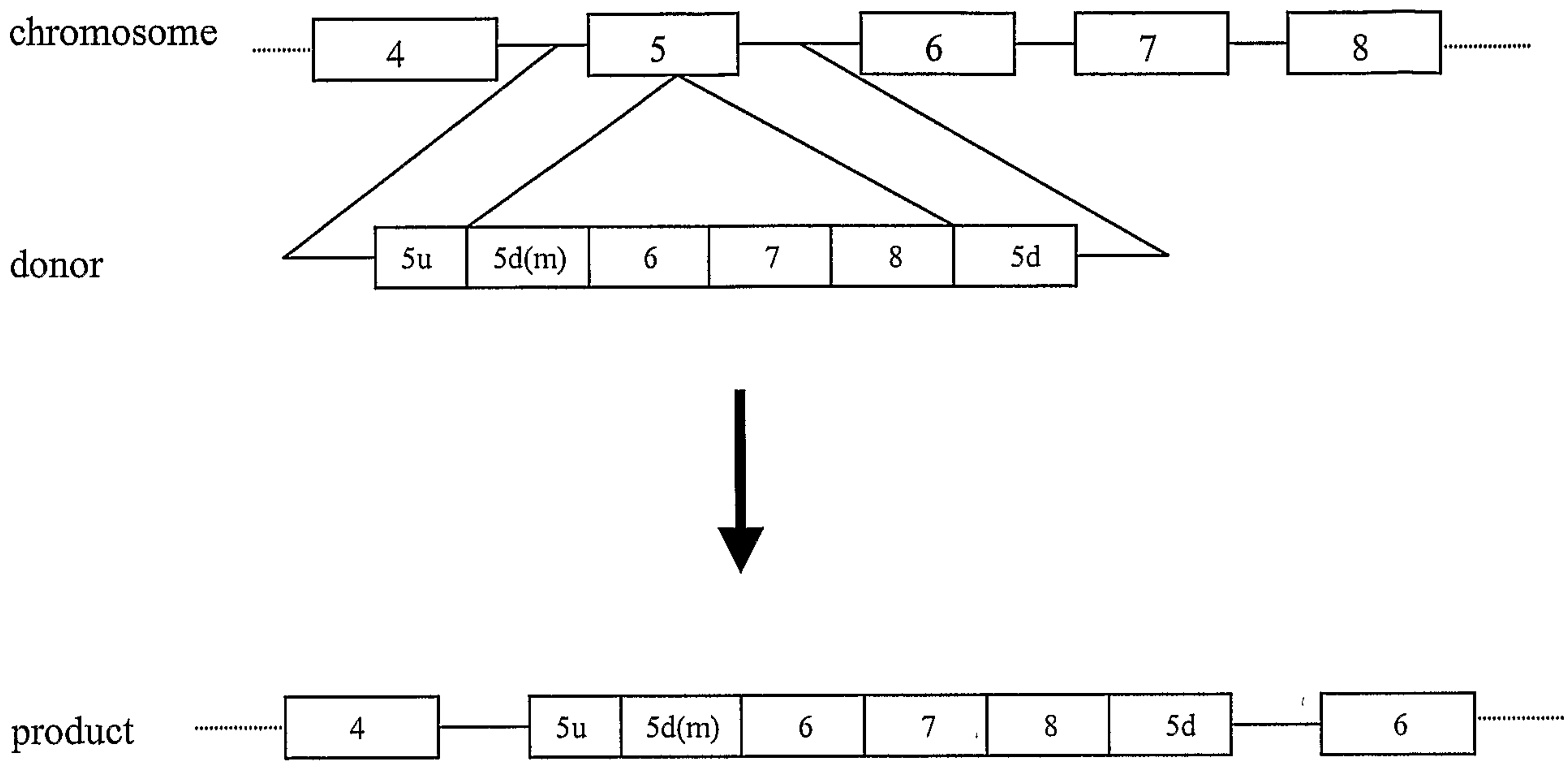


FIGURE 50

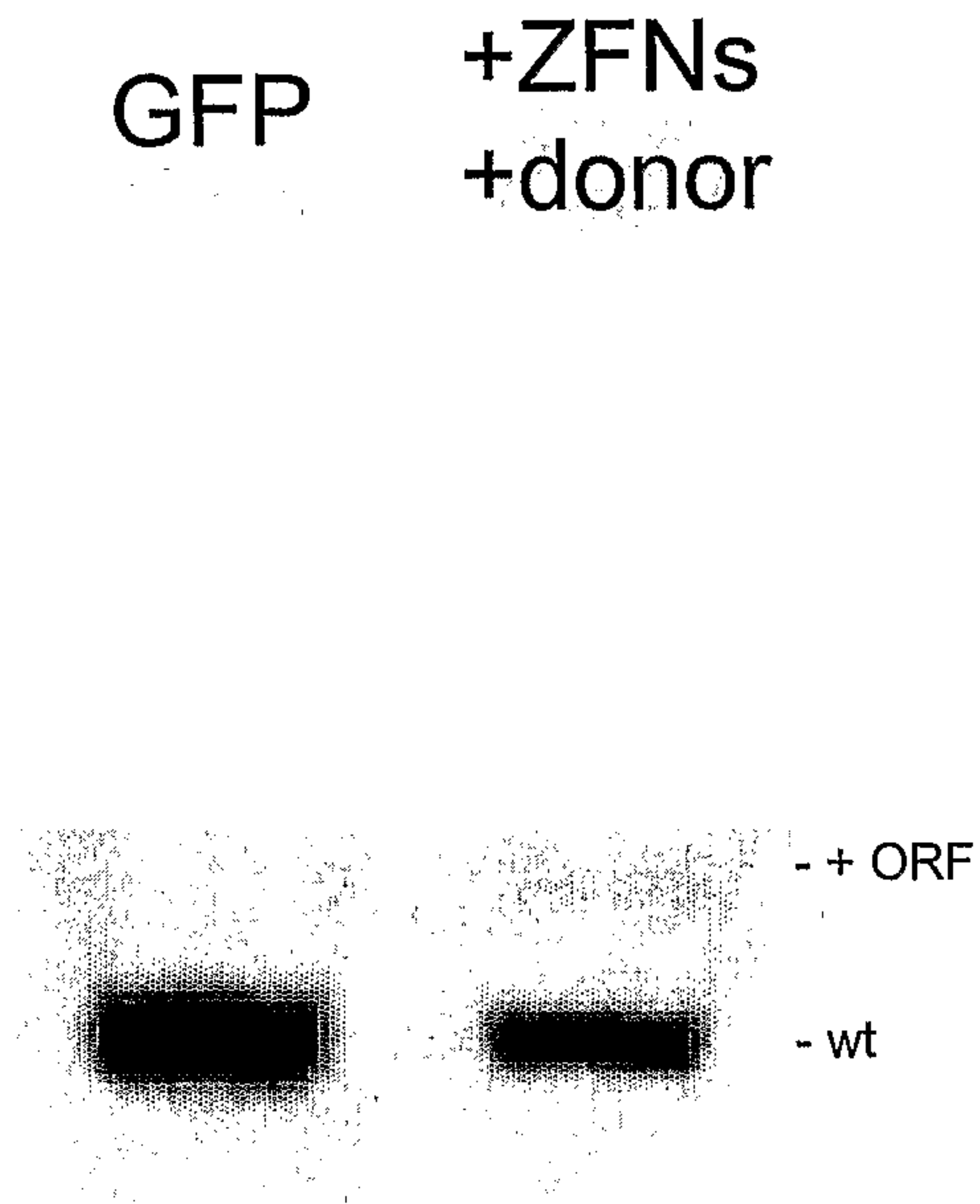


FIGURE 51

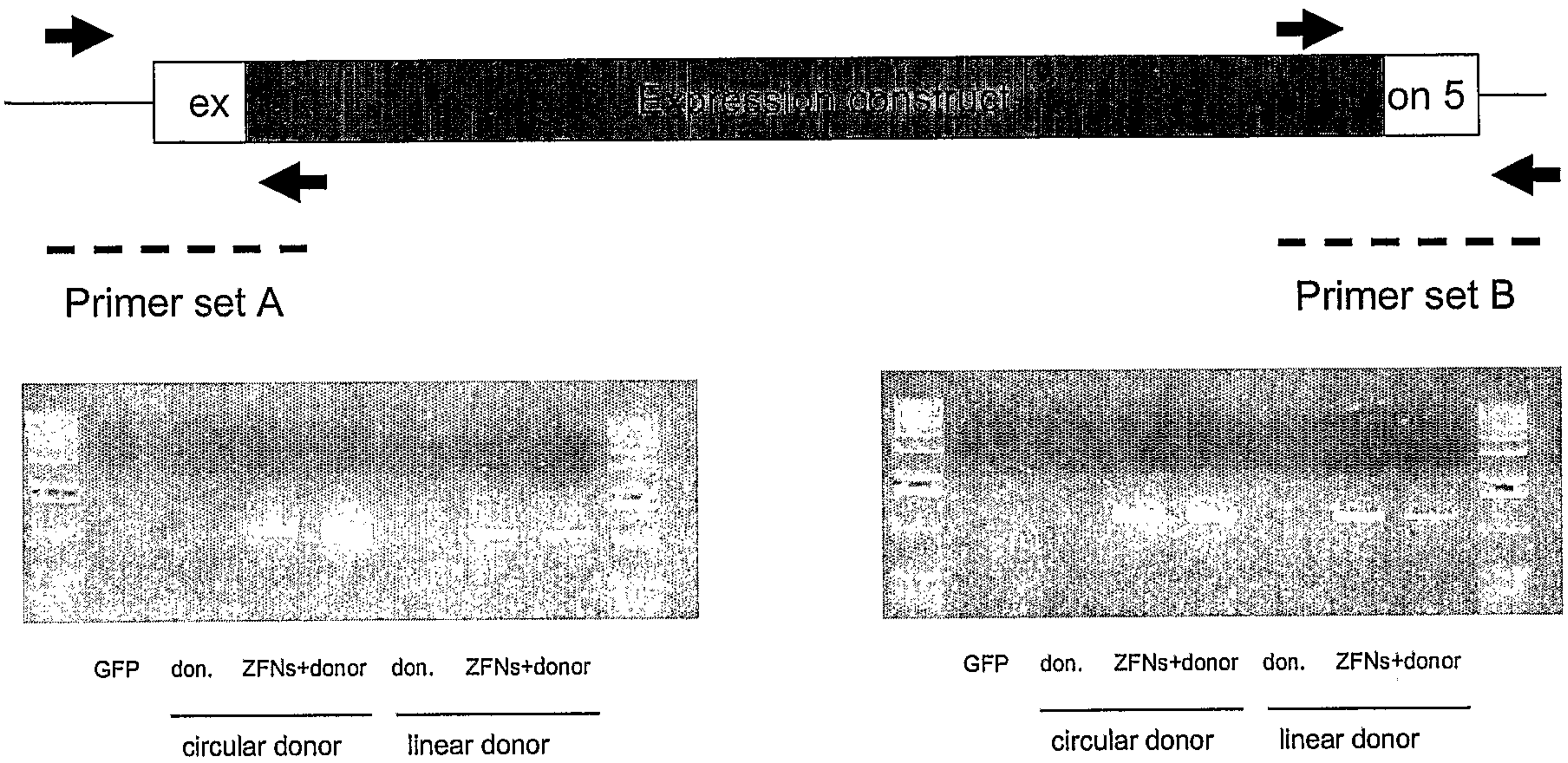

ex
Expression construct
on 5
Primer set A
Primer set B
GFP
don.
ZFNs+donor
don.
ZFNs+donor
circular donor
linear donor
GFP
don.
ZFNs+donor
don.
ZFNs+donor
circular donor
linear donor

FIGURE 52

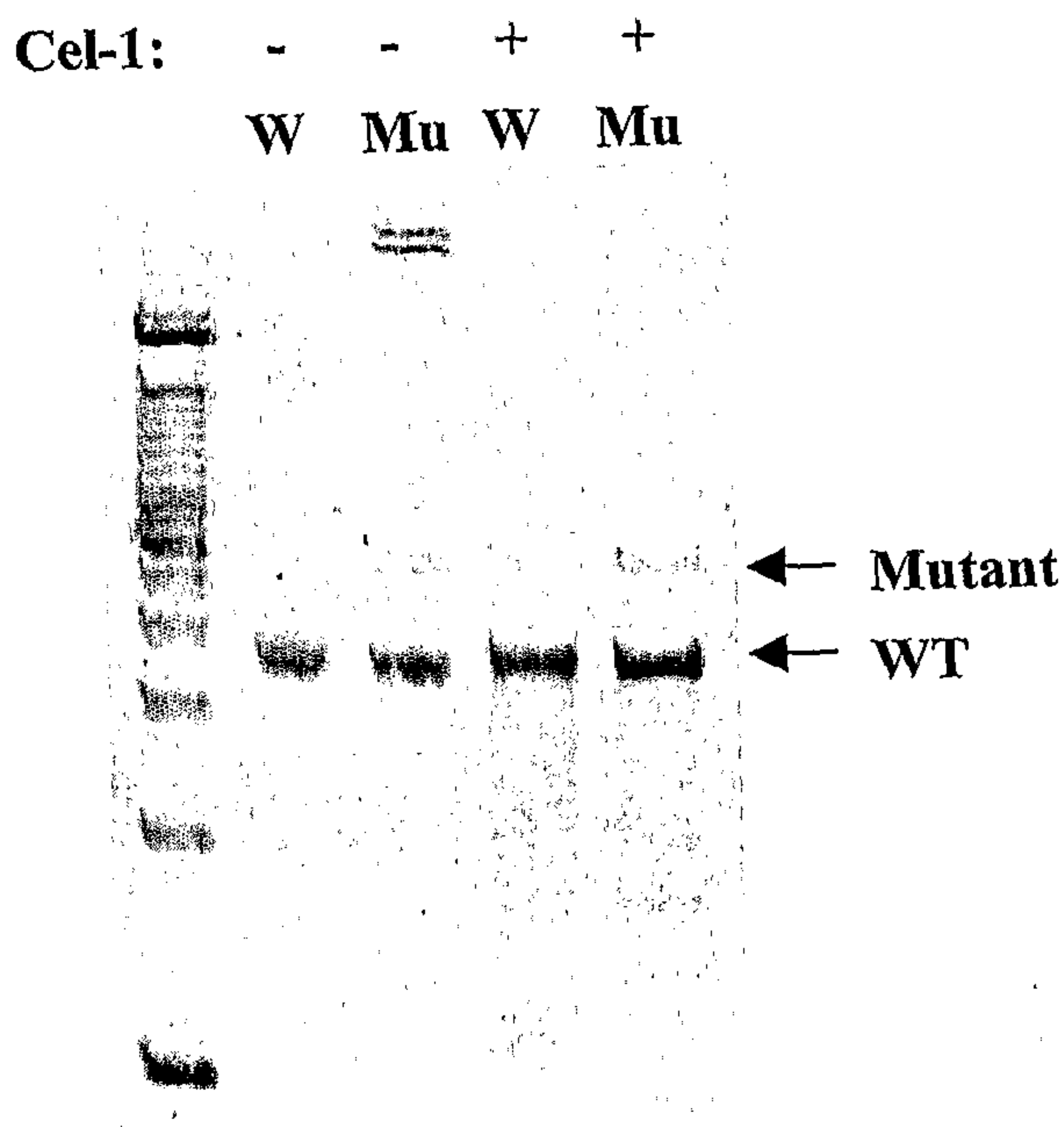


FIGURE 53

```
CTCGGTTTCC CTAACCCAAT CCAGCCAGTA CCTGAGCATT G: ::::::: ::::::::::
CTCGGTTTCC CTAACCCAAT CCAGCCAGTA CCTGAGCATT GTGGCACCTT CCAGGGTCAA

:::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
GGAAGGCACG GGGGAGGGGC AAACAACAGA TGGCTGGCAA CTAGAAGGCA CAGTCGAGGC

:::::::::: :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
TGATCAGCGG TTTAAACTTA AGCTTGGTAC CGAGCTCGGA TCCACTTTGA CGTCAATGGG

:::::::::: :::::::GCC AGGGAAGGTC TCCGTTCTTG CCGATGCCCA TATTCTGGGA(SEQ ID NO:233)
TGGAGTATTT ACGGTAAACC AGGGAAGGTC TCCGTTCTTG CCGATGCCCA TATTCTGGGA(SEQ ID NO:234)
```

FIGURE 54

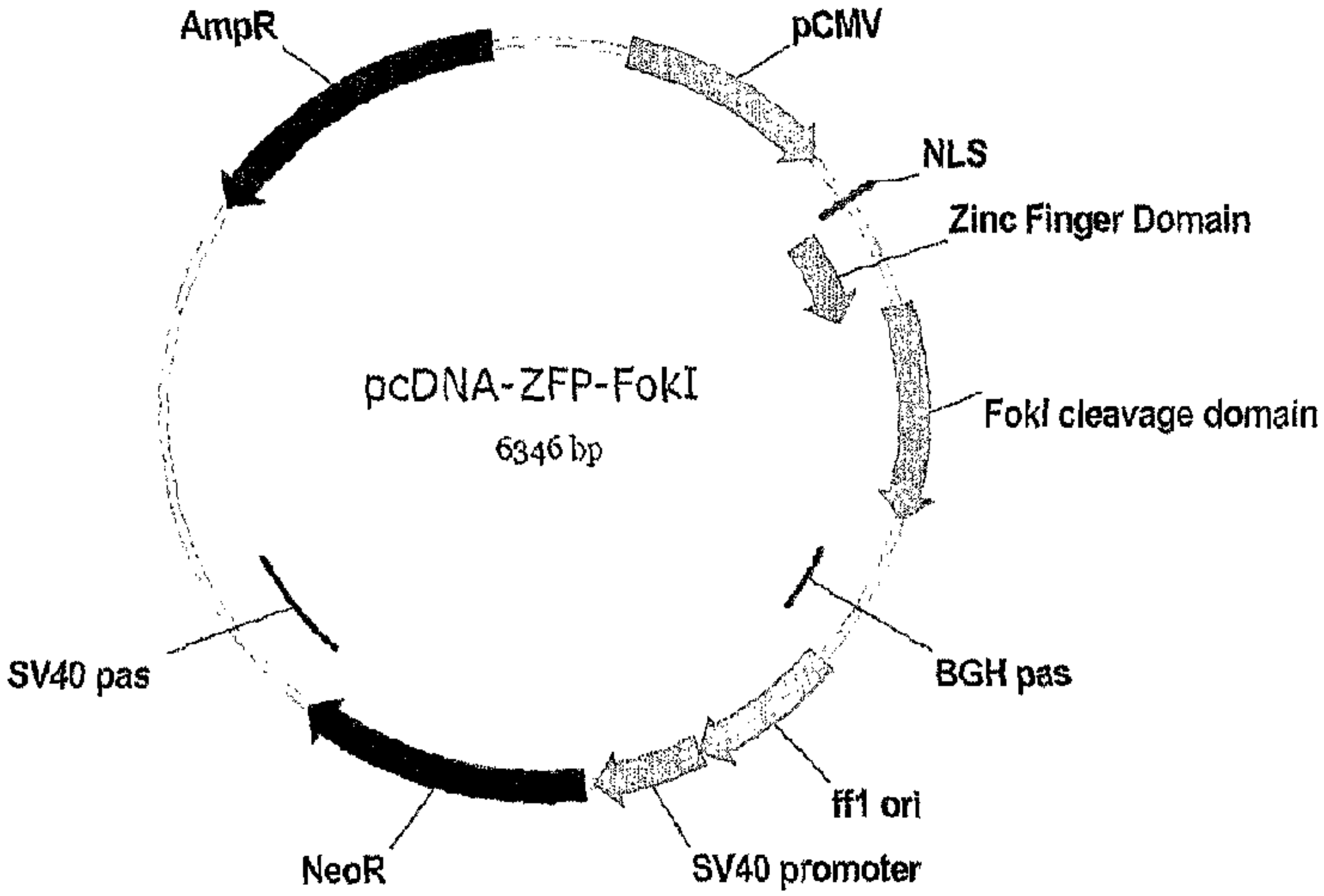
Wild-type *Fok*I cleavage half-domain
QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRG
KHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNK
HINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIG
GEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:249)

E490K
QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRG
KHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNK
HINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIG
GEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:250)

X2 (E490K:I538K)
QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRG
KHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVKENQTRNK
HINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHKTNCNGAVLSVEELLIG
GEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:251)

X3A (Q486E:E490K:I538K)
QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRG
KHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMERYVKENQTRNK
HINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHKTNCNGAVLSVEELLIG
GEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:252)

X3B (Q486I:E490K:I538K)
QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRG
KHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMIRYVKENQTRNK
HINPNEWWKVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHKTNCNGAVLSVEELLIG
GEMIKAGTLTLEEVRRKFNNGEINF (SEQ ID NO:253)


AmpR
pCMV
NLS
Zinc Finger Domain
pcDNA-ZFP-FokI
6346 bp
FokI cleavage domain
SV40 pas
BGH pas
ff1 ori
NeoR
SV40 promoter