

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
9 December 2004 (09.12.2004)

PCT

(10) International Publication Number
WO 2004/106513 A1

(51) International Patent Classification⁷: **C12N 15/00**,
1/21, 5/10, 1/20, 1/19

(21) International Application Number:
PCT/US2003/014657

(22) International Filing Date: 9 May 2003 (09.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicants (for all designated States except US): **THE GOVERNMENT OF THE UNITED STATES OF AMERICA** as represented by **THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; National Institutes Of Health, Office Of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852-3804 (US). **THE UNIVERSITY OF HONG KONG** [CN/CN]; Department of Biochemistry, 3/F Laboratory Block, Faculty of Medicine Building, 21 Sassoon Road, Hong Kong SAR (CN). **CHINESE ACADEMY OF MEDICAL SCIENCES (CAMS) & PEKING UNION MEDICAL COLLEGE (PUMC)** [CN/CN]; National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Beijing 100005 (CN).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **COURT, Donald, L.** [US/US]; 502 Magnolia Avenue, Frederick, MD 21702 (US). **LI, Xin-Tian** [CN/CN]; National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing 100005 (CN). **HUANG, Jian-Dong** [US/CN]; c/o Dept. of Biochemistry, the University of Hong Kong, 3/F Laboratory Block Faculty of Medicine Building, 21 Sassoon

Road, Hong Kong (CN). **COSTANTINO, Nina** [US/US]; 7913 Runnymede Drive, Frederick, MD 21702 (US). **LIU, Depei** [CN/CN]; National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing 100005 (CN).

(74) Agent: **SIEGEL, Susan, Alpert**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

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

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

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— with international search report

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.

(54) Title: HOST CELLS DEFICIENT FOR MISMATCH REPAIR

(57) Abstract: Methods are disclosed herein for inducing homologous recombination in a host cell comprising a target nucleic acid, using a single-stranded nucleic acid molecule. The single-stranded nucleic acid molecule has a sufficient number of nucleotides homologous to the target nucleic acid to enable homologous recombination with the target nucleic acid. The host cell includes a de-repressible promoter operably linked to a nucleic acid encoding a single-stranded binding protein and is deficient for mismatch repair. Isolated host cells of use in this method are also disclosed.



WO 2004/106513 A1

HOST CELLS DEFICIENT FOR MISMATCH REPAIR

5

FIELD

The present disclosure relates to methods to induce homologous recombination in cells in cells deficient for mismatch repair. It also relates to methods for modifying genomic DNA in bacterial artificial chromosomes (BACs) and to subcloning of genomic DNA from BACs into multi-copy plasmids. Cells of use for recombineering are disclosed that that are deficient for mismatch repair, and thus exhibit an increased frequency of homologous recombination.

BACKGROUND

15 Concerted use of restriction endonucleases and DNA ligases allows *in vitro* recombination of DNA sequences. The recombinant DNA generated by restriction and ligation may be amplified in an appropriate microorganism such as *E. coli*, and used for diverse purposes including gene therapy. However, the restriction-ligation approach has two practical limitations: first, DNA molecules can be precisely
20 combined only if convenient restriction sites are available; second, because useful restriction sites often repeat in a long stretch of DNA, the size of DNA fragments that can be manipulated is limited, usually to less than about 25 kilobases.

Homologous recombination, generally defined as an exchange between homologous segments anywhere along a length of two DNA molecules, provides an
25 alternative method for engineering DNA. In generating recombinant DNA with homologous recombination, a microorganism such as *E. coli*, or a eukaryotic cell such as a yeast or vertebrate cell, is transformed with exogenous DNA. The center of the exogenous DNA contains the desired transgene, whereas each flank contains a segment of homology with the cell's DNA. The exogenous DNA is introduced into
30 the cell with standard techniques such as electroporation or calcium phosphate-mediated transfection, and recombines into the cell's DNA, for example with the assistance of recombination-promoting proteins in the cell.

-2-

A recombination system (termed "recombineering") has been developed for efficient chromosome engineering in *Escherichia coli* using electroporated linear DNA (see published PCT Application No. WO 00/214495 A2, which is herein incorporated by reference). A defective prophage supplies functions (λ Red) that
5 protect and recombine an electroporated DNA substrate in the bacterial cell. This system can be used with single-stranded DNA, as well as with linear double-stranded DNA (dsDNA). The use of recombination eliminates the requirement for standard cloning as all novel recombination sites are engineered by chemical synthesis *in vitro*, and the linear DNA is efficiently recombined *in vivo*. In this
10 system, a temperature-dependent repressor tightly controls prophage expression, such that recombination functions can be transiently supplied by shifting cultures to 42°C. The efficient prophage recombination system does not require host RecA function and depends primarily on *exo*, *bet*, and *gam* functions expressed from the defective prophage. The defective prophage can be moved to other strains and can be
15 easily removed from any strain. Importantly, recombination in this system is proficient with DNA homologies as short as 30-50 base pairs, making it possible to use PCR-amplified fragments as the targeting cassette. Gene disruptions and modifications of both the bacterial chromosome and bacterial plasmids are possible, and the system has been shown to be of use in the bacterial artificial chromosome
20 libraries (see Published PCT Application No. WO 00/214495, herein incorporated by reference; Yu et al., *Proc. Natl. Acad. Sci. USA* 97:5978-5983, 2000).

This prophage system has been adapted for use in bacterial artificial chromosome (BAC) engineering by transferring it to DH10B cells, a BAC host strain. Fragments as large as 80 kb can be subcloned from BACs by gap repair
25 using this recombination system, obviating the need for restriction enzymes or DNA ligases. BACs can be modified with this recombination system in the absence of drug selection (Lee et al., *Genomics* 73:56-65, 2001). It has been suggested that recombineering in BACs allows modification or subcloning of large fragments of genomic DNA with precision. This ability facilitates many kinds of genomic
30 experiments that were difficult or impossible to perform previously and aid in studies of gene function. It has been suggested that this system is of use in

-3-

generating mouse models and providing a refined analysis of the mouse genome (Copland et al., *Nat. Rev. Genet.* 2:769-779, 2001).

Recombineering uses the *exo* and *bet* functions of the prophage lambda under the control of a temperature sensitive repressor. When the lambda functions
5 are turned on, cells become more “recombinogenic,” that is they take up DNA and recombination of the DNA occurs with a target sequence in the cell. This system has been adapted for use in bacterial artificial chromosome engineering, wherein inducible recombinases (e.g. cre or flpe) are introduced into host cells and BAC modification is accomplished using recombination sites (e.g. loxP or frt,
10 respectively). This system can be used to generate Cre-expressing transgenic mice for use in conditional knock-out studies. This system utilizes a targeting vector to introduce recombination sites (e.g. loxP) into a gene of interest. Expression of a recombinase from a specific promoter (such that expression occurs in a tissue of interest) results in recombination at the recombination sites, leading to a
15 “conditional knock-out.”

It has also been shown that synthetic single-stranded oligonucleotides (SSOs) can also be used in place of linear dsDNA fragments, to create sequence-specific mutations (Ellis et al., *Proc. Natl Acad Sci USA* 98:6742-6746, 2001; Swaminathan et al., *Genesis* 29:14-21, 2001). However, the mechanism(s) by
20 which SSOs effect these genetic changes remains to be established. In addition, although recombineering using SSOs provides a method for introducing homologous DNA into a target nucleic acid sequence, the frequency of recombination can be improved. Thus, methods for increasing recombination frequency using recombineering, and strains which produce recombinants at a high
25 frequency are disclosed herein.

SUMMARY

This disclosure provides methods for inducing homologous recombination using single-stranded or double-stranded oligonucleotides in host cells. Host cells
30 are of use in increasing the efficiency of recombination using the *bet* (and optionally *exo* and *gam*) functions of the prophage lambda under the control of a de-repressible promoter, and are deficient for mismatch repair. In one specific, non-limiting

-4-

example the promoter is lambda PL promoter, which is under the control of a temperature sensitive repressor. In one specific, non-limiting example, the nucleic acid nucleic acid is single-stranded.

The host cells of use in these methods are deficient for mismatch repair.

5 Thus, in one embodiment, the host cell is a bacterial cell that is deficient for mismatch repair. Specific, non-limiting examples are bacterial cells that have reduced or absent function of the polypeptides encoded by mutS, mutH, mutL, uvrD, and/or or dam, such that the bacterial cell is deficient for mismatch repair.

The methods include introducing a single-stranded or double-stranded
10 homologous nucleic acid of sufficient homology to a target sequence into a host cell deficient for mismatch repair. The homologous nucleic acid is of sufficient length to undergo homologous recombination with the target sequence, and can optionally include a mutation as compared to the target sequence. The host cell includes a de-repressible promoter operably linked to a nucleic acid encoding Beta. Activation of
15 the de-repressible promoter, such that the expression of Beta is induced, results in recombination of the homologous nucleic acid with the target nucleic acid. In additional embodiments, the de-repressible promoter is also operably linked to a nucleic acid encoding Exo and or Gam. In further embodiments, the host cell is a bacterial cell, a yeast cell, or a mammalian cell.

20 The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE FIGURES

25 **Fig. 1** is a schematic diagram of single stranded DNA at the replication fork. Both the leading and the lagging strand are shown. A single stranded nucleic acid is utilized for recombination that includes a mismatch. The annealing of the single-stranded nucleic acid, mediated by Beta, is indicated on the replication fork. MutS, MutH, MutL, and dam are shown at the replication fork. If present, these proteins
30 will correct the mismatch. A strain that is deficient for MutS, MutH, MutL, dam (or

-5-

uvrD) will not repair the mismatch, and thus an increased recombination frequency will be detected.

Fig. 2 is a schematic diagram of the use of a double stranded DNA including a 3' overhang. Two single stranded oligonucleotides are generated that are homologous to each other over a portion of their sequence, and thus can form an intermediate that is a double-stranded nucleic acid with 3' overhangs. The overhangs are single stranded. Beta (bet) binds the single-stranded ends of the intermediate, and can mediate recombination of this molecule into a target sequence. The annealing of intermediate to the replication fork at the lagging strand is shown.

Fig. 3 is a schematic diagram of branch migration at a replication fork that occurs following the step shown in Fig. 2. A region of mismatch is indicated. The mismatch repair proteins will correct the mismatch. Thus, recombination frequency will be increased in a host cell that is deficient for mismatch repair.

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. For double-stranded nucleic acids, only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. However, for single-stranded nucleic acids, the complementary strand is not included.

25

DETAILED DESCRIPTION

I. Abbreviations

Amp: ampicillin

BAC: bacterial artificial chromosome

30 **ddH₂O:** double distilled water

DNA: deoxyribonucleic acid

ds: double-stranded

-6-

- IPTG:** isopropyl-beta-D-thiogalactopyranoside
Kan: kanamycin
Lag: lagging
Lead: leading
5 **ml:** milliliter
min: minute(s)
MMR: mismatch repair
ng: nanograms
nt: nucleotide
10 **NT oligo:** non-template oligonucleotide
Oligo: oligonucleotide
PCR: polymerase chain reaction
ss: single-stranded
SSO or ss oligo: single-stranded oligonucleotide
15 **T oligo:** template oligo
μl: microliter
YAC: yeast artificial chromosome

II. *Terms*

- 20 Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al. (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers
25 (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments of this disclosure, the following explanations of specific terms are provided:

- 30 **3' overhang:** two nucleic acid sequences which when annealed are partially double-stranded and partially single-stranded. The single-stranded end or ends extend away from the double-stranded segment in a 5' to 3' direction.

-7-

5' overhang: two nucleic acid sequences which when annealed are partially double-stranded and partially single-stranded. The single-stranded end or ends extend away from the double-stranded segment in a 3' to 5' direction.

Bacterial artificial chromosome (BAC): Bacterial artificial chromosomes (BACs) have been constructed to allow the cloning of large DNA fragments in *E. coli*, as described in O'Conner et al., *Science* 244:1307-1312, 1989; Shizuya et al., *Proc. Natl. Acad. Sci. USA* 89:8794-8797, 1992; Hosoda et al., *Nucleic Acids Res.* 18:3863-3869, 1990; and *Current Protocols in Molecular Biology*, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates). This system is capable of stably propagating mammalian DNA over 300 kb. In one embodiment, a BAC carries the F replication and partitioning systems that ensure low copy number and faithful segregation of plasmid DNA to daughter cells. Large genomic fragments can be cloned into F-type plasmids, making them of use in constructing genomic libraries.

Beta: The 28 kDa lambda Beta ssDNA binding polypeptide (and nucleic acid encoding lambda Beta) involved in double-strand break repair homologous recombination. DNA encoding Beta (*bet*) and polypeptide chains having lambda Beta activity are also referred to herein as *bet* (see published PCT Application No. WO 00/214495 A2, herein incorporated by reference). The lambda Beta protein binds to single-stranded DNA and promotes renaturation of complementary single-strand regions of DNA (see Karakousis et al, *J. Mol. Biol.* 276:721-733, 1998; Li et al., *J. Mol. Biol.* 276:721-733, 1998; Passy et al., *PNAS* 96:4279-4284, 1999).

Functional fragments and variants of Beta include those variants that maintain their ability to bind to ssDNA and mediate the recombination function of lambda Beta as described herein, and in the publications referenced herein. It is recognized that the gene encoding Beta may be considerably mutated without materially altering the ssDNA binding function or homologous recombination function of lambda Beta. First, the genetic code is well-known to be degenerate, and thus different codons encode the same amino acids. Second, even where an amino acid mutation is introduced, the mutation may be conservative and have no material impact on the essential functions of lambda Beta. See Stryer, *Biochemistry* 3rd Ed.,

(c) 1988. Third, part of the lambda Beta polypeptide chain may be deleted without impairing or eliminating its ssDNA binding protein function, or its recombination function. Fourth, insertions or additions may be made in the lambda Beta polypeptide chain -- for example, adding epitope tags -- without impairing or
5 eliminating its essential functions (*Current Protocols in Molecular Biology*, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates)).

Biolistics: Insertion of DNA into cells using DNA-coated micro-projectiles. Also known as particle bombardment or microparticle bombardment. The approach
10 is further described and defined in U.S. Patent No. 4,945,050.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences that determine transcription. cDNA may be synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

15 **Cosmid:** Artificially constructed cloning vector containing the *cos* gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into *E. coli*; this permits cloning of larger DNA fragments (up to 45 kb) than can be introduced into bacterial hosts in plasmid vectors.

De-repressible Promoter: When a repressor is bound to a de-repressible
20 promoter transcription is substantially decreased as compared to transcription from the de-repressible promoter in the absence of repressor. By regulating the binding of the repressor, such as by changing the environment, the repressor is released from the de-repressible promoter, and transcription increases. In this definition, a de-repressible promoter does not require an activator for transcription. One specific,
25 non-limiting example is the PL promoter, which is regulated by the repressor cI, but is not activated by an activator. (The arabinose promoter is not a simple de-repressible promoter as arabinose inactivates the repressor AraC and converts it to an activator).

In one embodiment, the de-repressible promoter is a temperature sensitive
30 de-repressible promoter. For example, by increasing the temperature, the repressor is released from the promoter, or can no longer bind to the promoter with a high

affinity, and transcription is increased from the promoter. One specific, non-limiting example is the induction of PL promoter activity by increasing the temperature of the cell. Increased temperature inactivates a temperature-sensitive repressor cI, allowing genes that are operably linked to the PL promoter to be expressed at
5 increased levels. One of skill in the art can readily identify a de-repressible promoter.

In one embodiment, a de-repressible promoter is auto-regulated. For example, if only one copy of a gene encoding cI is present, yet many copies of the PL promoter are present, expression of cI is upregulated such that transcription is
10 blocked from any of the PL promoters.

Double-strand break repair recombination: A type of homologous recombination exemplified by the lambda recombination proteins Exo, Beta and Gam, and shared by numerous other recombinase systems. A double-strand break is the initiation point for concerted action of recombination proteins. Typically, an
15 exonuclease degrades processively from the 5' ends of these break sites, and ssDNA binding polypeptide binds to the remaining 3' single-strand tail, protecting and preparing the recessed DNA for homologous strand invasion (Szostak et al., *Cell* 33:25-35, 1983; Little, *J. Biol. Chem.* 242:679-686, 1967; Carter et al., *J. Biol. Chem.* 246:2502-2512, 1971; Lindahl et al., *Science* 286:1897-1905, 1999).
20 Examples of ssDNA binding polypeptides which bind to either ssDNA and/or dsDNA with 3' overhangs and promote double-strand break repair recombination include lambda Beta, RecT of *E. coli*, Erf of phage p22, and Rad52 in various eukaryotic cells including yeast and mammalian cells.

Electrocompetent: cells capable of macromolecular uptake upon treatment
25 with electroporation.

Electroporation: a method of inducing or allowing a cell to take up macromolecules by applying electric fields to reversibly permeabilize the cell walls. Various methods and apparatuses used are further defined and described in: U.S. Patent No. 4,695,547; U.S. Patent No. 4,764,473; U.S. Patent No. 4,882,28; U.S.
30 Patent No. 4,946,793; U.S. Patent No. 4,906,576; U.S. Patent No. 4,923,814; and U.S. Patent No. 4,849,089.

Eukaryotic cell: a cell having an organized nucleus bounded by a nuclear membrane. These include lower organisms such as yeasts, slime molds, and the like, as well as cells from multicellular organisms such as invertebrates, vertebrates, and mammals. They include a variety of tissue types, such as: endothelial cell, smooth muscle cell, epithelial cell, hepatocyte, cells of neural crest origin, tumor cell, hematopoietic cell, immunologic cell, T cell, B cell, monocyte, macrophage, dendritic cell, fibroblast, keratinocyte, neuronal cell, glial cell, adipocyte, myoblast, myocyte, chondroblast, chondrocyte, osteoblast, osteocyte, osteoclast, secretory cell, endocrine cell, oocyte, and spermatocyte. These cell types are described in standard histology texts, such as McCormack, *Introduction to Histology*, (c) 1984 by J.P. Lippincott Co.; Wheater et al., eds., *Functional Histology*, 2nd Ed., (c) 1987 by Churchill Livingstone; Fawcett et al., eds., Bloom and Fawcett: *A Textbook of Histology*, (c) 1984 by William and Wilkins.

Exo: the exonuclease of lambda (and the nucleic acid encoding the exonuclease protein) involved in double-strand break repair homologous recombination. See published PCT Application No. WO 00/214495 A2 and references therein for discussion.

Gam: a lambda protein (and nucleic acid encoding Gam) involved in double-strand break repair homologous recombination. It is believed to inhibit cellular nuclease activity such as that encoded by the recBCD and sbcC system of *E. coli*. See published PCT Application No. WO 00/214495 A2 for discussion. Over-expression of Gam function, when expressed in the cell, is extremely toxic to the cell, and prevents growth. For this reason tight controls over its expression are always required. PL and cI 857 are able to regulate Gam expression.

Functional fragments and variants of Exo and Gam: As discussed for Beta (see "Functional Fragments And Variants Of Beta"), it is recognized that genes encoding Exo or Gam may be considerably mutated without materially altering their function, because of genetic code degeneracy, conservative amino acid substitutions, noncritical deletions or insertions, etc. Unless the context makes it otherwise clear, the term lambda Exo, *Exo*, or lambda exonuclease are all intended to include the native lambda exonuclease, and all fragments and variants of lambda exonuclease.

-11-

Extrachromosomal: not incorporated into the chromosome or chromosomes of a cell. In the context of nucleic acids, extrachromosomal indicates a DNA oligonucleotide that is not covalently incorporated into the chromosome or chromosomes of a cell. **Intrachromosomal** refers to material such as an
5 oligonucleotide that is incorporated into the chromosome or chromosomes of a cell, such as a DNA oligonucleotide covalently incorporated into the chromosomal DNA of a cell.

Homologous arm: nucleotides at or near 5' or 3' end of a polynucleotide which are identical or similar in sequence to the target nucleic acid in a cell, and
10 capable of mediating homologous recombination with the target nucleic acid. Homologous arms are also referred to as homology arms. In one embodiment, a homology arm includes at least 20 bases of a sequence homologous to a nucleic acid of interest. In another embodiment, the homology arm includes at least 30 base pairs of a sequence homologous to a nucleic acid of interest. In yet another
15 embodiment, a homology arm includes at least 40 base pairs of a sequence homologous to a nucleic acid of interest. In a further embodiment, a homology arm includes from about 50 to about 100 base pairs of a sequence homologous to a nucleic acid of interest.

Homologous recombination: an exchange of homologous polynucleotide
20 segments anywhere along a length of two nucleic acid molecules. Homologous nucleic acids include nucleic acids that are 100% identical, as well as nucleic acids with limited differences in their nucleic acid sequence. For example, homologous recombination can occur between two nucleic acid sequences that are at least 90% identical, such as nucleic acid sequences that are at least 91%, at least 92%, at least
25 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, identical.

Host cell: a cell that is used in laboratory techniques such as DNA cloning to receive exogenous nucleic acid molecules. In one embodiment, a host cell is used to maintain or allow the reproduction of a vector, or to facilitate the manipulation of
30 nucleic acid molecules *in vitro*. A host cell can be a prokaryotic or a eukaryotic cell.

In one embodiment, a host cell is a bacterial cell, such as, but not limited to, an *E. coli* cell (e.g. DH10B, DY330, HME6, HME41, or DY380).

HVJ-mediated gene transfer: a method of macromolecular transfer into cells using inactivated hemagglutinating virus of Japan and liposomes, as described
5 in Morishita et al., *J. Clin. Invest.* 91:2580-2585, 1993; Morishita et al., *J. Clin. Invest.* 94:978-984, 1994.

Inducible promoter: a promoter whose activity may be increased by some change in the environment of the cell. Examples of inducible promoters abound in nature, and a broad range of environmental or hormonal changes may activate or
10 repress them. One specific, non-limiting example of an inducible promoter is the arabinose promoter.

Isolated: An "isolated" biological component (such as a nucleic acid or protein) has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs,
15 i.e., other chromosomal and extra-chromosomal DNA and RNA, and proteins. Thus, nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

20 **Lagging and Leading Strand of a double-stranded (ds) DNA:** the leading strand is the strand of a double-stranded DNA that is synthesized continuously during replication. The lagging strand is the strand of a double-stranded DNA that is synthesized discontinuously.

Okazaki fragments are short segments of newly synthesized DNA produced
25 during replication. All the known DNA polymerases only synthesize DNA in one direction, the 5' to 3' direction. However, as the strands separate, the replication fork moves along one parental strand in the 3' to 5' direction and 5' to 3' on the other parental strand. On the former, the leading strand, DNA is synthesized continuously in the 5' to 3' direction. On the other, the lagging strand, DNA synthesis only occurs
30 when a length of single-stranded (ss) DNA has been exposed and proceeds in the direction opposite to the movement of the replication fork (5' to 3'). Thus, it is

discontinuous, and the series of fragments are polymerized and then covalently linked by ligases to give a continuous strand. In eukaryotes, Okazaki fragments are typically a few hundred nucleotides in length, whereas in prokaryotes they can be several thousand nucleotides in length.

5 **Linear plasmid vector:** a DNA sequence (1) containing a bacterial plasmid origin of replication, (2) having a free 5' and 3' end, and (3) capable of circularizing and replicating as a bacterial plasmid by joining its free 5' and 3' ends. Examples of linear plasmid vectors include the linearized pBluescript vector and linearized pBR322 vectors.

10 **Lipofection:** the process of macromolecular transfer into cells using liposomes. See U.S. Patent No. 5,651,981.

Mini lambda: a derivative of lambda (λ) wherein most of the viral lytic genes, including those required for replication and lysis, are deleted. A mini-lambda maintains the Red functions (Beta, with, or without, Exo and/or Gam) for
 15 homologous recombination and maintains the integration/excision functions (e.g. att, integrase (int). and excisionase (xis)) to insert and excise its DNA from the chromosome.

Mismatch: a base-pair in a double-stranded nucleic acid that does not mate via normal Watson-Crick binding (e.g., G/C, A/T). For example, in the structure
 20 diagrammed below:



25 There is a base pair mismatch at the 3' end of the lower strand. Base pair mismatches include any homologous pairing of nucleotides (A/A, T/T, G/G, and C/C), as well as some heterologous pairings (G/A, G/T, C/A, C/G) of nucleotides. Mismatch is also a small insertion/deletion loop in double-stranded DNA. Thus, in a structure such as

-14-

5'----CCCCCGGGGG---3'

3'----GGGGCCCCC---5'

A

5 the A represents a base inserted into the bottom strand that is a “mismatch” as there is no corresponding base with which it forms Watson-Crick base pairs within the upper strand.

Although the examples shown are based on a single nucleotide difference, the a mismatch can be for a longer length of nucleic acid, such as but not limited to, 10 a mismatch of about one, two, three, four, five, or ten nucleotides.

Mismatch Repair and Polypeptides Involved in Mismatch Repair: the endogenous pathway in cells response for the repair of a mismatch. In prokaryotes (e.g.; *E. coli*), the MutS, MutL, MutH, dam, and uvrD proteins are involved in mismatch repair, amongst others.

15 Briefly, in a bacterial cell, the MutS protein recognizes and binds to mismatches in DNA duplexes. Briefly, a mismatch existing in the form of a base-base mispair of a small insertion/deletion loop in double-stranded DNA is bound by a MutS protein homodimer (e.g., see Genbank Accession No. NP_417213). Multiple MutS homodimers can bind to form a multimer complex until a hemi- 20 methylated GATC site is encountered. The MutL (e.g. see GenBank Accession No. NP_418591) protein binds to MutS, and the MutH protein (e.g. see GenBank Accession No. NP_417308) binds to hemimethylated dam methylation sites (which are methylated by dam, see, for example, GenBank Accession No. NP_417846). The MutS-MutL-DNA complex stimulates MutH to cleave the unmethylated DNA 25 strand at the GATC sequence (the GATC site can be either 5' or 3' of the recognized mismatch). One of two exonucleases (depending on whether cleavage was 5' or 3') chews away at the DNA to beyond the mismatch site, such that long patch repair synthesis can occur. Excision is accomplished by cooperation between the uvrD (helicase II, e.g. see GenBank Accession No. NP_418258) protein, which unwinds 30 from the nick in the direction of the mismatch, and a single-strand specific exonuclease of appropriate polarity, followed by resynthesis (polymerase III) and

ligation (DNA ligase). For a review of mismatch repair pathways, see Marti et al., *J. Cell. Physiol.* 191:28-41, 2002.

Eukaryotes have proteins with sequence similarity to MutS and MutL that are involved in a similar repair pathway. The eucaryotic MutS is a dimer of MSH2 and GTBP (now known as MSH3 or MSH6) proteins. Eucaryotic MutL also
5 consists of two polypeptides, MLH1 and PMS2. One of skill in the art can readily identify the eukaryotic proteins involved in mismatch repair. In addition, a table (Table 2) describing exemplary proteins is provided in the disclosure herein.

Nucleic acid: a deoxyribonucleotide or ribonucleotide polymer in either
10 single or double-stranded form, including known analogs of natural nucleotides unless otherwise indicated.

Oligonucleotide: a single-stranded nucleic acid ranging in length from 2 to about 500 bases, for example, polynucleotides that contain at least 20 or 40 nucleotides. Oligonucleotides are often synthetic but can also be produced from
15 naturally occurring polynucleotides.

Operably linked: a first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the
20 transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Phagemid artificial chromosome: also referred to as P1 artificial chromosome. A type of artificial chromosome allowing for stable cloning of very
25 large DNA fragments. Further described in Shepherd et al., *Proc. Natl. Acad. Sci. USA* 92:2629, 1994; Iannou et al., *Nature Genetics* 6:84-89, 1994.

Phage-based recombination systems: bacteria such as *E. coli* encode their own homologous recombination systems, which are used in repair of DNA damage and to maintain a functional chromosome. The viruses or phages that inhabit
30 bacteria often carry their own recombination functions. Phage λ carries the Red

-16-

recombination system. These phage systems can work with the bacterial recombination functions or independently of them.

PL promoter: the major leftward promoter of bacteriophage lambda. Once the lambda DNA is incorporated into the bacterial chromosome, transcription from this promoter is tightly repressed by the cI repressor. Upon inactivation of the cI repressor, for example by heat shock of a temperature sensitive mutant, transcription from the PL promoter is activated, leading to expression of lambda genes. See Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and Oppenheim et al., *Prog. Nucleic Acid Res. Mol. Biol.* 46:37-49, 1993.

Plasmid: autonomously replicating, extrachromosomal DNA molecules, distinct from the normal bacterial genome and nonessential for bacterial cell survival under nonselective conditions.

Polynucleotide: a double-stranded or single-stranded nucleic acid sequence of any length. Therefore, a polynucleotide includes molecules which are 15, 50, 100, 200 nucleotides long (oligonucleotides) and also nucleotides as long as a full length cDNA.

Polypeptide: any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation).

Prokaryote: cell or organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments. Prokaryotes include Archaea, and Bacteria. Bacteria include hydrogenobacteria, Thermotogales, green non-sulfur bacteria, the deioncoccus group, Cyanobacteria, planctomyces, spirochetes, spirilla, myxobacteria, lithotrophs, pseudomonads, enteric bacteria, vibrios, green sulfur bacteria, cytophagas, pyogenic cocci, and Gram Positive bacteria. Thus, prokaryotes include both gram positive and gram negative bacteria (e.g. *Staphylococcus aureus*, *Streptococcus pyogenes*, and *Streptococcus pneumonia*). In one specific, non-limiting example a prokaryote is a bacteria, such as a gram-negative bacteria, such as an enteric bacteria (e.g., *E. coli*).

Probes and primers: a nucleic acid probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include

-17-

radioactive isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and *Current Protocols in Molecular Biology*, ed. Ausubel
5 et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates)).

Primers are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and
10 the target DNA strand. The 3' hydroxyl end of the primer may be then extended along the target DNA strand through the use of a DNA polymerase enzyme. Primer pairs (one on either side of the target nucleic acid sequence) can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

15 Methods for preparing and using probes and primers are described, for example, in Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, Ausubel et al., *supra* (1987). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that
20 purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). Under appropriate conditions, the specificity of a particular probe or primer increases with its length. Thus, in order to obtain greater specificity, probes and primers may be selected that comprise 20, 25, 30, 35, 40, 50 or more consecutive nucleotides of related cDNA or gene sequence.

Purified: the term purified does not require absolute purity; rather, it is
25 intended as a relative term. Thus, for example, a purified lambda Beta preparation or ssDNA binding polypeptide is one in which the protein is more enriched than the protein is in its natural environment within a cell. Preferably, a preparation of lambda Beta is purified such that the polypeptide represents at least 50% of the total protein content of the preparation.

30 **RecA:** the RecA protein is a central protein that has an activity as in the recombination function of *E. coli*. Homologues are found in all other organisms.

RecA protein allows two homologous DNAs to find each other among non-homologous DNAs and then trade or transfer strands with each other. This exchange occurs by RecA binding to a single-stranded region in one of the DNAs and using that strand to search for its dsDNA homolog, binding to the dsDNA and causing the single-strand to pair with its complement in the dsDNA ultimately displacing the identical strand of the duplex. This strand transfer generates a key intermediate in the RecA-mediated recombination process.

***recE recT* genes and the Rac prophage:** *E. coli* and other bacteria contain in their chromosomes remnants of viruses. These viruses or prophages are for the most part defective and may contain only a few genes of the original virus. In *E. coli*, one defective prophage is called Rac. Two genes, *recE* and *recT* of the Rac prophage, encode homologous recombination functions. These genes are normally silent but the *sbcA* mutation activates their constitutive expression. Thus, the *sbcA* mutant is active for recombination.

Recombinases: proteins that, when included with an exogenous targeting polynucleotide, provide a measurable increase in the recombination frequency between two or more oligonucleotides that are at least partially homologous.

Recombineering: the use of a recombinase to mediate homologous recombination between linear DNA introduced into a living cell and a replicon in the cell. The linear DNA can be either single or double stranded. A recombineering system is described in published PCT Application No. WO 00/214495 A2, which is herein incorporated by reference. In one embodiment, recombineering uses the *bet* functions (and optionally *exo* and *gam*) of the prophage lambda under the control of a de-repressible promoter, such as but not limited to, a promoter regulated by a temperature sensitive repressor. When the lambda functions are turned on, cells become more "recombinogenic," that is they take up DNA and recombination of the DNA occurs with a target sequence in the cell. This system has been adapted for use in bacterial artificial chromosome engineering, and can also be used with yeast artificial chromosomes. Recombineering can be modified to use other functions, such as, but not limited to 34.1 gene from *Bacillus subtilis* phage SPP1, which encodes the polypeptide Chu (the product of 34.1) or herpes simplex virus (HSV) UL29 (Vellani and Meyers, *J. Bacteriol.* 185: 2465-74, 2003). Recombineering can

be used to introduce mutations into a target sequence in both eukaryotic and prokaryotic cells.

Selection markers or selectable markers: nucleic acid sequences which upon intracellular expression are capable of being detected, such a nucleic acid
5 sequence that confers either a positive or negative selection marker or phenotypic characteristic for the cell expressing the sequence. The term "selection marker" or "selectable marker" includes both positive and negative selection markers. A "positive selection marker" is a nucleic acid sequence that allows the survival of cells containing the positive selection marker under growth conditions that kill or
10 prevent growth of cells lacking the marker. Examples of a positive selection marker is a nucleic acid sequence which promotes expression of the neomycin resistance gene, ampicillin resistance gene, and kanamycin resistance gene. Cells not containing the neomycin resistance gene are selected against by application of G418, whereas cells expressing the neomycin resistance gene are not harmed by G418
15 (positive selection). A "negative selection marker" is a nucleic acid sequence that kills, prevents growth of or otherwise selects against cells containing the negative selection marker, usually upon application of an appropriate exogenous agent. An example of a negative selection marker is the sacB gene, which encodes a function causing sensitivity to the addition of sucrose to the culture. Another example of a
20 negative marker is a nucleic acid sequence which promotes expression of the thymidine kinase gene of herpes simplex virus (HSV-TK). Cells expressing HSV-TK are selected against by application of ganciclovir (negative selection), whereas cells not expressing the gene are relatively unharmed by ganciclovir. The terms are further defined, and methods further explained, by U.S. Patent No. 5,464,764.

25 **Sequence identity:** the similarity between two nucleic acid sequences, or two amino acid sequences is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar are the two sequences.

30 Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman,

-20-

Adv. Appl. Math. 2:482, 1981; Needleman and Wunsch, *J. Mol. Bio.* 48:443, 1970; Pearson and Lipman, *Methods in Molec. Biology* 24:307-331, 1988; Higgins and Sharp, *Gene* 73:237-244, 1988; Higgins and Sharp, *CABIOS* 5:151-153, 1989; Corpet et al., *Nucleic Acids Research* 16:10881-10890, 1988; Huang et al.,
5 *Computer Applications in BioSciences* 8:155-165, 1992; and Pearson et al., *Methods in Molecular Biology* 24:307-331, 1994. Altschul et al. (1994) presents a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., *J. Mol. Biol.* 215:403-410, 1990) is available from several sources, including the
10 National Center for Biological Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at the NCBI website, together with a description of how to determine sequence identity using this program.

Homologues of lambda Beta, Exo and Gam, and ssDNA binding proteins
15 typically possess at least 60% sequence identity counted over full-length alignment with the amino acid sequence of the protein being evaluated (that is, lambda Beta, Exo or Gam, or ssDNA binding protein such as P22 Erf, RecT, and Rad52) using the NCBI Blast 2.0, gapped blastp set to default parameters. For comparisons of amino acid sequences of greater than about 30 amino acids, the Blast 2 sequences function is
20 employed using the default BLOSUM62 matrix set to default parameters, (gap existence cost of 11, and a per residue gap cost of 1). When aligning short peptides (fewer than around 30 amino acids), the alignment should be performed using the Blast 2 sequences function, employing the PAM30 matrix set to default parameters (open gap 9, extension gap 1 penalties). Proteins with even greater similarity to the
25 reference sequence will show increasing percentage identities when assessed by this method, such as at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, at least 98%, or at least 99% sequence identity. When less than the entire sequence is being compared for sequence identity, homologs will typically possess at least 75% sequence identity over short windows of 10-20 amino acids, and may possess
30 sequence identities of at least 85% or at least 90% or 95% depending on their similarity to the reference sequence. Methods for determining sequence identity over such short windows are described at the NCBI website.

Similarly the identity between a target nucleic acid sequence and a homologous single-stranded nucleic acid sequence using the NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., *J. Mol. Biol.* 215:403-410, 1990), as available from several sources, including the National Center for Biological Information (NCBI, Bethesda, MD) and on the Internet. It can be accessed at the
5 NCBI website, together with a description of how to determine sequence identity of nucleotide sequences.

One of skill in the art will appreciate that these sequence identity ranges are provided for guidance only; sequences of sufficient identity to provide homologous
10 recombination activity could be obtained that fall outside of the ranges provided.

Single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA):
ssDNA is DNA in a single polynucleotide chain; the DNA bases are not involved in Watson-Crick base pairing with another polynucleotide chain. dsDNA involves two or more complementary polynucleotide chains, in which the two polynucleotide
15 chains are at least partially Watson-Crick base-paired to each other. Double-stranded DNA can also include a segment of DNA that is partially ssDNA and partially dsDNA, for example if there are gaps in one polynucleotide chain of a segment of dsDNA, such as a DNA including 5' or 3' overhangs. ssDNA and dsDNA may contain nucleotide analogs, nonnaturally occurring or synthetic
20 nucleotides, biotin, or epitope or fluorescent tags. ssDNA or dsDNA may be labeled; typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

Site-specific recombinase: a recombinase whose activity is limited to DNA of a specific sequence. Examples include the Cre, FLP and FLPE recombinases.
25 Lambda Int is specific for its att site. Cre recombinases are site-specific for *loxP* recombination sites in a DNA sequence, whereas FLP and FLPE recombinases are site-specific for *FRT* recombination sites. A recombination site is a nucleic acid sequence specifically recognized by a recombinase. For example, the Cre recombinase specifically binds a *loxP* recombination site, and thereby induces
30 recombination.

Substantially purified: a polynucleotide or polypeptide which is substantially free of other nucleotides, proteins, lipids, carbohydrates or other materials with which it is naturally associated. For example, a polypeptide may be at least 50%, 80% or 90% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In another example, a polynucleotide
5 may be isolated from the nucleotide sequences found 5' and 3' from the sequence in a wild-type cell.

Target nucleic acid sequence: the nucleic acid segment which is targeted for homologous recombination. Typically, this is a segment of chromosomal or
10 extrachromosomal DNA in a cell. Extrachromosomal DNA harboring target nucleic acid sequences may include episomal DNA, plasmid DNA, bacterial artificial chromosome, phagemid artificial chromosomes, yeast artificial chromosomes, cosmids, and the like. In one embodiment, the target nucleic acid sequence harbors a gene or gene fragment which will be mutated in some fashion upon homologous
15 recombination. Examples of target nucleic acid sequences include DNA sequences surrounding the tyr 145 UAG amber mutation of galK, as described in Yu et al., *Proc. Natl. Acad. Sci.* 97:5798-5983, 2000, and in Example 3 of this application; the second exon of mouse hox 1.1 gene, as described in U.S. Patent No. 5,464,764; the human hemoglobin S gene mutation as described in Example 15 of published PCT
20 Application No. WO 00/214495 A2, which is herein incorporated by reference.

Targeting frequency: the frequency with which a target nucleic acid sequence undergoes homologous recombination. For example, extrachromosomal DNA is introduced into a eukaryotic cell. The extrachromosomal DNA has sequences capable of undergoing homologous recombination with a target
25 intrachromosomal DNA sequence. After introducing the extrachromosomal DNA and allowing homologous recombination to proceed, the total number of cells may be determined, and the number of cells having the target DNA sequence altered by homologous recombination may be determined. The targeting frequency is the number of cells having the target DNA sequence altered, divided by the total
30 number of cells. For example, if there are a total number of one million cells, and 1,000 of these cells have the target DNA sequence altered, then the targeting frequency is 1 in 1,000, or 10^{-3} .

Transformed: as used herein, the term transformation encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transfection with viral vectors, transformation with plasmid vectors, and introduction of DNA (including DNA linked to Beta protein) by electroporation, lipofection, and biolistics.

Upstream: nucleic acid sequences 5' to a nucleic acid sequence of interest. In one embodiment, "upstream" refers to nucleic acid sequences that precede the codons that are transcribed into a RNA of interest. Similarly, downstream refers to nucleic acid sequences 3' of a nucleic acid of interest, such as, but not limited to, a nucleic acid sequence that follows codons that are transcribed into a RNA of interest.

Variants of Amino Acid and Nucleic Acid Sequences: the production of lambda Beta, Exo or Gam, or another ssDNA binding polypeptide can be accomplished in a variety of ways. DNA sequences which encode for the protein, or a fragment of the protein, can be engineered such that they allow the protein to be expressed in eukaryotic cells, bacteria, insects, and/or plants. In order to accomplish this expression, the DNA sequence can be altered and operably linked to other regulatory sequences. The final product, which contains the regulatory sequences and the nucleic acid encoding the therapeutic protein, is operably linked into a vector, allowing stable maintenance in a cell. This vector can then be introduced into the eukaryotic cells, bacteria, insect and/or plant. Once inside the cell, the vector allows the protein to be produced.

One of ordinary skill in the art will appreciate that the DNA can be altered in numerous ways without affecting the biological activity of the encoded protein. For example, PCR may be used to produce variations in the DNA sequence which encodes lambda Beta, Exo or Gam, or other ssDNA binding proteins. Such variants may be variants that are optimized for codon preference in a host cell that is to be used to express the protein, or other sequence changes that facilitate expression.

Two types of cDNA sequence variant may be produced. In the first type, the variation in the cDNA sequence is not manifested as a change in the amino acid sequence of the encoded polypeptide. These silent variations are simply a reflection

-24-

of the degeneracy of the genetic code. In the second type, the cDNA sequence variation does result in a change in the amino acid sequence of the encoded protein. In such cases, the variant cDNA sequence produces a variant polypeptide sequence. In order to preserve the functional and immunologic identity of the encoded polypeptide, such amino acid substitutions are ideally conservative in highly conserved regions. Conservative substitutions replace one amino acid with another amino acid that is similar in size, hydrophobicity, etc. Outside of highly conserved regions, non-conservative substitutions can more readily be made without affecting function of the protein. Examples of conservative substitutions are shown in Table 1 below.

Table 1

	Original Residue	Conservative Substitution
	Ala	Ser
	Arg	Lys
15	Asn	Gln, His
	Asp	Glu
	Cys	Ser
	Gln	Asn
	Glu	Asp
20	Gly	Pro
	His	Asn; Gln
	Ile	Leu, Val
	Leu	Ile; Val
	Lys	Arg; Gln; Glu
25	Met	Leu; Ile
	Phe	Met; Leu; Tyr
	Ser	Thr
	Thr	Ser
	Trp	Tyr
30	Tyr	Trp; Phe
	Val	Ile; Leu

Variations in the cDNA sequence that result in amino acid changes, whether conservative or not, should be minimized in order to preserve the functional and immunologic identity of the encoded protein. The immunologic identity of the protein may be assessed by determining whether it is recognized by an antibody to
5 the protein; a variant that is recognized by such an antibody is immunologically conserved. Particular examples of cDNA sequence variants introduce no more than 20, and fewer than 10 amino acid substitutions, into the encoded polypeptide. Variant amino acid sequences may, for example, be at least 80, 90 or even 95% identical to the native amino acid sequence.

10 **Yeast artificial chromosome (YAC):** a vector used to clone DNA fragments (up to 400 kb); it is constructed from the telomeric, centromeric, and replication origin sequences needed for replication in yeast cells (see *Current Protocols in Molecular Biology*, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates)).

15 **Wild-type:** the genotype or phenotype that is found in nature or in the standard laboratory stock for a given organism.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art
20 to which this disclosure belongs. The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are
25 approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The term “comprises” means “includes.” All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In
30 case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Recombineering using Host Cells Deficient for Mismatch Repair

Recombineering, or the use of a recombinase to mediate recombination using homology sufficient to induce recombination, as disclosed herein, can be performed using single-stranded oligonucleotides (oligos) as the targeting cassette. At certain
5 positions on the chromosome, greater than 1% of the SS oligo-treated cells have been shown to be recombinant, making it possible to screen for mutants without using a selectable marker (Ellis et al., *Proc. Natl Acad Sci USA* 98:6742-6746, 2001; Swaminathan et al., *Genesis* 29:14-21, 2001). Single stranded oligos have also been used to effect genetic modifications in yeast (Moerschell et al., *Proc. Natl Acad Sci
10 USA* 85:524-528, 1988; Liu et al., *Mol. Cell. Biol.* 22:3852-3863, 2002; Liu et al., *Nucleic Acids Res* 29:4238-4250, 2001) and in mammalian cells (Gamper et al., *Nucleic Acids Res* 28:4332-4339, 2002; Igoucheva et al., *Gene Ther.*, 8:391-399, 2001). Recombineering can also be performed using double-stranded nucleic acid, or double stranded nucleic acid with a 5' and or a 3' overhang (see published PCT
15 Application No. WO 00214495A2, which is herein incorporated by reference). However, a need remains to increase the frequency of recombination in both eukaryotic and prokaryotic cells. This disclosure provides methods and host cells can be used to increase the efficiency of recombineering.

The recombineering methodology utilizes recombination functions (e.g.
20 phage recombination functions) under control of a de-repressible promoter to generate recombination products using homologies of at least 20 base pairs. Thus, in one embodiment, recombineering uses a cell including Beta under the control of a de-repressible promoter. In specific, non-limiting example, expression of Beta alone (without Exo and Gam) is under the control of the de-repressible promoter (e.g. the
25 nucleic acid encoding Beta is operably linked to the de-repressible promoter). In another embodiment, expression of Beta, in addition to Gam and/or Exo, is under the control of the de-repressible promoter. In further embodiments, RecT, P22 Erf, or Rad52 is operably linked to a de-repressible promoter. In yet another
embodiment, DNA bound to a Beta protein is introduced into a host cell.

30 In recombineering, phage recombination functions can be used to introduce recombination into a target nucleic acid sequence in a host cell. The host cell can be eukaryotic or prokaryotic. In specific non-limiting examples, the host cell is a

-27-

mammalian cell, a yeast cell, or a bacterial cell (e.g. *E. coli*). The target can be on the chromosome, or can be on an extra-chromosomal element. In several specific, non-limiting examples, the target nucleic acid can be included in a plasmid, a bacterial artificial chromosome (BAC), a yeast artificial chromosome, a cosmid or a
5 vector, including but not limited to a viral vector. In one specific non-limiting example recombination is induced in a BAC strain or a BAC DNA is introduced into strain carrying recombination functions.

The length of the homologous sequence can be varied. In several embodiments, the homology is at least 20, at least 25, at least 30, at least 40, at least
10 50, at least 75 or at least 100 nucleotides in length. However, larger regions of homology can also be utilized. Thus, in one embodiment, between about 20 and about 1,000 nucleotides of homologous sequence is utilized, or between about 40 and about 1,000 nucleotides of homologous sequence is utilized. In one specific, non-limiting example, the ssDNA is about 20, about 25, about 30, about 40, about
15 50, about 75 or about 100 nucleotides in length. In embodiment, the homologous nucleic acid is a single-stranded nucleic acid. In another embodiment, the homologous nucleic acid is a double stranded nucleic acid. Double stranded nucleic acids include molecules that are completely double stranded, as well as nucleic acid molecules that have a 5' or a 3' overhang.

20 A single-stranded nucleic acid or double-stranded nucleic acid including sufficient homology to the target sequence is introduced into the host cell. "Sufficient homology" is any region of sufficient identity to the target sequence such that recombination can occur. In several embodiments, sufficient homology includes a sequence of at least 20 nucleotides in length, wherein at most five, at most
25 three, at most two, at most one nucleotide, or no nucleotides differ from the target nucleic acid sequence. In additional embodiments, sufficient homology includes a sequence of at least 25 nucleotides in length, wherein at most five, at most three, at most two, at most one nucleotide, or no nucleotides differ from the target nucleic acid sequence. Similarly, sufficient homology can readily be determined for a
30 nucleic acid of at least 30, at least 40, at least 50, or at least 100 nucleotides in length.

If the single-stranded nucleic acid or double-stranded nucleic acid differs from the target nucleic acid, these differences can be clustered (i.e. at one area in the target nucleic acid) or can be scattered in the sequences (e.g. two nucleotide differences from the target sequence, wherein each difference is located at different area in the sequence. In another embodiment, sufficient homology includes about a 5 100%, 99%, 98%, or 97% sequence identity between the homologous nucleic acid (e.g., the single-stranded or the double-stranded nucleic acid) and the target nucleic acid sequence. In another specific, non-limiting example, sufficient homology includes at least 90% sequence identity between the single-stranded or double- 10 stranded nucleic acid and the target nucleic acid, such as nucleic acid sequences that are at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, identical. It should be noted that a homologous nucleic acid sequence can differ from a target nucleic acid by substitutions, deletions, and/or additions of nucleotides. In another embodiment, the 15 single stranded nucleic acid (or double-stranded nucleic acid) is labeled, such as with a biotinylated nucleotide, a methylated nucleotide, or a DNA adduct.

The homologous nucleic acid (e.g., the single-stranded nucleic acid or double-stranded nucleic acid) can be introduced into the host cell by any means known to one of skill in the art. As disclosed herein, it is advantageous that the host 20 cell is deficient in mismatch repair. A host cell "deficient" for mismatch repair can repair mismatched nucleotides at a reduced frequency as compared to a wild-type cell, such as at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% reduction in mismatch repair as compared to a wild-type cell. In one specific, non-limiting example, mismatch repair is reduced at least 90% as compared to a wild-type cell. 25 A host cell deficient for mismatch repair can include a mutation in a nucleic acid sequence encoding a protein involved in mismatch repair, such that the protein has reduced function (or its function is eliminated). Thus, the function of one or more mismatch repair proteins can be reduced at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% as compared to the function of the protein in a wild-type cell. In 30 several embodiments, the function of one or more mismatch repair proteins is decreased at least 80%, such as 90%, 95%, 96%, 97%, 98%, 99%, or is completely absent in the host cell deficient for mismatch repair as compared to a wild-type cell.

In this context, a wild-type cell is a cell of the same species that does not include a mutation in the gene encoding the protein involved in mismatch repair.

In one embodiment, mismatch repair can be constitutively reduced in the host cell. Thus, if the cell is a prokaryotic cell, a cell that is deficient for mismatch
5 repair can have a mutation in one or more nucleic acids encoding mutS, mutH, mutL, uvrD, or dam. The mutS, mutH, mutL, uvrD, or dam protein produced from the mutated gene has a substantially reduced (or no) function in mismatch repair. Thus, a corresponding wild-type cell does not have a mutation in the nucleic acid encoding MutS, MutH, MutL, uvrD, or dam, respectively. A cell deficient for
10 mismatch repair can also have more than one mutation or the nucleic acid encoding MutS, MutH, MutL, uvrD, or dam, or can have mutations in more than one of these genes. The mutation can be an insertion, deletion, or a point mutation. Thus, in several specific, non-limiting examples, a prokaryotic cell deficient for mismatch repair has a mutation in a nucleic acid encoding Mut S (mutS-, or Δ mutS), MutH
15 (mutH- or Δ mutH), MutL (mutL- or Δ mutL), UvrD (uvrD- or Δ uvrD), or Dam (dam- or Δ dam), or a combination (e.g. mutS-mutH- (Δ mutS Δ mutH), mutS-mutL- (Δ mutS Δ mutL), mutH-mutL- (Δ mutH Δ mutL), mutH-uvrD- (Δ mutH Δ uvrD), etc.).

The proteins involved in mismatch repair are also known for eukaryotic cells. Eukaryotes have proteins with sequence similarity to MutS and MutL that are
20 involved in a similar repair pathway. The eukaryotic MutS is a dimer of MSH2 and GTBP (now known as MSH3 or MSH6) proteins. Eucaryotic mutL also consists of two polypeptides, MLH1 and PMS2. Eukaryotes lack homologs of MutH and uvrD. Exemplary genes encoding mismatch repair proteins are disclosed in the table shown below:

Table 2
Genes encoding enzymes of mismatch repair

<i>E. coli</i>	<i>S. cerevisiae</i>	Human	Functions of Eukaryotic proteins
		?	DNA repair in mitochondria
<i>MutS</i>	MSH1		
<i>MutS</i>	MSH2	MHS2	Single mismatch and small loop repair (with MSH6 to form MutSalpha); loop repair (with MSH3 to form MutSbeta)
<i>MutS</i>	MSH3	MSH3	Loop repair (with MSH2 to form MutSbeta)
<i>MutS</i>	MSH4	MSH4	Meiosis (with MLH1)
<i>MutS</i>	MSH5	MSH5	Meiosis (with MLH1)
<i>MutS</i>	MSH6	MSH6	Single mismatch and small loop repair (with MSH2 to form MutSalpha)
<i>MutL</i>	MLH1	MLH1	Mismatch repair
<i>MutL</i>	PMS1	PMS2	Mismatch repair (with MLH1 to form MutLalpha)
<i>MutL</i>	MLH2	PMS1	Interacts with MLH1 to form MutLbeta. Not involved in mismatch repair in yeast.
<i>MutL</i>	MLH3	MLH3	Loop repair (with MLH1)
<i>MutH</i>	?	-	-
<i>uvrD</i>	?	-	Helicase, generates single strand
<i>recJ</i>	Exonuclease 1	Exonuclease 1	Mismatch repair (5' to 3' polarity)
?	RAD27	DNase IV FEN-1	Mismatch repair (Flap endonuclease)
<i>xonA</i>			Exonuclease (3' to 5' polarity)
<i>dam</i>	-	-	methylase

- 5 It should be noted that *S. pombe* also has two genes encoding MutS functions (MSH2, SW14, and MSH6), two genes encoding MutL functions (MLH1 and PMS1), and an EXO1 gene. *C. elegans* has two genes encoding MutS function (MSH2 and MSH6), and two genes encoding MutL functions (MLH1 and PMS1). *D. melanogaster* has two genes encoding MutS function (SPEL1 and MSH6) and

two genes encoding MutL functions (MLH1 and PMS2), and the TOSCA gene. In addition, *A. thaliana* has four genes encoding MutS functions (MSH2, MSH3, MSH6, MSH7), three genes encoding MutL functions (PMS1 and MLH3), and two additional genes involved in mismatch repair (Q9C7N8 and AAK913436, which
5 may be homologous to EXO1) (for review see Marti et al., *J. Cell. Physiol.* 191:28-41, 2002). As disclosed herein, eukaryotic host cell deficient for mismatch repair can have one or more mutations in a gene encoding mismatch repair, or can have mutations in more than one gene involved in mismatch repair.

In another embodiment, the method utilizes a cell that is transiently deficient
10 in mismatch repair. Thus, a cell deficient for mismatch repair can also be a cell treated to inactivate mismatch repair for a specified period of time. In one specific, non-limiting example, a prokaryotic cell (e.g., *E. coli*) can be treated with a compound, such as 2-aminopurine, to induces a reversible phenotype of DNA mismatch repair deficiency (see Matic et al., *J. Bacteriol.* 185:1459-1461, 2003). In
15 another specific, non-limiting example, a cell can be treated with an effective amount of multi-copy single-stranded DNA (see Maas et al., *Molec. Microbiol* 19:505-509, 1996). Multi-copy single-stranded DNA is a single-stranded DNA covalently linked to DNA that forms stem-loop structures. Treatment with multi-copy single-stranded DNA with mismatched base pairs is known to inhibit mismatch
20 repair. Without being bound by theory, this effect is believed to be due to a titration of MutS (see Maas et al, *supra*, 1996).

In another embodiment, antisense, small inhibitory RNAs, ribozymes, or other nucleotide based strategies are used to induce a decrease in mismatch repair. Antisense nucleic acids are DNA or RNA molecules that are complementary to at
25 least a portion of a specific mRNA molecule (Weintraub, *Scientific American* 262:40, 1990), such a nucleic acid encoding a protein involved in mismatch repair. In the cell, the antisense nucleic acids hybridize to the corresponding mRNA, forming a double-stranded molecule. The antisense nucleic acids interfere with the translation of the mRNA, since the cell will not translate a mRNA that is double-
30 stranded. Antisense oligomers of about 15 nucleotides are preferred, since they are easily synthesized and are less likely to cause problems than larger molecules when introduced into the target cell. The use of antisense methods to inhibit the *in vitro*

translation of genes is well known in the art (Marcus-Sakura, *Anal. Biochem.* 172:289, 1988).

Use of an oligonucleotide to stall transcription is known as the triplex strategy since the oligomer winds around double-helical DNA, forming a
5 three-strand helix. Therefore, these triplex compounds can be designed to recognize a unique site on a chosen gene (Maher et al., *Antisense Res. and Dev.* 1:227, 1991; Helene, *Anticancer Drug Design* 6:569, 1991), such as a gene encoding a protein involved in mismatch repair.

Ribozymes are RNA molecules possessing the ability to specifically cleave
10 other single-stranded RNA in a manner analogous to DNA restriction endonucleases. Through the modification of nucleotide sequences which encode these RNAs, it is possible to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, *J. Amer. Med. Assn.* 260:3030, 1988). A major advantage of this approach is that, because they are
15 sequence-specific, only mRNAs with particular sequences are inactivated.

There are two basic types of ribozymes namely, *tetrahymena*-type (Hasselhoff, *Nature* 334:585, 1988) and "hammerhead"-type. *Tetrahymena*-type ribozymes recognize sequences which are four bases in length, while
A hammerhead"-type ribozymes recognize base sequences 11-18 bases in length.
20 The longer the recognition sequence, the greater the likelihood that the sequence will occur exclusively in the target mRNA species. Consequently, hammerhead-type ribozymes are preferable to *tetrahymena*-type ribozymes for inactivating a specific mRNA species and 18-based recognition sequences are preferable to shorter recognition sequences.

25 A "transfected cell" or a "transformed cell" is a cell into which has been introduced, by means of recombinant DNA techniques, a nucleic acid molecule. The nucleic acid molecule can be single-stranded or double-stranded (including a double-stranded nucleic acid with an overhang). The introduction of a nucleic acid molecule into a host cell can be carried out by conventional techniques as are well
30 known to those skilled in the art. Where the host is prokaryotic, such as *E. coli*, competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the CaCl₂

method using procedures well known in the art. Alternatively, MgCl₂ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell if desired, or by electroporation.

When the host is a eukaryote, such methods of transfection of DNA as
5 calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors may be used. Eukaryotic cells can also be cotransformed with DNA sequences including sequence of interest, and a second foreign DNA molecule encoding a selectable phenotype, such as neomycin resistance (see for example,
10 *Eukaryotic Viral Vectors*, Cold Spring Harbor Laboratory, Gluzman ed., 1982). Other specific, non-limiting examples of viral vectors include adenoviral vectors, lentiviral vectors, retroviral vectors, and pseudorabies vectors.

The homologous nucleic acid (e.g. the single-stranded nucleic acid or double-stranded nucleic acid) is introduced into the host cell deficient for mismatch
15 repair, the de-repressible promoter is activated, and recombinants are generated *in vivo*. Thus, in one specific, non-limiting example, if the de-repressible promoter is PL, the host cell is treated with heat to induce the expression of Beta (see Copeland et al., *Nature Reviews* 2:769, 2001, and Ellis et al., *Proc. Natl. Acad. Sci.* 98:6742-6746, 2001, which are herein incorporated by reference). Generally, the
20 homologous nucleic acid, whether it is a single-stranded nucleic acid or a double-stranded nucleic acid, differs from the target nucleic acid by at least one nucleotide, but is sufficiently homologous to undergo recombination with the target sequence (see above).

Recombinants can be detected by any means known to one of skill in the art.
25 If recombination has occurred in a nucleic acid encoding a marker, such as a nucleic acid encoding a polypeptide involved in antibiotic resistance, detection can be performed by selection or counterselection. However, detection can also be performed by direct screening (e.g. colony hybridization or sequencing). Detection can also be performed by detecting a label on the nucleic acid (e.g. when DNA
30 includes a DNA adduct or a marker such as biotin).

As has been described in *E. coli* (see published PCT application No. WO 00/214495 A2, herein incorporated by reference), a single base change has been

-34-

substituted in the *galk* gene and a 3.3 kbp insertion removed from the *galk* gene using single-stranded oligos. Single-stranded oligos have also been used to cure 5 different *Th10* insertions at different places on the *E. coli* chromosome. Whereas Exo, Beta, and Gam facilitate recombination of PCR amplified dsDNA cassettes with flanking homologies, only Beta is required for ssDNA recombination.

Recombination with either of two complementary DNA oligos has revealed that although either strand can be efficiently used for recombination, one strand is more competent for recombination than the other. This strand bias has been examined at several positions around the bacterial chromosome with the result that the preferred strand correlates with the lagging strand of DNA replication for each site tested. Without being bound by theory, these results indicated that strand bias is associated with the replication direction through the region being targeted and that ssDNA recombination occurs efficiently near the replication fork. The process of DNA replication results in transient regions of ssDNA that may be accessible to Beta-mediated annealing of the ssDNA oligo. Although recombination occurs on the leading strand, the increased recombination efficiency of the lagging strand oligos could reflect the increased frequency of single-stranded regions during lagging versus leading strand synthesis. DNA polymerase and DNA ligase could then complete the joining of the annealed oligo to the lagging strand (see PCT application No. WO 00214495 A2, which is herein incorporated by reference). Without being bound by theory, mutations in a mismatch repair protein, such that mismatch repair is reduced in the host cell, do not change the stand bias for recombination of single stranded DNA. However, a deficiency in mismatch repair results in an increased detection of recombinants, whether a recombination occurs with the leading or the lagging strand of the target double-stranded DNA (see Fig. 1).

Fig. 1 shows is a schematic diagram of repair of a mismatch when single-stranded nucleic acid is utilized. When double-stranded DNA that does not include 3' or 5' overhangs is used for linear recombination, *exo*, *bet*, and *gam* gene products are required. However, when single stranded nucleic acid, or when double-stranded nucleic acid with a 3' and/or 5' overhang is utilized, only Beta is required. Without being bound by theory, a DNA structure is suggested that is shown in the top of Fig.

-35-

2. In this model, a linear double-stranded DNA with flanking 3' single-stranded DNA overhangs is generated as an annealing intermediate with Beta bound. Such a structure made *in vitro* and electroporated into a cell is recombinogenic; this recombination does not require Exo. Experiments have been performed in which
5 the 3' overhangs were constructed by co-electroporating two oligonucleotides that are 70 nucleotides in length that pair only in the middle. As predicted by the model shown in Fig. 2, this substrate recombined and more importantly, required only Beta. The presence or absence of Exo had no effect. This recombination was also found to be *recA*-independent. Because single-stranded DNA ends are involved, this
10 intermediate, like single stranded DNA, is likely to initiate recombination at a DNA replication fork (Figs. 2 and 3). A mismatch may be left in both daughter DNA molecules.

Cells of Use

15 Isolated host cells of use in the methods disclosed herein include a de-repressible promoter operably linked to a nucleic acid encoding a recombinase. These cells are also deficient for mismatch repair. In one embodiment, the cell is constitutively deficient for mismatch repair. The cell can be a eukaryotic cell or a prokaryotic cell. In one embodiment, the cell is a bacterial cell, such as a gram
20 positive or a gram negative bacterial cell. In one specific, non-limiting example, the bacteria is a gram negative bacteria, such as an enteric bacteria (e.g. *E. coli*). In another embodiment, the cell is a eukaryotic cell. Thus, the cell can be a yeast cell or a mammalian cell. other specific, non-limiting examples, the cell can be a *S. cerevisiae* cell, and *S. pombe* cell, a *C. elegans* cell, a *D. melanogaster* cell, an *A. thaliana* cell, a murine cell, or a human cell.
25

Thus, in one embodiment, the cell has a mutation in a gene encoding a protein of mismatch repair (see Table 2), in addition to including a de-repressible promoter operably linked to a nucleic acid encoding a recombinase. Mutations in a gene encoding the mismatch repair protein can be made or selected by any means
30 known to one of skill in the art. Methods for introducing a mutation in a gene, or for "knocking out" a gene are known for both prokaryotic and eukaryotic cells (see for example, *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, ed. Sambrook

et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, and *Current Protocols in Molecular Biology*, ed. Ausubel et al., Greene Publishing and Wiley-Interscience, New York, 1987 (with periodic updates)).

Targeted disruption of a gene *in vivo* with complete loss of function has been
5 achieved by any transgenic technology familiar to those in the art. In one
embodiment, gene knock-outs are utilized in which a target gene has been rendered
nonfunctional by an insertion targeted to the gene to be rendered non-functional by
homologous recombination. As used herein, the term “transgenic” includes any
transgenic technology familiar to those in the art which can produce a cell or an
10 organism carrying an introduced transgene or one in which an endogenous gene has
been rendered non-functional or knocked out.

In one specific, non-limiting example, a specific mutation is introduced into
a mismatch repair gene in the host cell. In another embodiment, the entire gene may
be deleted. Optionally, a disruption or deletion in a gene encoding a protein
15 involved in mismatch repair may be accompanied by insertion of or replacement
with other DNA sequences, such as a nucleic acid encoding a selectable marker
(such as, but not limited to, an nucleic acid conferring resistance to an antibiotic). In
other embodiments, the host cell includes a DNA antisense to the coding sequence
for the nucleic acid encoding the mismatch repair protein, or a ribozyme that
20 specifically targets the nucleic acid encoding a mismatch repair protein.

In one embodiment, a gene encoding mismatch repair (e.g. see Table 2
above) is disrupted by homologous targeting in a eukaryotic cell. These include
lower organisms such as yeasts, slime molds, and the like, as well as cells from
multicellular organisms such as invertebrates, vertebrates, and mammals. Suitable
25 cells include, but are not limited to, stem cells such as hematopoietic progenitor
cells, hematopoietic stem cells, and embryonic stem cells. Suitable cells also
include differentiated cells, including tumor cells and cell lines. The cell can be a
terminally differentiated cell, such as a differentiated cell from any organ, including
but not limited to cells from the circulatory system, respiratory system, reproductive
30 system, lymphoid organs, excretory system. These include cells in a variety of
tissue types, such as: endothelial cell, smooth muscle cell, epithelial cell, hepatocyte,
cells of neural crest origin, tumor cell, hematopoietic cell, lymphoid cell, T cell, B

cell, monocyte, macrophage, dendritic cell, fibroblast, keratinocyte, neuronal cell, glial cell, adipocyte, myoblast, myocyte, chondroblast, chondrocyte, osteoblast, osteocyte, osteoclast, secretory cell, endocrine cell, oocyte, and spermatocyte.

In another embodiment, the cell is a prokaryotic cell. Thus, in one specific, non-limiting example, the cell is a gram negative bacteria, such as, but not limited to, an enteric bacteria. Thus, in one specific, non-limiting example the cell is an *E. coli* cell that is deficient for mismatch repair and includes a nucleic acid encoding a single-stranded binding protein, operably linked to a de-repressible promoter. In one embodiment the single-stranded binding protein is Beta. The cell can optionally also include a nucleic acid sequence encoding Exo and/or Gam.

The bacterial cell can be generated using a mini-lambda (see PCT WO 00/214495 A2, which is herein incorporated by reference). Thus, in this embodiment, the host cell is deficient for mismatch repair and includes a mini-lambda. In another specific, non-limiting example, the cell is an isolated bacterial cell comprising a defective lambda prophage of genotype $\lambda cI857 \Delta(cro-bioA)$, wherein the bacterial cell further comprises a mutation in at least one nucleic acid sequence encoding a mismatch repair polypeptide. Thus, in addition to including a de-repressible promoter operably linked to a nucleic acid encoding a recombinase (e.g. encoding Beta), the bacterial cell also includes a mutation in one or more of MutS, MutH, MutL, uvrD, and/or dam. Methods for introducing mutations in a bacterial cell are well known in the art (e.g. see Yu et al. *Proc. Natl Acad. Sci. USA* 97:5978-5983, 2000, herein incorporated by reference), and exemplary methods are disclosed in the examples section below). The isolated bacterial cell can have mutations in two or more of these genes. Thus, in several specific, non-limiting examples, a prokaryotic cell deficient for mismatch repair has a mutation in two or more of MutS (MutS-), MutH (mutH), MutL (mutL-), uvrD (uvrD-), or dam (dam-), for example, the cell is an isolated bacterial cell having a MutS-MutH-, MutS-MutL-, MutH-MutL-, or a MutH-uvrD-, genotype. In bacterial cells, there are additional proteins that have one or more functions of the dam methylase. Thus, suitable host cells can also include mutations in additional genes that encode polypeptides with a dam methylase function. Thus, in one specific, non-limiting example, the host cell includes mutation in a gene encoding a polypeptide with a dam methylase function

-38-

(that is not dam), one or more mutations in MutS, MutH, MutL, or uvrD, and/or dam. In another specific non-limiting example, the host cell includes a mutation in a gene encoding a polypeptide with a dam methylase function that is not dam, and a mutation in dam.

5 In another specific, non-limiting example the bacterial cell can be derived from HME6, DY380 or a DY330 cell, such that a mutation is introduced into a gene encoding a protein involved in mismatch repair. One of skill in the art can readily produce these cells using known methods. The generation of several exemplary cells, and the genotype of these cells, is disclosed in the Examples section below
10 (e.g., see Tables 3 and 7).

Exemplary bacterial strains have been deposited as ATCC NO. ____ (HME6 (Mut+)), ATCC No. ____ (HME63 (MutS)), ATCC No. ____ (HME 61 (MutH)), ATCC No. ____ (HME62 (MutL)), and ____ (HME64 (uvrD)), all of which were deposited on May 7, 2003.

15 These cells, and exemplary specific methods to produce cells deficient in a protein involved in mismatch repair, are further described in the Examples below.

The disclosure is illustrated by the following non-limiting Examples.

20

EXAMPLES

Example 1

Materials and Methods

The lambda Red recombination system has been extensively used to modify the bacterial chromosome with single-strand oligonucleotides (ss oligo). As shown
25 previously, a strand bias of the ss oligo can be demonstrated with regards to chromosome replication but not to transcription direction. Both the host MMR system and sequence of the single-strand oligonucleotide play an important role when a single base change is repaired.

30 Single-strand DNA oligonucleotides (ss oligos) have been used to modify the chromosome in both yeast and *E. coli* (Ellis et al., *Proc. Natl. Acad. Sci. USA* 98:6742-6746, 2001). The lambda Red recombination system has been used with ss oligos to create mutations, to correct single base mutations, to create deletions, and

to remove large heterologies from the *E. coli* chromosome. The λ Red system includes the Beta, Exo and Gam proteins. However, recombination with ss oligos requires only the Beta function. Oligos with homologies as short as 20 bases can be used to recombine in *E. coli* and as little as 10 ng of oligo yields saturating levels of
5 recombinants (Ellis et al., *supra*).

For any allele the oligo corresponding to either of the two complementary strands carrying that allele can be used to initiate recombination. For several cases tested, there was a marked strand bias for one of the two complementing oligos used as a recombination substrate (Ellis et al., *supra*). This strand preference appeared to
10 be correlated to the direction of DNA replication through the region being recombined; the direction of transcription through the allele did not appear to cause the bias. As disclosed herein, the primary bias observed is not caused by transcription but is determined by the direction of DNA replication.

Beta can anneal complementary strands but cannot cause a ssDNA to invade
15 a duplex DNA; RecA function can cause strand invasion but is not required for Beta-mediated recombination. Therefore, Beta-mediated recombination with a ss oligo is likely to be directed to a single-strand region of the chromosome. A DNA replication fork provides an ideal area of the chromosome with exposed single-strand regions. The strand bias observed for oligo recombination is dependent upon
20 replication direction. The oligo that corresponds in sequence to Okazaki fragments generates the highest efficiency. Such oligos are referred to herein as the lagging strand oligo and their complement as the leading strand oligo with reference to the replication fork.

Many host proteins are involved in progression of the replication fork. As
25 disclosed herein, a decrease in the proof-reading system of methyl-directed mismatch repair (MMR) results in an increase in recombinant yield when a ss oligo is utilized. The host MMR system acts to correct small base pair mismatches by first recognizing, then excising the incorrect base. Without being bound by theory, the changes that the ss oligos introduce at the replication fork may trigger mismatch
30 repair. Thus, decreasing MMR can affect the frequency that homologous recombination can be detected. The following material and methods were used in the Examples.

-40-

Genotype of strains: HME5 is W3110 $\Delta(\text{argF-lac})\text{U169 gal}^+$, $\{\lambda\text{cI857}\Delta\text{cro-bioA}\}$. HME6 is HME5 *galK tyr145UAG*, HME31 is HME5 *galK<>cat-sacB*, and HME41 is HME6 with the entire *gal* operon inverted without inverting the adjacent genes. These strains have previously been described (Ellis et al., *supra*). HME6,
 5 HME31, and HME41 were made *recA*⁻ by P1 transduction to move in $\Delta\text{srl-recA}::\text{Tn10}$.

Deletion of host factor genes: Host MMR genes *mutH*, *mutL*, *mutS*, *uvrD*, and *dam* were inactivated by inserting PCR generated antibiotic resistant cassettes in place of the coding region of each gene (see Table 3 below).

10

Table 3**Knock outs of methyl-directed mismatch repair genes¹**

	<u>dam</u> <>kan	(5'dam AGGGGGCAAGTA-kan-TTCTCAAGGAGA 3'dam) (5'dam SEQ ID NO: 1-kan-SEQ ID NO:2- 3'dam)
15	<u>mutH</u> <>amp	(5'mutH AGGTATCATGAC-amp-AGTGCACTACTG 3'mutH) (5'mutH SEQ ID NO: 3-amp-SEQ ID NO:4- 3'mutH)
	<u>mutL</u> <>amp	(5'mutL CAACTGGCGAAC-amp-TACATCCGGCGA 3'mutL) (5'mutL SEQ ID NO:5-amp-SEQ ID NO:6- 3'mutL)
20	<u>mutS</u> <>amp	(5'mutS GGACATAACCCC-amp-TAATAACAATTC 3'mutS) (5'mutS SEQ ID NO:7-amp-SEQ ID NO:8- 3'mutS)
	<u>uvrD</u> <>kan	(5'urvD GGACGTTTCTTA-kan-TAACGTTGCCGG 3'mutS) (5'urvD SEQ ID NO:9-kan-SEQ ID NO:10- 3'mutS)

¹Sequence in capital letters indicates chromosomal regions near the designated gene that flank the inserted antibiotic cassette.

25

Either a kanamycin or ampicillin resistance cassette was amplified by PCR using oligos that contained at their 5' ends 45-55 bases of homologies to the target MMR gene. The PCR cassettes with flanking homologies were used for recombination (Yu et al., *Proc. Natl. Acad. Sci. USA* 97:5978-5983, 2000);
 30 recombinants were selected for antibiotic resistance and tested by analytical PCR. Once the substitution was confirmed, it was moved by P1 transduction into the appropriate strains.

E. coli transformation: Electroporation-competent cells were prepared according to Yu et al., *Proc. Natl. Acad. Sci. USA* 97:5978-5983, 2000, herein
 35 incorporated by reference. Briefly, overnight cultures inoculated from a single

-41-

colony were diluted 50-fold in LB medium (total volume 10ml), and grown to $OD_{600} = 0.6$. Expression of Red functionality was then induced by shifting the cells to 42°C for 15 minutes followed by chilling on ice for 20 minutes. Cells were harvested by centrifugation, washed with ice-cold sterile water (3x 1.5ml). The cell
5 pellets were then resuspended in sterile water (50µl) and transformed with DNA by electroporation at 1.75 kV, 25 mF using the *E. coli* Pulser (BIO-RAD). LB-media (1 ml) was added then the cells were incubated at 32°C for 1.5 hours with shaking, before being spread onto the appropriate selective LB-agar media plates or minimal galactose (see Ellis et al., *Proc. Natl Acad. Sci. USA* 98:6742-6746, 2001).

10 *Materials:* Oligonucleotides and PCR primers were supplied by Invitrogen as salt free but otherwise not purified. All ss oligos used to correct the *galK* mutations were 70 bases in length. The sequence of Oligo 100 which corrects the *TAG* stop to a *TAT* tyrosine codon is: 5'
AAGTCGCGGTCGGAACCGTATTGCAGCAGCTTTATCATCTGCCGCTGGAC
15 GGCGCACAAATCGCGCTTAA 3' (SEQ ID NO: 11). Oligo 101 is the complementary strand to 100. Oligos 144 and 145 were identical to 100 and 101 respectively except that they correct the *TAG* stop to a *TAC* tyrosine codon. The detailed sequence of these ss oligos around the *galK* amber stop is shown in Example 2. Taq polymerase HiF and Concert Rapid PCR purification kits were
20 supplied by Invitrogen.

Assay for recombination to gal⁺ phenotype: Strains were induced at 42°C for 15 minutes to express λ Red functions and immediately made electro-competent as previously described (Yu et al., *Proc. Natl. Acad. Sci. USA* 97: 5978-5983, 2000). Saturating levels (100 ng) of each ss oligo were used per electroporation. Gal⁺
25 recombinant colonies were selected on minimal galactose plates and viable cells were counted on L plates as described previously (Ellis et al., 2000, *supra*).

*Plasmid construction, p(+)*mKan* and p(-)*mKan*:* DNA inserts for the plasmid constructs were amplified by PCR with Expand High-fidelity polymerase using primers containing restriction enzyme sites suitable for cloning. All inserts
30 were sequenced to check for possible mutations introduced by PCR. Primers mKan-1: 5'-GGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTAGGACTG-3',

-42-

SEQ ID NO: 12 and mKan-2 5'-TATTCGGCAAGCAGGCATCG-3', SEQ ID NO: 13 were used to amplify a 559 bp fragment from plasmid pGK-frt. The *EagI-NcoI* digested 559 bp PCR fragment was then ligated into the 4,287 bp *EagI-NcoI* fragment from pGK-frt to form pmKan. To construct p(+)mKan and p(-)mKan, the purified *EcoRI-ClaI* fragment from pmKan was ligated to the 2942 bp *EcoRI-ClaI* fragment of pBSKS and pBSSK to form p(+)mKan and p(-)mKan, respectively.

Construction of additional mutants: To construct *E. coli* strain DY380(+), PCR was used to amplify a 2,397 bp fragment from p(+)mKan using the primers
10 DY380(+)-1 (GGCGCTGCAAAAATTCTTTGTCGAACAGGGTG
TCTGGATCTAATGCGCCGCTACAGGGCGCGTAA, SEQ ID NO; 14)
and DY380(+)-2 (GGCGCTGCAAAAATTCTTTGTCGAACAGGG
TGTCTGGATCTAATGCGCCGCTACAGGGCGCGTAA, SEQ ID NO: 15)

The purified PCR fragment, which contained the mKan gene and an Amp
15 selectable marker flanked by two regions homologous to the chromosomal target,
was inserted into DY380 via Red-mediated homologous recombination as described
previously (Yu et al. *Proc. Natl Acad. Sci. USA* 97:5978-5983, 2000; Ellis et al.,
Proc. Natl Acad. Sci. USA 98:6742-6746, 2001). The recombination event inserts
the drug cassettes between the λ prophage *cro* gene and the *bioA* gene adjacent to
20 the prophage (see Lee et al., *Genomics* 73:56-65, 2001). Amp^r colonies were
selected then screened by PCR, using the primers DY380(+)-detect
(ACGTTAAATCTATCACCGCAAGGG-3', SEQ ID NO: 16) and PGK-frt-2
(5'-TATTCGGCAAGCAGGCATCG-3', SEQ ID NO: 17) to positively identify
those that contained the correctly inserted cassettes. Similarly, the other mutant
25 strains were generated by Red-mediated recombination using appropriate PCR-
generated cassettes. The normal and inverted *gal* operon with the *galKam* mutation
have been described by Ellis et al., *Proc. Natl Acad. Sci. USA* 98:6742-6746, 2001.

Table 4
Additional Bacterial strains

Strain	Genotype
DH10B	F ⁻ <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ80 <i>dlacZ</i> Δ <i>M15</i> Δ <i>lacX74 deoR recA1 endA1 araD139</i> Δ (<i>ara, leu</i>) 7649 <i>galU galK rspL nupG</i>
DY330	W3310 Δ <i>lacU169 gal490 λcI857</i> Δ (<i>cro-bioA</i>)
DY380	DH10B [<i>λcI857 (cro-bioA) < > tet</i>] ^a
DY330(+)	DY330 [(<i>cro-bioA</i>) < > (<i>mKan-Amp^r</i>)]
DY330(-)	DY330 [(<i>cro-bioA</i>) < > (<i>Amp^r-mKan</i>)] ^b
DY380(+)	DY380 [(<i>cro-bioA</i>) < > (<i>mKan-Amp^r</i>)]
DY380(-)	DY380 [(<i>cro-bioA</i>) < > (<i>Amp^r-mKan</i>)] ^b
DY380Δ <i>mutS</i>	DY380 [<i>mutS</i> < > <i>Cm^r</i>]
DY380(+) <i>ΔmutS</i>	DY380(+) [<i>mutS</i> < > <i>Cm^r</i>]
DY380(-) <i>ΔmutS</i>	DY380(-) [<i>mutS</i> < > <i>Cm^r</i>]
DY380(+) <i>ΔmutL</i>	DY380(+) [<i>mutL</i> < > <i>Cm^r</i>]
DY380(-) <i>ΔmutL</i>	DY380(-) [<i>mutL</i> < > <i>Cm^r</i>]
DY380(+) <i>ΔmutH</i>	DY380(+) [<i>mutH</i> < > <i>Cm^r</i>]
DY380(-) <i>ΔmutH</i>	DY380(-) [<i>mutH</i> < > <i>Cm^r</i>]
DY380(+) <i>Δmfd</i>	DY380(+) [<i>mfd</i> < > <i>Cm^r</i>]
DY380(-) <i>Δmfd</i>	DY380(-) [<i>mfd</i> < > <i>Cm^r</i>]
DY380(+) <i>ΔuvrB</i>	DY380(+) [<i>uvrB</i> < > <i>Cm^r</i>]
DY380(-) <i>ΔuvrB</i>	DY380(-) [<i>uvrB</i> < > <i>Cm^r</i>]
HME6	W3310 <i>galK_{tyr145UAG} λcI857</i> Δ (<i>cro-bioA</i>)
HME41	W3310 IN <i>gal</i> < > (<i>galM⁺ K_{tyr145UAG} T⁺ E⁺</i>) <i>λcI857</i> Δ (<i>cro-bioA</i>) ^c

^a The symbol < > indicates a replacement generated by recombinering technology. For example, (*cro-bioA*) < > *tet* indicates the substitution of *cro-bioA* with *tet*.

5 ^b *mKan-Amp^r* and *Amp^r-mKan* indicates that the *mKan* gene has been inserted in opposite directions in the two DY380 or DY330 strains.

^c IN*gal* indicates that the *gal* gene has been inserted in opposite directions in the HME6 strain.

10 Deletion of the *mfd* or *uvrB* genes (i.e. strains defective in TCR) led to only minor changes in the recombination efficiency. By deleting *mfd* or *uvrB* genes, it was demonstrated that the transcription-coupled repair pathway does not affect for Red/SSO-mediated recombination. As disclosed herein, SSO mediate recombination even in *recA* and *mutS* double mutant strains.

-44-

Example 2**Sequence affects recombination efficiency**

Oligo 100 and its complement 101 are designed to correct the *galK*_{tyr145am} to the original *TAT* tyrosine codon. Two other oligos are also made to correct the *galK* mutation but to a different tyrosine codon; Oligo 144 and its complement 145 create a *TAC* tyrosine codon (see below).

	OLIGO 100	CAG CTT TA t CAT CTG (SEQ ID NO:12)
	5'	or
10	OLIGO 144	CAG CTT TA c CAT CTG (SEQ ID NO: 13)
	5'	CAG CTT TAG CAT CTG 3' (SEQ ID NO: 14)
	3'	GTC GAA ATC GTA GAC 5' (SEQ ID NO: 15)
15	OLIGO 101	GTC GAA AT a GTA GAC (SEQ ID NO: 16)
		or
	OLIGO 145	GTC GAA AT g GTA GAC (SEQ ID NO: 17)

20

In these sequences, the G:C pair which creates an amber codon is shown in bold. The correcting base brought in by the repair oligos is shown in lower case.

Although Oligos 100 and 101 generate many *gal*⁺ recombinants, the relative efficiency of Oligo 144 recombination was much greater than that for Oligo 100 and Oligo 101 (see Table 5 below).

25

Table 5**Oligo differences in recombination frequency**

Strain	galK gene	oligo	DNA replic	Gal+ recombinants
30	HME6	100	lagging	2.9 x 10 ⁵
		144	lagging	4.7 x 10 ⁷
35		101	leading	4.3 x 10 ⁴
		145	leading	1.8 x 10 ⁴

-45-

HME31	◁ <u>cat</u> <u>sacB</u>	100	lagging	1.2×10^5
		144	lagging	9.5×10^4
5		101	leading	3.5×10^3
		145	leading	2.0×10^3

The frequency of Gal⁺ recombinants with Oligo 144 was 4.7×10^7 ,
 10 approximately 100-fold greater than for Oligo 100, which is identical except for the
 one base difference in the tyrosine codon. On the other hand, the frequency
 recombinants with Oligo 145 was similar to that of Oligo 101.

Oligo 144 and Oligo 145 correct *galK*◁*cat-sacB* with the same efficiency
 as Oligo 100 and Oligo 101, respectively. Thus, Oligo 144 yields a dramatic
 15 increase in recombinant yield with the *galK*_{tyr145am} point mutant but the same
 enhanced effect is not provided during recombination with the *cat-sacB* heterology
 containing the 3.3kbp insertion in *galK*. This result suggests that the pairing
 difference of Oligo 100 and Oligo 144 with the complementary strand of the
*galK*_{tyr145am} mutation might generate the differences in recombination efficiency
 20 observed.

The Oligo 100 *TAT* sequence generates a T/C mismatch whereas the Oligo
 145 *TAC* sequence generates a C/C mismatch. The MutS protein of the methyl-
 directed mismatch repair (MMR) system binds and corrects T/C as well as most
 other single base mismatches but does not bind and correct C/C mismatches. This
 25 suggests that C/C mismatch can be used to generate changes in a target nucleic acid
 sequence by homologous recombination using this system. The MMR system does
 not recognize mismatches created between large heterologies like those expected for
cat-sacB repair by the oligos. Thus, the low efficiency of repair of *cat-sacB* by these
 oligos is caused by an intrinsically poor recombination rate relative to that of the
 30 point mutation. This indicates that mismatch repair corrects base pair mismatches
 inserted during homologous recombination.

Example 3

Mismatch repair genes are important in correcting *galk_{tyr145am}* but not *cat-sacB*

The effect of recombination by the components of the MMR system was tested in an HME6 background (see Table 3, above). These strains were used to test the ability of the lagging strand Oligo 100 and leading strand Oligo 101 to correct the amber mutation. For both oligos, elimination of the *mutH*, *mutL*, *mutS*, or *uvrD* genes also increases recombination efficiency by ~100-fold (Table 6).

Table 6

Effect of MMR genes on Recombination with point and insertion mutations
Gal⁺ recombinants²

MMR	HME6 (amber)	HME31 (cat-sacB ₋)
wild-type	4.7 x 10 ⁵	1.2 x 10 ⁵
Δ <i>mutH</i>	2.1 x 10 ⁷	5.2 x 10 ⁴
Δ <i>mutL</i>	2.2 x 10 ⁷	1.0 x 10 ⁵
Δ <i>mutS</i>	3.6 x 10 ⁷	1.4 x 10 ⁵
Δ <i>uvrD</i>	2.7 x 10 ⁷	not done
Δ <i>dam</i>	2.7 x 10 ⁶	1.6 x 10 ⁵

²Data is given from single experiments and wild-type data falls within the expected values based on mean and standard deviation from Table 5.

A Dam methylase mutant was also tested that was created (*dam* \leftrightarrow *amp*). In the HME6 *dam* \leftrightarrow *amp* strain only a 10-fold increase in recombination efficiency was observed relative to HME6. In other mismatch repair mutants there is a 100-fold effect. This suggests that there is a redundancy in recognizing the parental strand. Additional mutations in proteins involved in this recognition could also be used to increase recombination frequency. Comparable increases in efficiency were also seen when the complementary Oligo 101 was used. However, the MMR

-47-

deficient strains did not show a difference in recombination frequency when either Oligo 100 or 101 was used to remove the large heterology, *cat-sacB*.

Example 4

5 **Replication bias is independent of oligo sequence and deletion of *mutS* gene product eliminates oligo sequence bias**

The four ss oligos were tested in strains HME6 and HME41, both of which contain the *galK_{tyr145am}*. Because these strains are identical, except for the orientation of the *gal* operon, the recombination efficiency can be examined for each
10 oligo when it acts as either a lagging or leading strand oligo. As can be seen in Table 7, each oligo gives a higher recombination frequency when it acts as the lagging strand oligo irrespective of the presence or absence of the MMR system.

Table 7

15 **Recombination at galK amber – sequence and mutS effect³**

		Gal+ Recombinants			
		Wild-type		Δ mutS	
		Lagging	Leading	Lagging	Leading
20	Oligo				
	100	4.7x10 ⁵	<u>1.1x10⁴</u>	3.6x10 ⁷	<u>1.4x10⁶</u>
	101	<u>1.6x10⁶</u>	4.8x10 ⁴	<u>5.5x10⁷</u>	2.1x10 ⁶
	144	5.2x10 ⁷	<u>1.3x10⁶</u>	5.1x10 ⁷	<u>8.9x10⁵</u>
	145	<u>8.1x10⁴</u>	1.0x10 ⁴	<u>3.0x10⁷</u>	2.1x10 ⁶

25

³Values underlined were obtained in strain HME41, *gal* operon inverted. The two strains HME6 and 41 are identical except that *gal* is precisely inverted in 41. Thus, replication that reaches the *gal* operon passes through *gal* in different relative directions in the two strains. In HME 6 the oligos of 100 and 144 are the same as the Okazaki fragments whereas in HME41 they are the same as the
30 leading strand DNA. This is reversed for the oligos 101 and 145. As both strains were utilized all combinations were tested. In a strain including the *mutS* deletion, the effects of sequence are eliminated.

As shown in Table 7, the variability in lagging strand recombination is broad in
35 MMR proficient strains, ranging from 8x10⁴ to 5x10⁷, depending on which oligo is

-48-

used. Diversity in recombination efficiency is also observed with the leading strand where values range from 1×10^4 to 1×10^6 . Thus it was determined if these variations would persist in a strain deficient for MMR. Therefore all four oligos were tested in both HME6*mutS* $\langle \rangle$ *amp* and HME41*mutS* $\langle \rangle$ *amp*.

5 The recombination efficiency for lagging strand oligos, regardless of the mismatch created, is approximately 4×10^7 in the mismatch deficient strains. The recombination with leading strand oligos is also more uniform, averaging 1.6×10^6 . Beta alone in a strain defective for the *exo* and *gam* genes gave the same number of recombinants in both *mut+* and *mutS* strains when the ssDNA corresponded to the
10 lagging strand, whereas the when the ssDNA corresponded to the leading strand slightly lower recombination levels for Beta without Exo and Gam were detected.

 Thus, a deficiency in mismatch repair (MMR) eliminates sequence diversity effects. However, a deficiency in MMR, does not affect the lagging and leading strand bias. It is demonstrated herein that a reduction in the ability to perform
15 mismatch repair can increase the frequency of homologous recombination by as much as 100-fold.

Example 5

ss-oligo Recombination with the *galK*_{tyr145am} and *galK* $\langle \rangle$ *cat-sacB* mutants:

20 **RecA function affects recombination with a large heterology**

 An ss oligo is capable of correcting either the *galK*_{tyr145am} or *galK* $\langle \rangle$ *cat-sacB*. When the lagging strand oligo (Oligo 100) was used to correct the *galK*_{tyr145am} of HME6, there were 3.5×10^5 recombinants per 10^8 cells surviving the electroporation. This same oligo was also capable of removing a 3.3kb *cat-sacB*
25 heterology, from HME31 with nearly the same rate – 8.6×10^4 recombinants per 10^8 survivors. The ability of a leading strand oligo (Oligo101) to correct the same lesions was tested. The efficiency of recombination was reduced 10-fold for both HME6 and HME31.

 Previous work had shown ssDNA recombination was independent of RecA
30 function when the *galK*_{tyr145am} point mutation was corrected with an ss oligo (Ellis, *supra*, 2001). A *recA* mutation was moved into HME31 to determine if there was an effect of RecA function on recombination with of a large heterology. A slight

decrease of Gal⁺ colonies was found with lagging strand Oligo 100, but a 7-fold decrease in Gal⁺ recombinants was found when Oligo 101 (leading strand) is used.

No significant influence of transcription on strand bias was detected with the chromosomal gal operon, under conditions known to have a large effect on
5 transcription levels. Without being bound by theory, it is believed that the mismatch repair system recognizes and repairs mismatches at the replication fork. Thus, as demonstrated herein MMR limits SSO-mediated recombination. Thus, in order to increase recombination frequency using recombineering technology, strains can be utilized that include one or more mutations in the MMR system, such as, but
10 not to mutS, mutH, uvrD, dam, and/or mutL genes. These strains are disclosed herein, as are methods to maximize recombination frequency using SSO or double stranded DNA in these strains.

It will be apparent that the precise details of the methods or compositions
15 described may be varied or modified without departing from the spirit of the described invention. We claim all such modifications and variations that fall within the scope and spirit of the claims below.

-50-

CLAIMS

1. A method for inducing homologous recombination in a host cell comprising a target nucleic acid, the method comprising:
 - 5 introducing a single-stranded nucleic acid molecule or double-stranded nucleic acid molecule into the host cell,
 - wherein the single-stranded nucleic acid molecule or double-stranded nucleic acid comprises a sufficient number of nucleotides homologous to the target nucleic acid to enable homologous
10 recombination with the target nucleic acid, and wherein the host cell comprises a de-repressible promoter operably linked to a nucleic acid encoding a single-stranded binding protein and wherein the host cell is deficient for mismatch repair; and
inducing expression of the single-stranded binding protein in the host cell;
 - 15 thereby inducing homologous recombination of the single-stranded nucleic acid molecule or the double stranded nucleic acid molecule with the target nucleic acid in the cell.
2. The method of claim 1, wherein the single-stranded nucleic acid molecule
20 or double-stranded nucleic acid molecule comprises at least about 25 nucleotides homologous to the target nucleic acid.
3. The method of claim 1, wherein the single-stranded DNA binding polypeptide is selected from the group consisting of lambda Beta, RecT, P22 Erf,
25 and Rad52.
4. The method of claim 3, wherein the single-stranded DNA binding polypeptide is Beta.
- 30 5. The method of claim 1, wherein the de-repressible promote is the PL promoter.

6. The method of claim 1, wherein the host cell is a bacterial host cell.
7. The method of claim 1, wherein the host cell is a eukaryotic host cell.
- 5 8. The method of claim 6, wherein the bacterial host cell comprises a mutation in a nucleic acid encoding MutH, dam, MutS, uvrD or MutL, rendering the bacterial host cell deficient for mismatch repair.
- 10 9. The method of claim 8, wherein the bacterial cell is deficient for the production of functional MutH polypeptide as compared to a wild-type cell.
10. The method of claim 8, wherein the bacterial host cell is deficient for the production of functional MutS polypeptide as compared to a wild-type cell.
- 15 11. The method of claim 8, wherein the bacterial host cell is deficient for the production of functional MutL polypeptide as compared to a wild-type cell.
12. The method of claim 8, wherein the bacterial host cell is deficient for the production of functional uvrD polypeptide as compared to a wild-type cell.
- 20 13. The method of claim 1, wherein single-stranded nucleic acid molecule or the double-stranded nucleic acid molecule comprises about 25 nucleotides to about 1,000 nucleotides homologous to the target nucleic acid sequence.
- 25 14. The method of claim 1, wherein a bacterial artificial chromosome comprises the target nucleic acid.
15. A method for inducing homologous recombination between a single-
30 stranded nucleic acid molecule of at least 25 nucleotides in length and a target nucleic acid in a host cell, the method comprising

-52-

inducing the expression of a single-stranded binding protein in a host cell wherein the host cell comprises a de-repressible promoter operably linked to a nucleic acid encoding a single-stranded binding protein and wherein the host cell is deficient for mismatch repair,

5 thereby inducing homologous recombination of the single-stranded nucleic acid molecule with the target nucleic acid in the cell, wherein the nucleic acid sequence of the single-stranded nucleic acid molecule differs from a homologous section of the target nucleic by about one nucleotide to about five nucleotides.

10 16. The method of claim 15, wherein the host cells comprises an extrachromosomal element comprising the target nucleic acid sequence.

15 17. The method of claim 15, wherein the nucleic acid sequence of the single-stranded nucleic acid molecule differs from a homologous section of the target nucleic by one nucleotide.

18. The method of claim 15, wherein the nucleic acid sequence of the single-stranded nucleic acid molecule differs from a homologous section of the target nucleic by about one to about three nucleotides.

20

19. The method of claim 15, wherein the host cell is a bacterial cell.

20. The method of claim 19, wherein the bacterial cell is an *E. coli* cell.

25 21. The method of claim 15, wherein the single-stranded binding protein is Beta.

22. The method of claim 15, wherein the bacterial host cell comprises a mutation in a nucleic acid encoding MutH, MutS, dam, uvrD or MutL, rendering the
30 bacterial host cell deficient for mismatch repair.

23. The method of claim 15, wherein the bacterial host cell is deficient for the production of functional MutL polypeptide as compared to a wild-type cell.

24. The method of claim 15, wherein the bacterial host cell is deficient for
5 the production of functional MutS polypeptide as compared to a wild-type cell.

25. The method of claim 15, wherein the bacterial host cell is deficient for the production of functional MutH polypeptide as compared to a wild-type cell.

10 26. The method of claim 15, wherein the bacterial host cell is deficient for the production of functional uvrD polypeptide as compared to a wild-type cell.

27. The method of claim 15, wherein the bacterial cell is a DY380 cell, or an HME6 cell, or DY330 cell comprising a mutation in a nucleic acid encoding one or
15 more of MutS, MutH, uvrD or MutL.

28. The method of claim 15, wherein the de-repressible promoter is PL.

29. A method for inserting a nucleic acid molecule into a target nucleic acid
20 in a bacterial host cell, the method comprising

introducing into the bacterial host cell a single-stranded nucleic acid of at least 25 nucleotides in length sufficiently homologous for recombination to occur with the target nucleic acid, but not identical to the target nucleic acid, wherein the bacterial host cell comprises a nucleic acid sequence encoding Beta operably linked
25 to a de-repressible promoter, and wherein the bacterial cell has a mutation in a nucleic acid encoding MutS, MutH, uvrD, dam or MutL such that mismatch repair is impaired in the bacterial host cell; and

inducing expression of Beta from the de-repressible promoter, thereby inducing homologous recombination between the single-stranded DNA and the
30 target nucleic acid, and thereby introducing the nucleic acid molecule into the target nucleic acid in the bacterial cell.

30. The method of claim 29, wherein the single-stranded nucleic acid is homologous to a lagging strand of the target nucleic acid.

5 31. The method of claim 29, wherein the single-stranded nucleic acid is homologous to a leading strand of the target nucleic acid.

32. The method of claim 29, wherein the target nucleic acid is a bacterial artificial chromosome (BAC), P1 artificial chromosome, or yeast artificial
10 chromosome.

33. The method of claim 29, wherein the bacterial host cell comprises a nucleic acid encoding Exo or Gam.

15 34. The method of claim 29, wherein the bacterial host cell does not comprise a nucleic acid encoding Exo or Gam.

35. The method of claim 29, wherein the bacterial host cell comprises an extrachromosomal element comprising the target nucleic acid sequence.

20

36. The method of claim 29, wherein a chromosome of the cell comprises the target nucleic acid sequence.

25 37. A method of introducing a mutation into a target nucleic acid in a bacterial host cell, the method comprising

introducing a first single-stranded nucleic acid of at least 30 nucleotides in length homologous to the target nucleic acid into the bacterial host cell; wherein the single-stranded nucleic acid is not identical to the target nucleic acid, and wherein the cell comprises a de-repressible promoter operably linked to a nucleic acid

-55-

sequence encoding Beta, and wherein the bacterial host cell does not express functional *uvrD*, *dam* *MutS*, *MutH*, *MutL* polypeptide, or a combination thereof;

inducing the expression of Beta from the de-repressible promoter,
thereby inducing recombination of the single-stranded nucleic acid with the
5 target nucleic acid in the cell, wherein the recombination of the single-stranded
nucleic acid with the target nucleic acid introduces the mutation into the target
nucleic acid.

38. The method of claim 37, wherein the single-stranded nucleic acid is
10 homologous to a lagging strand of the target nucleic acid.

39. The method of claim 37, wherein the first single-stranded nucleic acid is
homologous to a leading strand of the target nucleic acid.

15 40. The method of claim 37, wherein the mutation is a deletion.

41. The method of claim 37, wherein the mutation is a point mutation.

42. The method of claim 37, wherein the mutation is an insertion.
20

43. An isolated bacterial cell comprising a defective lambda prophage of
genotype $\lambda cI857 \Delta(cro-bioA)$, wherein the bacterial cell further comprises a
mutation in at least one gene encoding a mismatch repair polypeptide.

25 44. The isolated bacterial cell of claim 43, wherein the mismatch repair
polypeptide comprises a *MutS* polypeptide.

-56-

45. The isolated bacterial cell of claim 43, wherein the mismatch repair polypeptide comprises a MutH polypeptide.
46. The isolated bacterial cell of claim 43, wherein the mismatch repair
5 polypeptide comprises a MutL polypeptide.
47. The isolated bacterial cell of claim 43, further comprising a nucleic acid sequence encoding a selectable marker adjacent to the 3' end of *cl857*.
- 10 48. The isolated bacterial cell of claim 43, wherein the nucleic acid sequence encoding the selectable marker is operably linked to a de-repressible promoter.
- 15 49. The isolated bacterial cell of claim 43, wherein the de-repressible promoter is PL.
50. The isolated bacterial cell of claim 43, wherein the bacterial is an *E. coli*.
- 20 51. The isolated bacterial cell of claim 43, wherein the bacterial cell is DY330, HME6, or DY380.
52. A bacterial cell deposited as ATCC No. (HME61) ATCC No. (HME62) ATCC No. (HME63), or ATCC No. (HME64).
- 25 53. A bacterial cell, wherein the bacterial cell comprises a mini-lambda, and wherein the bacterial cell is deficient for mismatch repair.
55. The isolated bacterial cell of claim 53, wherein the bacterial cell
30 comprises a mutation in a nucleic acid encoding MutS.

-57-

56. The isolated bacterial cell of claim 53, wherein the bacterial cell comprises a mutation in a nucleic acid encoding MutH.

5 57. The isolated bacterial cell of claim 53, wherein the bacterial cell comprises a mutation in a nucleic acid encoding MutL.

58. The isolated bacterial cell of claim 53, wherein the bacterial cell comprises a mutation in a nucleic acid encoding uvrD.

10

59. The isolated bacterial cell of claim 53, wherein the bacterial cell comprises a mutation in a nucleic acid encoding dam.

Repair of Mismatched Base

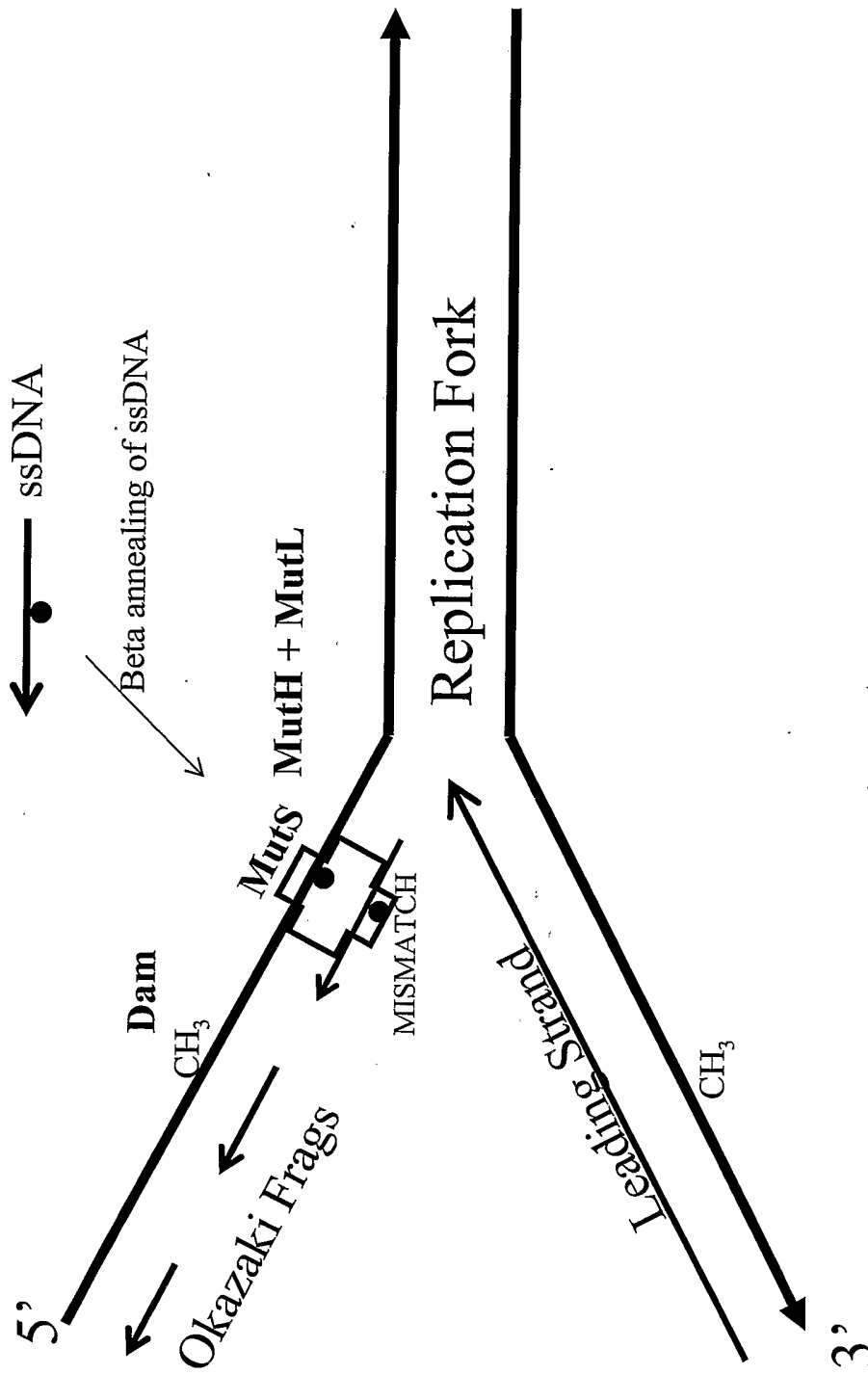


FIG. 1

Anneal ssDNA of Intermediate to Lagging Strand

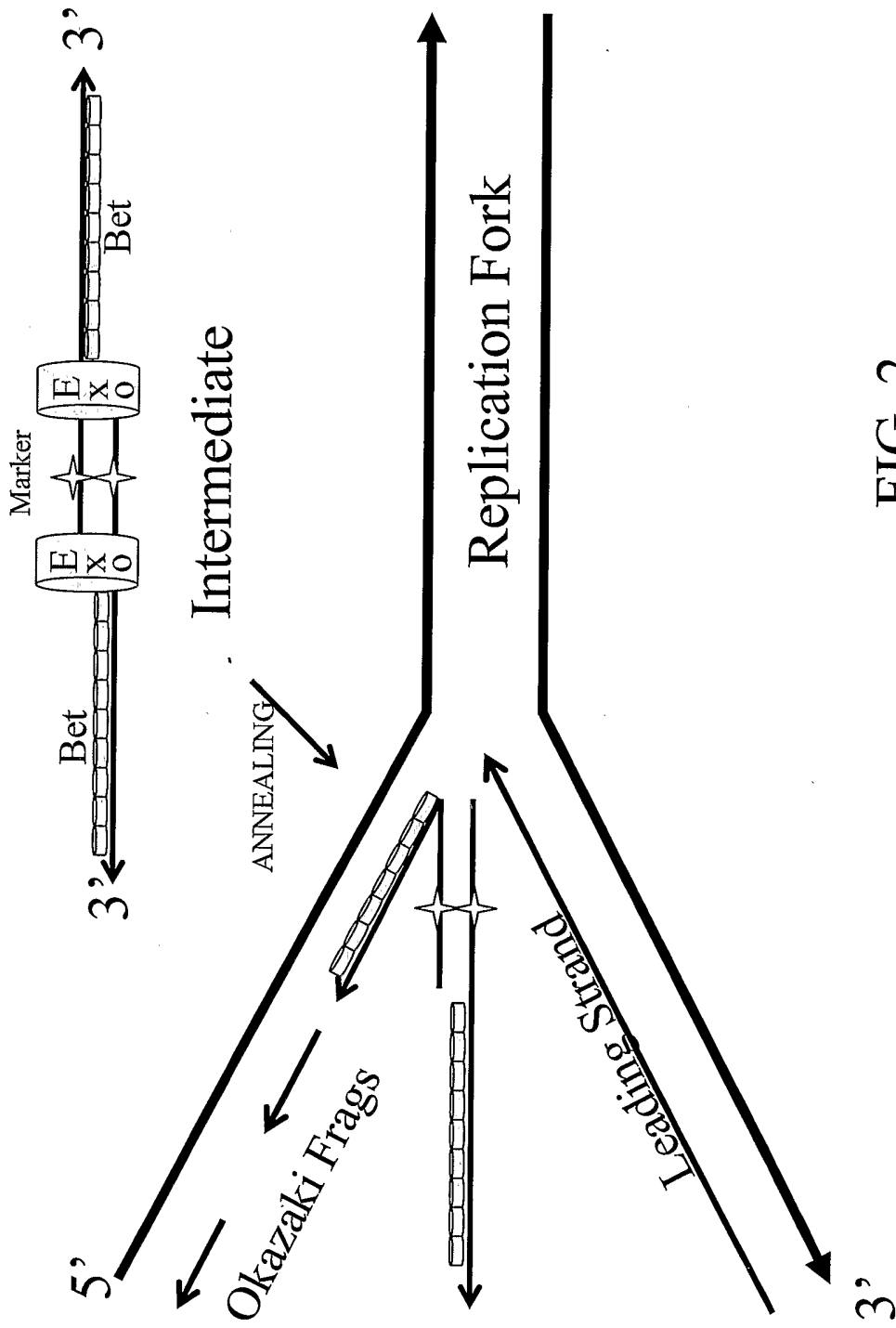


FIG. 2

Branch Migration to Regenerate Replication Fork

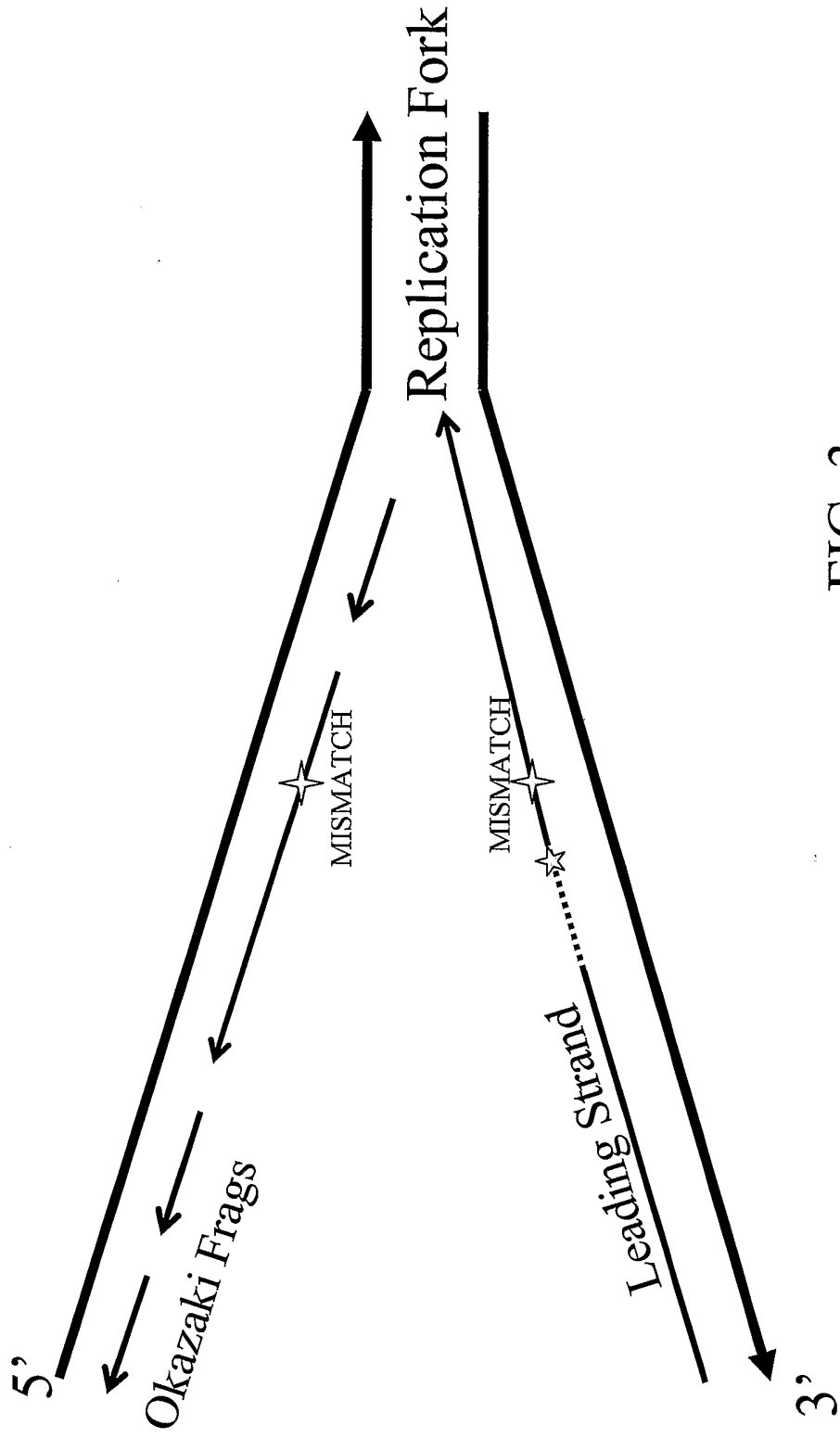


FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/14657

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(7) : C12N 15/00, 1/21, 5/10, 1/20, 1/19
 US CL : 435/455, 465, 471, 490, 252.1, 252.3, 252.33, 254.2
 According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
 Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 435/455, 465, 471, 490, 252.1, 252.3, 252.33, 254.2

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 02/14495 (COURT et al) 21 February 2002 (21.02.2002), see entire document.	1-59
Y	ZAHRT et al. Proceedings of the National Academy of Sciences, USA, September 1997, Vol. 94, pages 9786-9791, see entire document, especially abstract.	1-6, 8-59
Y	US 2002/0013956 (BORTS et al) 31 January 2002 (31.01.2002), see entire document, especially abstract and claims.	7
Y	US 2002/0151059 (TE RIELE) et al) 17 October 2002 (17.10.2002), see entire document, especially abstract and claims.	7

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search: 11 July 2003 (11.07.2003)
 Date of mailing of the international search report: 05 MAR 2004

Name and mailing address of the ISA/US: Mail Stop PCT, Attn: ISA/US, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450
 Facsimile No.:
 Authorized officer: Nancy T. Vogel
 Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

PCT/US03/14657

Continuation of B. FIELDS SEARCHED Item 3:

EAST, STN, CAPLUS, BIOSIS, MEDLINE, EMBASE

search terms: homologous recombination, mismatch repair, beta, mut? uvr? dam, RecT, P22 Erf, Rad52

Box No. VIII (iv) DECLARATION: INVENTORSHIP (only for the purposes of the designation of the United States of America)

The declaration must conform to the following standardized wording provided for in Section 21-4; see Notes to Boxes Nos. VIII, VIII (i) to (v) (in general) and the specific Notes to Box VIII (iv). If this Box is not used, this sheet should not be included in the request.

**Declaration of inventorship (Rules 4.17(iv) and 51bis.1(a)(iv))
for the purposes of the designation of the United States of America**

I hereby declare that I believe I am the original, first and sole (if only one inventor is listed below) or joint (if more than one inventor is listed below) inventor of the subject matter which is claimed and for which a patent is sought.

This declaration is directed to the international application of which it forms a part (if filing declaration with application).

This declaration is directed to international application No. PCT/US03/14657 (if furnishing declaration pursuant to Rule 26ter).

I hereby declare that my residence, mailing address, and citizenship are as stated next to my name.

I hereby state that I have reviewed and understand the contents of the above-identified international application, including the claims of said application. I have identified in the request of said application, in compliance with PCT Rule 4.10, any claim to foreign priority, and I have identified below, under the heading "Prior Applications," by application number, country or Member of the World Trade Organization, day, month and year of filing, any application for a patent or inventor's certificate filed in a country other than the United States of America, including any PCT international applications designating at least one country other than the United States of America, having a filing date before that of the application on which foreign priority is claimed.

Prior Applications:

I hereby acknowledge the duty to disclose information that is known by me to be material to patentability as defined by 37 C.F.R. § 1.56, including for continuation-in-part applications, material information which became available between the filing date of the prior application and the PCT international filing date of the continuation-in-part application.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name: COURT, Donald L.

Residence: Frederick, Maryland
(city and either US state, if applicable, or country)

Mailing Address: 502 Magnolia Avenue, Frederick, Maryland 21702, United States of America

Citizenship: US

Inventor's Signature:.....
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date:.....
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)

Name: COSTANTINO, Nina

Residence: Frederick, Maryland
(city and either US state, if applicable, or country)

Mailing Address: 7913 Runnymede Drive, Frederick, Maryland 21702, United States of America

Citizenship: US

Inventor's Signature:.....
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date:.....
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)

This declaration is continued on the following sheet, "Continuation of Box No. VIII (iv)".

Continuation of Box No. VIII (i) to (v) DECLARATION

If the space is insufficient in any of Boxes No. VIII (i) to (v) to furnish all the information, including in the case where more than two inventors are to be named in Box No. VIII (iv), in such case, write "Continuation of Box No. VIII..." (indicate the item number of the Box) and furnish the information in the same manner as required for the purposes of the Box in which the space was insufficient. If additional space is needed in respect of two or more declarations, a separate continuation box must be used for each such declaration. If this Box is not used, this sheet should not be included in the request.

Continuation of Box No. VIII (iv) DECLARATION: INVENTORSHIP

Name: Li, Xin-tian

Residence: Beijing, China
(city and either US state, if applicable, or country)

Mailing Address: c/o National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) and Peking Union Medical College (PUMC), Beijing 100005, P.R. China

Citizenship: China

Inventor's Signature:
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date:
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)

Name: HUANG, Jian-Dong

Residence: Hong Kong SAR, China
(city and either US state, if applicable, or country)

Mailing Address: c/o Department of Biochemistry, The University of Hong Kong, 3/F Laboratory Block, Faculty of Medicine Building, 21 Sassoon Road, Hong Kong SAR, China

Citizenship: US

Inventor's Signature:
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date:
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)

Name: LIU, Depei

Residence: Beijing, China
(city and either US state, if applicable, or country)

Mailing Address: c/o National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing 100005 P. R. CHINA

Citizenship: China

Inventor's Signature:
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date: May 17, 2004
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)

Name:

Residence
(city and either US state, if applicable, or country)

Mailing Address:

Citizenship:

Inventor's Signature:
(if not contained in the request, or if declaration is corrected or added under Rule 26ter after the filing of the international application. The signature must be that of the inventor, not that of the agent)

Date:
...
(of signature which is not contained in the request, or of the declaration that is corrected or added under Rule 26ter after the filing of the international application)