

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 December 2010 (09.12.2010)

PCT

(10) International Publication Number
WO 2010/140066 A2

- (51) International Patent Classification: Not classified
- (21) International Application Number: PCT/IB2010/002396
- (22) International Filing Date: 4 June 2010 (04.06.2010)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 0909660.3 4 June 2009 (04.06.2009) GB
- (71) Applicant (for all designated States except US): **GENE BRIDGES GMBH** [DE/DE]; BioInnovationsZentrum Dresden, Tatzberg 47-51, 01307 Dresden (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **STEWART, Adrian, Francis** [AU/DE]; Anton-Graff-Str 24, 01309 Dresden (DE). **IACONE, Roberto** [IT/CH]; c/o Gene Bridges GmbH, BioInnovationsZentrum Dresden, Tatzberg 47-51, 01307 Dresden (DE). **MARESCA, Marcello** [IT/US]; c/o Gene Bridges GmbH, BioInnovationsZentrum Dresden, Tatzberg 47-51, 01307 Dresden (DE). **ANASTASIADIS, Konstantinos** [GR/DE]; c/o Gene Bridges GmbH, BioInnovationsZentrum Dresden, Tatzberg 47-51, 01307 Dresden (DE).
- (74) Agents: **GOODFELLOW, Hugh, Robin** et al.; Carpmaels & Ransford, One Southampton Row, London WC1B 5HA (GB).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report (Rule 48.2(g))



WO 2010/140066 A2

(54) Title: METHOD OF ALTERING NUCLEIC ACIDS

(57) Abstract: The invention provides a method for altering the sequence of a target nucleic acid molecule to incorporate a replacement nucleic acid sequence. The method comprising the steps of: a) introducing a nucleic acid fragment into the target nucleic acid molecule by homologous recombination, wherein the nucleic acid fragment comprises: i) a first region of homology to the target nucleic acid molecule; ii) a 5' splice site; iii) a selectable marker; iv) a 3' splice site; v) a second region of homology to the target nucleic acid molecule; wherein components i) to v) are ordered from 5' to 3'; and the replacement nucleic acid sequence is positioned within or between the first and second regions of homology but external of the splice sites; b) selecting for the introduction of the nucleic acid fragment using the selectable marker; c) incubating the product of step b) under conditions suitable for a splicing reaction to occur such that the selectable marker is excised and from the target nucleic acid, and thereby generating the desired altered target nucleic acid sequence.

Method of altering nucleic acids

This application is related to a method for alteration of a nucleic acid molecule based on linked steps of recombination and intron splicing, particularly by self-splicing ribozyme excision.

All publications, patents and patent applications cited herein are incorporated in full by reference.

Background

The engineering of nucleic acid molecules, particularly DNA molecules, is of fundamental importance to Life Science research. For example, the construction and precise manipulation of nucleic acid molecules is required in many studies and applications in the research fields of, for example, functional genomics (for review, see Vukmirovic and Tilghman, *Nature* 405 (2000), 820-822), structural genomics (for review, see Skolnick et al., *Nature Biotech* 18 (2000), 283-287) and proteomics (for review, see Banks et al., *Lancet* 356 (2000), 1749-1756; Pandey and Mann, *Nature* 405 (2000), 837-846).

A number of methods are currently available for engineering nucleic acid molecules, particularly DNA molecules. Conventional methods, which are still the most widely used, rely wholly on restriction digestion, followed by ligation (see Sambrook J and Russell D.W. *Molecular Cloning, a laboratory manual*, 3rd ed. (2000) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Progress in our understanding of the various mechanisms of nucleic acid recombination has allowed conventional cloning techniques to be complemented and partially replaced by more advanced strategies utilising homologous recombination (in prokaryotes, see below; in eukaryotes, see, for example, Bode et al., *Biol Chem* 381 (2000), 801-813; Joyner, *Gene Targeting, a practical approach*, (2000) second edition, Oxford University Press Inc. New York), PCR-directed mutagenesis (see Ling and Robinson, *Anal. Biochem.* 254 (1997), 157-178), site-specific recombination (for example, Hauser et al., *Cells Tissues Organs* 167 (2000), 75-80) and transposon mutagenesis (see, for example, Martienssen, *Proc. Natl. Acad. Sci. USA* 95 (1998), 2021-2026; Parinov and Sundaresan, *Curr Opin Biotechnol* 11 (2000), 157-161).

One particular area of mutagenesis which is restricted by current technologies is site directed mutagenesis (SDM). SDM is a fundamental methodology in molecular biology whereby a single nucleotide, or a few nucleotides at one site, are altered. This

methodology is especially relevant for the analysis of protein function because alteration of a codon in a protein coding gene is the way to change a single amino acid in the context of the rest of the protein. SDM is also powerful for other types of functional enquiry via mutagenesis, including discrete analyses of *cis* elements in DNA or RNA molecules. SDM requires a high degree of precision. Ideally only one to a few nucleotides are altered in the target DNA molecule, which can be as small as an *E.coli* plasmid of 2000 bps, or as large as a higher eukaryotic genome of 10,000 million bps. This application addresses the current technical limitations of SDM. Existing methods, particularly those that are popular in commercial practice, are severely limited in size and practice to *E.coli* plasmids of less than 8000 bps. These methods are limited to small sized plasmids by two different limitations. Firstly random mutagenesis, and secondly cost in terms of time and money. These limitations are described in more detail below.

In two-step PCR, which encompasses the frequently applied 'overlap PCR' technique, the final product is assembled through 2 rounds of PCR. In the first round one or more PCR fragments are produced which are then assembled, or extended, using a differing combination of primers in a second round of PCR. The product of the second round is then cloned into a vector using standard molecular biology techniques. Whilst this method is simple, and can be easily performed with commonplace techniques, it is disadvantaged by the increased mutational frequencies inherent with multiple PCR rounds.

A second set of methods are based on amplification of the whole target plasmid. The first of these applies 'inverse PCR'. Here the whole plasmid is amplified as a single linear fragment, wherein the desired mutations are introduced through the use of mutagenic primers. Following amplification, the product may be digested with the DpnI enzyme to cleave methylated and hemimethylated DNA, i.e. the unmutated template DNA. If the DpnI digestion step is not complete, a proportion of the plasmids recovered are likely to be unmutated parent. However, completion of digestion cannot be confirmed easily. In this situation a gel-extraction step may be performed, although this step is likely both to extend and complicate the process, and additionally contribute to substantial loss of the product. The linear fragment is then ligated back to itself to reform a plasmid, before transformation into a host cell. To increase the efficiency of the re-ligation, the fragment is treated with a kinase, or cleaved by a restriction enzyme that recognises a sequence at the termini to generate termini with phosphates attached.

A further method that amplifies the plasmid backbone is the “Quick Change” PCR mutagenesis method (Stratagene). This kit is the most popular method for SDM. In this method two complementary primers, which also contain the desired mutations, are used to prime extension around the plasmid, from the point of annealing. In this method, all amplification is from the parent molecule, which significantly reduces unwanted mutations. The product is then treated with DpnI, as detailed above, and in doing so suffers from the same disadvantages detailed above. As this method produces double-stranded PCR products with complementary single-stranded regions at their termini no restriction enzyme cleavage/kinase/ligation steps are required for re-circularisation prior to transformation into the host cell.

Although PCR-based *in vitro* strategies which amplify around the plasmid backbone allow precise site-directed mutagenesis to be effected, such methods suffer from the introduction of unwanted artefactual secondary mutations in the targeted molecule during amplification of the mutated nucleic acid product. In order to verify that these molecules are free of erroneous mutations, the entire molecule must be sequenced. This imposes a significant bottleneck on the mutagenesis process in respect of time and/or costs. The costs of screening may not be prohibitive for the mutation of one base to another specified base, but with increasing number of permutations in a single mutagenic reaction, for example when attempting saturation mutagenesis at a number of positions, the costs may become quite significant. Furthermore, methods amplifying the whole target plasmid with high efficiency and fidelity are presently limited to molecules of a maximal size of around 10-15 kilobasepairs, and typically around 8 kilobasepairs.

A further limitation of this method stems from the fact that the PCR reactions in which the mutant sequences are generated are primed by synthetic oligonucleotides. Should it be desired to introduce multiple mutations, the maximum distance between the mutations which can be achieved in a single “Quick Change” reaction is limited by current oligonucleotide synthesis technologies, that is a distance of 100-200 oligonucleotides. Furthermore, longer oligonucleotide syntheses are more prone to produce unwanted mutant products. This adds weight to the reality that the products must be sequenced to ensure that the desired, and only the desired, mutations have been incorporated into the target.

Other techniques for mutagenesis do not allow flexible DNA engineering at any chosen position, but instead require specific sequence elements (site-specific recombination target sites or restriction enzyme recognition sites based methods) or are based on random

targeting (transposon based methods). The application of homologous recombination in DNA engineering has been pioneered in *S. cerevisiae* (for review see Shashikant et al., Gene 223 (1998), 9-20). However, since several inherent complications limit the usefulness of yeast as a DNA engineering host, homologous recombination based DNA engineering has recently been established in the premier cloning host *E. coli*. (for review see Muyrers et al., Trends in Bioch Sci, (2001) 26(5): 325-331). The most intensively studied pathway is the RecA-dependent recombination pathway, which is responsible for the majority of recombinogenic processes in the bacterial cell. A second recombination pathway is the RecF-pathway. In the third pathway, recombination requires the expression of both components of the RecE/RecT protein pair, or of its functional homologues derived from the lambda phage, Red α /Red β . In the last few years, a technology termed ET, Red/ET or recombineering has been developed that uses the phage RecE/RecT or Red α /Red β protein pairs for precise DNA engineering (Zhang et al., Nature Genetics 20 (1998), 123-128; Muyrers et al., Nucl Acids Res 27 (1999), 1555-1557; International patent application WO99/29837; for review see Muyrers et al., Trends Bioch Sci 2001) 26(5): 325-331). This system is immensely powerful and may be used to introduce substitutions, deletions and insertions into nucleic acid molecules, as desired.

However, another disadvantage of the methods described above is that there is no simple, generally applicable method for identifying products which contain the mutation from those that do not. In the absence of any easily selectable marker or physical selection systems, PCR-based methods can be used to detect the mutation. Approaches based on these methods do not employ a selectable gene but simply rely on finding the intended mutation amongst a usually large to very large background of unaltered substrate. Such approaches are labour intensive and not efficient.

Recombinogenic mutagenesis (recombineering) has been used to incorporate selectable markers into the target DNA in conjunction with the desired mutation. The ability to select by a physical characteristic, for example resistance to an antibiotic, greatly simplifies the process of screening. To solve this problem, methods have been developed to incorporate sites for further site specific recombinases (e.g Bloor & Cranenburgh, Appl. Env. Micro. (2006) 2520-2525). By designing these sites in the correct orientation relative to one another, the selectable marker can be removed from the construct by the action of the further recombinase. In this case, however, artefacts comprising one copy of the site-specific recombinase sites are left behind in the targeted construct, which can significantly affect the structure and function of the DNA into which they are inserted.

This is particularly disadvantageous in methods of site directed mutagenesis, where often only a single nucleotide change is desired.

Attempts have been made to perform a second recombineering step to remove the marker. These attempts have only been met with success in single copy systems, for example BACs (Warming et al., Nucl. Acid Res. (2005) e36). When applying this approach in multicopy systems, it was observed that intermolecular recombination was observed, as opposed to the desired intramolecular reaction. This resulted in a mix of desired product, the intermediate product produced in the first recombineering step, higher multimers of these plasmid and also hybrid molecules, such as a multimer containing the sequence of both the desired product and the intermediate product produced in the first recombineering step (Thomason et al., Plasmid (2007) 58: 148-158).

Hence, there is an important need to develop better methods for SDM that are not fundamentally limited to minor tasks, in other words, methods that are not constrained by the maximum size of either the target, or the mutagenic oligonucleotide, or by excessive costs of time and money spent in screening for the desired mutants. A further reason for developing better methods lies with high-throughput (HT) methodology. The current methods, particularly the screening steps, which in some cases may require the sequencing of the whole plasmid backbone, are labour intensive, expensive and not easily adaptable to HT processing and automation. Given the vast challenges presented by biological diversity and the clear need to magnify the scale of analysis using HT methodologies, SDM methods that permit HT scale-up will create new applications. That is not to say that such methods are limited to performing solely SDM. Any method of mutagenesis that is amenable to HT applications will be of great use in the field.

It is the object of the current invention to provide a method for the swift and simple alteration of target nucleic acid molecules, which by virtue of these properties is amenable to high-throughput applications.

Summary of the invention

According to the invention, there is provided a method for altering the sequence of a target nucleic acid molecule to incorporate a replacement nucleic acid sequence, said method comprising the steps of:

a) introducing a nucleic acid fragment into the target nucleic acid molecule by homologous recombination, wherein the nucleic acid fragment comprises:

i) a first region of homology to the target nucleic acid molecule;

- ii) a 5' splice site;
- iii) a selectable marker;
- iv) a 3' splice site;
- v) a second region of homology to the target nucleic acid molecule;

wherein components i) to v) are ordered from 5' to 3'; and

the replacement nucleic acid sequence is positioned within or between the first and second regions of homology but external of the splice sites;

b) selecting for the introduction of the nucleic acid fragment using the selectable marker;

c) incubating the product of step b) under conditions suitable for a splicing reaction to occur such that the selectable marker is excised from the target nucleic acid, and thereby generating the desired altered target nucleic acid sequence.

This method is herein termed intron recombineering (IRN). Methods according to the invention are swift, simple, efficient and amenable to high-throughput methodologies. The methods exploit recombineering to introduce splice sites into a transcriptionally-active nucleic acid fragment along with a replacement nucleic acid sequence. Subsequent splicing generates seamlessly mutagenised products of the desired sequence. The method introduces the required mutation in a single step and takes advantage of positive selection. The spliced RNA can be translated to a protein that contains no scars except the chosen mutation. Due to the high fidelity of recombineering and the use of selection, the approach can be easily applied in a high-throughput pipeline in liquid culture without any need for a physical step of screening such as PCR screening.

To achieve the inventive breakthrough described here, it was necessary to bypass some of the existing limitations of homologous recombination, as well as to combine homologous recombination using a selectable gene with the employment of a step using splicing in a very specific way.

Summary of the method

The method can be illustrated by reference to Figure 2 herein, which depicts a method according to the invention being used for site-directed mutagenesis. This method is based on the selection of a sequence in the target nucleic acid molecule (for example a chromosome, episome or plasmid) that will be subject to mutagenesis. It will be understood by the skilled addressee that this sequence may be of any length, and designed

such that it is appropriate to the strategy employed. Further, whilst this Figure illustrates the use of this method in mutating a protein coding region through the use of a self-splicing ribozyme in mRNA, it is equally applicable to mutagenesis using introns that are processed by endogenous cellular machineries, or tRNA and rRNA genes that are processed through spliceosome mediated or tRNA intron splicing.

The Figure should be read in anti-clockwise direction starting from the upper right part.

In this example, the target nucleic acid has been prepared as a PCR product. In this PCR reaction, first and second PCR primers have been designed which contain 5' and 3' regions of homology to the specific region on the target nucleic acid where it is desired to alter the sequence, which in this case involves the insertion of a sequence mutation. Here, the depicted primers contain the mutation to be introduced into the target nucleic acid, although this replacement sequence may lie anywhere external to the acceptor and donor splice sites, within or between the first and second regions of homology. By "within" is meant that the replacement sequence is placed such that it is flanked and is abutted on either side by stretches of the homology region. By "between" is meant that the replacement sequence abuts either, or both, of the first homology region and the 5' splice site or the 3' splice site and the second homology region. In Figure 2 this mutation is denoted by an asterisk, and may be followed through the mutagenesis protocol using this annotation. The two regions of homology flank an intron template which comprises splice acceptor and donor sites into which a selectable marker has been inserted. In the example given, a self-splicing ribozyme has been used. Of course, instead of using a PCR product, the introduced nucleic acid fragment may be prepared by digestion of a plasmid in which the desired fragment has been constructed. This embodiment has the further advantage that the fragment may be sequenced, to ensure that there are no erroneous mutations.

In step a) of the method, illustrated in the top left quarter of Figure 2, a nucleic acid fragment containing a gene for selection, usually a gene conferring antibiotic resistance, (here shown as blasticidin) is introduced into the target nucleic acid by homologous recombination. The nucleic acid fragment includes, in order from 5' to 3', a first region of homology to the target nucleic acid molecule, also termed a "homology arm"; a 5' splice site (here, the first portion of a self-splicing ribozyme); a selectable marker; a 3' splice site (here, the second portion of the self-splicing ribozyme); and a second region of homology to the target nucleic acid molecule.

In addition, the nucleic acid fragment includes a replacement nucleic acid sequence that is positioned within or between the first and second regions of homology. These two

homology arms must flank the replacement sequence in the incorporated nucleic acid molecule to ensure that the replacement sequence is introduced into the target nucleic acid. Additionally, the replacement sequence must flank, *i.e.* lie outside, the splice sites. The replacement sequence may involve a deletion, a substitution or an insertion, and may be of any practical length, ranging from a single nucleotide to kilobases or megabases in theory. Here, a single substitution is illustrated.

As detailed above, the replacement sequence may be placed entirely between the first region of homology and the 5' splice site, or entirely between the 3' splice site and the second region of homology. In an alternative, the replacement sequence may be placed within one or both of the first and/or second regions of homology. Further, a first portion of the replacement sequence may be placed between the first region of homology and the 5' splice site, and a second portion between the 3' splice site and the second region of homology, or indeed in any combination of the positions listed above. The precise positioning of the replacement sequence will be dependent upon the desired strategy, and will be evident to the skilled artisan following the teachings provided herein.

The introduced nucleic acid fragment is preferably introduced into a transcriptionally-active nucleic acid. By "transcriptionally-active" is meant that the nucleic acid is used as a template to produce a single-stranded RNA transcript. The template may be a ribonucleic acid or a deoxyribonucleic acid.

The nucleic acid formed by homologous recombination between the target nucleic acid molecule and the introduced nucleic acid fragment represents a stably integrated intermediate in the method. This intermediate is selected for using the introduced selectable marker. A further advantage of this method is that the selectable marker may be permanently integrated into the target nucleic acid. Therefore, in addition to use in selecting for the introduction of the replacement nucleic acid, the selectable marker may also be used to maintain the presence of the introduced nucleic acid in the target nucleic acid over time. For example, should the target nucleic acid be an episome, the marker may be used to ensure that this episome is not lost from the host cell. Therefore, at this stage the mutagenesis is not in any way scarless. In other words, sequence that is not desired in the final product still remains in the target nucleic acid. In contrast to other methods of mutagenesis, the removal of the extraneous sequence is not performed at this stage, but following a splicing reaction that occurs after transcription.

Upon transcription, a single-stranded RNA transcript is formed and a splicing reaction then takes place which involves excision of all sequence between the 5' and 3' splice

sites. The idea is applicable to any type of intron. Here, a self-splicing ribozyme is illustrated. In this transcript, the three-dimensional structure of the ribozyme will form spontaneously. In order for the ribozyme to form in the correct confirmation, complementary sections of the ribozyme must base-pair together to form short double-stranded sections which comprise the framework of the ribozyme. These sections of the ribozyme must be preserved. The selectable marker must therefore be inserted into a loop portion of the ribozyme. In other words, the marker should be inserted into a region of the ribozyme which is a single-stranded portion between the double-stranded portions of the ribozyme, but in a fashion such that it does not interfere with the self-splicing catalytic mechanism. This is shown schematically in the bottom part of Figure 2.

The product thus obtained will be a mutated target nucleic acid. In Figure 1, a single point mutation is illustrated, however the same approach can be applied to mutate any of the base pairs that fall within the replacement sequence created within or between the homology regions. For example, mutations may comprise substitutions, insertions or deletions. The number of base pairs changed may range from a single point mutation, to multiple substitutions, insertions or deletions, for example, 2, 3, 5, 10, 20, 50 or more base pairs.

Following the excision of the region between the two splice sites, here defined by the components of a self-splicing ribozyme, the transcript now contains only the natural sequence and the introduced mutation, and no extraneous sequences from the self-splicing ribozyme, or the selectable marker encoded within it. In Figure 2, following excision of the self-splicing ribozyme, the mRNA is translated into a protein which incorporates the mutation in the replacement sequence of the introduced nucleic acid fragment.

Homologous recombination

In step a) of the method, a nucleic acid fragment is introduced into the target nucleic acid molecule by homologous recombination. Design of appropriate homology arms will be within the ambit of the skilled artisan, imbued with knowledge of the present invention.

A preferred homologous recombination technique for use in these methods employs the Red operon from phage lambda and is commonly termed 'recombineering' (see Zhang et al., *Nature Genet* 20 (1998), 123-128; Muyrers et al., *Nucl Acids Res* 27 (1999), 1555-1557; co-pending International patent application WO99/29837; co-pending European patent application EP1399546; also co-pending International application PCT/IB2009/000488, entitled "Method of nucleic acid recombination"; the contents of all these documents is hereby incorporated by reference). Homologous recombination uses

the Red β protein, or combinations of proteins comprising the Red β protein, for example the Red α , Red β and Red γ proteins, to mediate homologous recombination. As an alternative, homologous recombination uses combinations of proteins related to the Red β protein, for example the RecT, Erf or Plu β proteins.

In order for homologous recombination to take place, the nucleic acid fragment to be introduced must be brought into contact with the target nucleic acid molecule in the presence of a phage annealing protein, or a functional equivalent or fragment thereof. Suitable phage annealing proteins for use in the invention (as known at the time of writing) include RecT (from the rac prophage), Red β (from phage λ), and Erf (from phage P22). The identification of the recT gene was originally reported by Hall et al., (J.Bacteriol. 175 (1993), 277-287). The RecT protein is known to be similar to the λ bacteriophage β protein or Red β (Hall et al. (1993); Muniyappa and Radding, J.Biol.Chem. 261 (1986), 7472-7478; Kmiec and Holloman, J.Biol.Chem.256 (1981), 12636-12639). The Erf protein is described by Poteete and Fenton, (J Mol Biol 163 (1983), 257-275) and references therein. Erf is functionally similar to Red β and RecT (Murphy et al., J Mol Biol 194 (1987), 105-117), and in some cases can substitute for the lambda phage recombination system (Poteete and Fenton, Genetics 134 (1993), 1013-1021). The invention also includes the use of functional equivalents of the molecules that are explicitly identified above as RecT, Red β and Erf, provided that the functional equivalents retain the ability to mediate recombination, as described herein, as described by bioinformatics methods to identify similar proteins based on sequence similarities (for example Iyer, L. M., Koonin, E. V. & Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* 3, 8) and in European patent application EP1399546. Such functional equivalents include homologues of elements of recombination systems that are present in bacteriophages, including but not limited to large DNA phages, T4 phage, T7 phage, small DNA phages, isometric phages, filamentous DNA phages, RNA phages, Mu phage, P1 phage, defective phages and phagelike objects, as well as the functional homologues of elements of recombination systems that are present in viruses (e.g. Datta et al., 2008 PNAS 105: 1626-1631). Of course, as and when additional, functionally equivalent annealing proteins are discovered, for example, as a result of genome sequencing projects of other coliphages and lambdoid phages, it is envisaged that these annealing proteins will be equally suitable to those that are explicitly recited above.

In any context, homologous recombination, even using an efficient system like the Red operon, is a rare event and so according to the methods of the invention, a selection step needs to be included to ensure high efficiency. Usually the selection step should be based on the insertion of an antibiotic or other gene so that the product can be distinguished from substrate. In conventional methods, the selectable gene remains at the site of the mutagenesis and must be removed, as the insertion of such a gene is incompatible with most genetic engineering strategies, for example, SDM, which requires only a very small change at the site of mutagenesis. The desire to ensure seamless removal of the selectable marker is equally valid for other types of mutagenesis, such as where it is desired to maintain the mutated construct as close as possible to its original form incorporating only the necessary mutations, for example the placement of a whole domain in a coding sequence with another similar domain. As described above, previous applications of recombineering to SDM attempted to solve this problem by use of counterselectable genes and additional site specific recombination sites, which were met with numerous difficulties. The methods of the present invention overcome all these problems. Here, even though the selectable marker remains in the target nucleic acid, it is excised spontaneously upon the formation of the self-splicing ribozyme. No further, often complicated, steps of homologous recombination are needed to produce a scarless mutated transcript.

Selectable markers

Any selectable marker may be used, either conferring resistance, sensitivity, causing fluorescence and so on. The selectable marker may be an antibiotic resistance marker. Alternatively, the selectable marker may be an enzyme which complements an auxotrophy. In other words the selectable marker may be an enzyme which produces an essential nutrient. In a host which lacks the ability to produce this essential nutrient, only those cells containing the selectable marker will be able to grow in media which lack the nutrient. Examples of auxotrophic markers are well known in the art. A non-limiting list of examples includes *ura3*, *pyrG*, *niaD* and *trpC*. Incorporating a marker ensues simple screening because targets incorporating the introduced nucleic acids can be identified by phenotypic screening, using selectable growth media, rather than by sequence or indeed size, on a gel.

Use of a fluorescence marker may be particularly applicable for high-throughput methodologies. By using such marker it will be possible to isolate cells containing the desired product by Fluorescence-activated cell sorting (FACS).

Particularly preferred for use in the methods of the invention is a genetic organisation wherein the transcript produced from the transcriptional active target nucleic acid is in the same sense as the transcript produced from the selectable marker. That is to say if transcription to produce the single-stranded RNA encoding, for example, a self-splicing ribozyme uses the Watson strand as template, then transcription to produce the single-stranded RNA encoding the selectable marker should use the Watson strand as template, and *vice versa*. The advantage of this orientation is that complementary transcripts are not produced. Accordingly, a double-stranded RNA molecule will not be formed. Formation of double-stranded RNA molecule should be avoided as such molecules are known to be targeted for degradation. Therefore, expression of both the product encoded by the altered nucleic acid and the selectable marker will both be higher. This, in turn, permits easier screening at higher concentration of selective agent, and higher expression of the altered transcribed nucleic acid.

Other methods of selection will be known to those of skill in the art. For example, a primer region may be included for PCR amplification of a selectable gene.

Homology arms

According to the invention, the introduced nucleic acid fragment should possess at least two regions of sequence homology (homology arms) with regions of sequence on the target nucleic acid molecule. By "homology" is meant that when the sequences of the introduced and target nucleic acid molecules are aligned, there are a number of nucleotide residues that are identical between the sequences at equivalent positions. Degrees of homology can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Such regions of homology are preferably at least 9 nucleotides each, more preferably at least 15 nucleotides each, more preferably at least 20 nucleotides each, even more preferably at least 30 nucleotides each. Particularly efficient recombination events may be effected using longer regions of homology, such as 50 nucleotides or more. Preferably, the degree of homology over these regions is at least 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% or 99% or more identity, as determined using BLAST version 2.1.3 using the default parameters specified by the NCBI (the National Center for

Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) [Blosum 62 matrix; gap open penalty=11 and gap extension penalty=1].

The regions of sequence homology may be located on the introduced nucleic acid fragment so that one region of homology is at one end of the molecule and the other is at the other end. However, one or both of the regions of homology may also be located internally. The two sequence homology regions should thus be tailored to the requirements of each particular experiment. There are no particular limitations relating to the position for the two sequence homology regions located on the target DNA molecule, except that for circular double-stranded DNA molecules, the repair recombination event should not abolish the capacity to replicate. As the skilled reader will appreciate, the sequence homology regions can be interrupted by non-identical sequence regions, provided that sufficient sequence homology is retained to allow the repair recombination reaction to occur.

In some embodiments, as discussed in more detail above, one or both of the homology regions may contain the replacement sequence.

Splicing sites

The introduced nucleic acid molecule should also include 5' and 3' splice sites, also referred to herein as "donor" and "acceptor" splice sites. The 5' and 3' splice sites demark the boundaries of the excised intron sequence.

Splicing is a mechanism whereby fragments of an RNA transcript are excised, and the regions external to the excised fragment are joined together.

A number of splicing mechanisms occur in nature. All mechanisms may be used in the method provided herein. Each different mechanism requires varying structural elements provided by spontaneously formed loops in the transcript, and some further require proteins and RNAs.

The majority of splicing in Eukaryotes is performed in reactions involving a protein/RNA complex called a spliceosome, which removes introns from RNA transcripts. A 5' splice site in such introns typically has the sequence GU, and the 3' splice site has the sequence AG. Introns with these splice sequences (GU-AG introns) are excised by the major spliceosome. In a typical splicing reaction, the 2'-OH of a specific Adenosine residue in the RNA attacks the 5' splice site and cleaves it. In this attack the intron forms a loop structure. The exposed 3'-OH of the 5' splice site, created in the first reaction, now attacks the 3' splice site. In doing so, the portions of the transcript external to the intron

are joined together, with the simultaneous release of the intron. Removal of these introns requires the presence of the spliceosome; it does not occur spontaneously. Introns with sequences which vary from the canonical format detailed above (AU-AC introns) are known to be removed from transcripts by the minor spliceosome. The splice donor (5' splice site) is the site which provides the 3'-OH for attack of the splice acceptor (3' splice site) to generate the processed transcript.

A further form of splicing is tRNA splicing. In some Eukaryotes, introns are removed from tRNA genes by a three step process, each catalysed by introns. Here, the 5' and 3' splice sites are defined by their distance from the recognition sequence of an endonuclease (Abelson et al., 2003 J. Biol Chem 273: 12685-12688). The joining of the fragments generated by the action of the endonuclease is performed by a ligase enzyme.

The preceding splicing mechanisms are characterised by their requirement for host proteins to effect the splicing. Introns exist which are capable of self-splicing. That is to say they do not require any host proteins to do so, though host factors may increase the rate and efficiency of the process. The catalysis of the splicing reaction is performed by structures formed by the intron itself. Group II and group III introns perform self splicing via a lariat structure mechanism, similar to intron excision as catalysed by the spliceosome. A 2'-OH of a defined residue initiates the splicing by attack of the 5' splice site to form the lariat, which is followed by a second reaction which joins the 3-OH of the 5' splice site and the 3' splice site.

A further type of self-splicing intron is a group I self splicing intron. As opposed to group II and group III, there is no absolute requirement for host proteins for the reaction mechanism to occur. The reaction mechanism differs in that no lariat structure is formed. In this mechanism an external guanosine provides the hydroxyl group that initiates the splicing mechanism by attacking the 5' splice site. Then, in common with the other mechanisms, the free 3'-OH of the 5' splice site attacks the 3' splice site, which brings about joining of the 5' and 3' splice sites with simultaneous excision of the intron.

In a preferred embodiment, the donor and acceptor splice sites are components of a self-splicing ribozyme. A self-splicing ribozyme is a ribonucleic acid which is defined by its capability to excise itself from a transcript in which it is located. The presence of this self-splicing ability is a preferable feature of the methods of the invention because this removes the need for additional components (*e.g.* the spliceosome) for a successful splicing reaction to occur. Such ribozymes are observed as introns in transcribed nucleic acids in a wide range of organisms, including prokaryotic and eukaryotic species, and

also in viruses (e.g. Ko et al, 2002, *J. Bact.* 184: 3917-3922; Bonocora and Shub, 2004, *J. Bact.* 186: 8153-8155; Yamada et al., 1994 *Nucleic Acids Research* 22: 2532-2537).

In the transcribed nucleic acid, the functional structure of the ribozyme forms. The ribozyme then proceeds, through a series of reactions, to catalyse the cleavage and rejoining of the transcript such that the ribozyme is excised. Self-splicing ribozymes can in theory be incorporated anywhere within an encoding nucleic acid, but will only be removed from a transcribed region. Ribozymes only fold into their active form in any single-stranded RNA. The methods of this invention are equally applicable to the mutagenesis of messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA). Use of self-splicing ribozymes is particularly advantageous as it permits the insertion of a nucleic acid fragment into a target nucleic acid, which will be excised when the sequence is present in a single-stranded ribonucleic acid, for example where the sequence to be removed is incorporated into the ribozyme.

As described above, there are 3 main groups of self-splicing intron, all of which may be exploited in the methods of the present invention. These ribozymes differ in their use for proteins to aid self-splicing. Particularly preferred for use in methods of the invention are Group I and Group II introns; of these, Group I introns are preferred.

The reaction pathway of a Group I self-splicing intron is shown schematically in Figure 1. Group I introns are characterised by a secondary structure of 9 paired regions formed by the self annealing of the ribonucleic acid sequence to complementary regions in the ribozyme. It is preferred that 1, 2, 3, 4, 5, 6, 7, 8, or all 9 of these regions are maintained in a Group I intron self-splicing ribozyme used in the invention.

Also included as self-splicing ribozymes are Group I-like introns. These ribozymes are similar to Group I introns and possess the ability to self-splice, but have a different core structure. Also included are twintrons (for example Drager and Hallick 1993 *Nucleic Acids Research* 25: 2389-2394) which are nested self splicing introns.

Also included are functional variants of self-splicing ribozymes. Examples of functional variants are self-splicing ribozyme variants that have been optimised and/or evolved. Such functional variants may be evolved such that they are capable of splicing at elevated temperatures, or under extreme physiological conditions, such as pH or salt concentration (Guo & Cech, 2002. *RNA* 8: 647-658). The properties of the self-splicing ribozyme should be appropriate to the experiment being performed, for example, thermostable

ribozymes will be most appropriate when the mutagenesis strategy is performed in a thermophilic host.

Another example of a functional variant preferred for use in a method of the invention is a self-splicing ribozyme in which sequence that is not essential for self-splicing has been removed. Particularly preferred for use in a method of the invention are non-motile self-splicing ribozymes. As will be evident to the skilled person incorporating the teachings herein, a wide range of naturally occurring introns can be applied to the method of the invention. Often native self-splicing ribozymes also encode reverse transcriptases and homing endonucleases (for example Odom et al, 2001. *Mol. Cell Biol* 21: 3472-3481). These proteins, if expressed in a host cell, cause the ribozyme to become mobile, whereupon it may randomly integrate into another part of the host cell genome. Such genome instability is not preferred for high-throughput, stable mutagenesis. Therefore, for use in a method of the invention, such naturally mobile self-splicing ribozymes should be engineered to remove their mobility functions. Such engineering can be accomplished using standard molecular biology techniques (see Sambrook and Russell, *Molecular Cloning*, Third Edition (2000), Cold Spring Harbor Laboratory Press).

Conditions suitable for splicing to occur

For splicing to occur, the target mutated nucleic acid must be transcribed. In other words, a single-stranded RNA transcript must be produced using the target nucleic acid as a template. Upon transcription, the single stranded nucleic acid should spontaneously form the structures, with host proteins and other nucleic acids if necessary, required for the splicing reactions. It is only upon transcription that step c) of the method occurs. Reactions involving the 5' and 3' splice sites will not occur when the sites are present in other nucleic formats than the prerequisite single-stranded RNA.

Transcription to produce the single stranded nucleic acid may occur *in vivo* or *in vitro*. In both cases the introduced nucleic acid should be inserted into a region of the target nucleic acid such that it is placed downstream of a promoter, and wherein there is no transcriptional terminator between the promoter and the mutated sequence.

Once the single-stranded transcript has been produced, splicing may occur. The conditions appropriate for splicing are dependent upon the 5' and 3' splice sites employed. That is to say when a 5' and 3' splice sites for a eukaryotic intron are utilised, then the milieu should also contain the constituent parts of the spliceosome. *In vitro*, these constituent parts should be provided as further reagents when the method of the invention is put into practise. *In vivo*, the host cell should be engineered such that these constituent

parts are expressed. When the 5' and 3' splicing sequences are from a tRNA intron, the milieu should contain the enzymes required to catalyse the reaction, and be of appropriate conditions for these enzymes to be functional.

Similarly, whilst further protein or nucleic acid components are not required for action of the self-splicing ribozymes, the milieu, *in vitro* or *in vivo*, should be of appropriate conditions for the ribonucleotide enzymes to function, such as the presence of divalent metal cations for group I and II introns (e.g. Sontheimer et al., 1999 *Genes & Development* 13: 1729-1741, Zhang & Leibowitz 2001 *Nucleic Acids Research* 29: 2644-2653).

Applications of the method

The applications described herein are appropriate to episome engineering in both prokaryotes and eukaryotes, although they are also very suited to engineering large molecules, and even genomes.

Site-directed mutagenesis

According to one embodiment of the invention, the method can be used for site-directed mutagenesis of a target nucleic acid molecule. The method can be illustrated by reference to Figure 2 herein. This method is based on the selection of a sequence in the target nucleic acid molecule (for example a chromosome, episome or plasmid) that will be subject to mutagenesis. The Figure should be read in anti-clockwise direction starting from the upper right part.

In this example, the target nucleic acid has been prepared as a PCR product. In this PCR reaction, first and second PCR primers have been designed which contain 5' and 3' regions of homology to the specific region on the target nucleic acid where it is desired to alter the sequence, which in this case involves the insertion of a sequence mutation. Here, the depicted primers contain the mutation to be introduced into the target nucleic acid, although this replacement sequence may lie anywhere external to the acceptor and donor splice sites. In Figure 2 this mutation is denoted by an asterisk, and may be followed through the mutagenesis protocol using this annotation. The two regions of homology flank an intron template which comprises splice acceptor and donor sites into which a selectable marker has been inserted. In the example given, a self-splicing ribozyme has been used.

In step a) of the method, illustrated in the top left quarter of Figure 2, a nucleic acid fragment containing a gene for selection, usually a gene conferring antibiotic resistance,

(here shown as blasticidin) is introduced into the target nucleic acid by homologous recombination. The nucleic acid fragment includes, in order from 5' to 3', a first region of homology to the target nucleic acid molecule, also termed a "homology arm"; a 5' splice site (here, the first portion of a self-splicing ribozyme); a selectable marker; a 3' splice site (here, the second portion of the self-splicing ribozyme); and a second region of homology to the target nucleic acid molecule.

In addition, the nucleic acid fragment includes a replacement nucleic acid sequence that is positioned within or between the first and second regions of homology. These two homology arms must flank the replacement sequence in the incorporated nucleic acid molecule to ensure that the replacement sequence is introduced into the target nucleic acid. Additionally, the replacement sequence must flank, *i.e.* lie outside, the splice sites. The replacement sequence may involve a deletion, a substitution or an insertion, and may be of any practical length, ranging from a single nucleotides (for example 0, 1, 2, 3, 4, 5, or 5 or more nucleotides, for example 10, 20, 40, 50), to kilobases or megabases in theory. The number of base pairs changed may thus range from a single point mutation, to multiple substitutions, insertions or deletions. Here, a single substitution is illustrated.

As detailed above, the replacement sequence may be placed entirely between the first region of homology and the 5' splice site, or entirely between the 3' splice site and the second region of homology. In an alternative, the replacement sequence may be placed within one or both of the first and/or second regions of homology. Further, a first portion of the replacement sequence may be placed between the first region of homology and the 5' splice site, and a second portion between the 3' splice site and the second region of homology, or indeed in any combination of the places listed above. The precise positioning of the replacement sequence will be dependent upon the desired strategy, and will be evident to the skilled artisan following the teachings provided herein.

The introduced nucleic acid fragment is preferably introduced into a transcriptionally-active nucleic acid. The nucleic acid formed by homologous recombination between the target nucleic acid molecule and the introduced nucleic acid fragment represents a stably integrated intermediate in the method, which is selected for using the introduced selectable marker. Upon transcription, a single-stranded RNA transcript is formed and a splicing reaction then takes place which involves excision of all sequence between the 5' and 3' splice sites. Here, a self-splicing ribozyme is illustrated. In this transcript, the three-dimensional structure of the ribozyme will form spontaneously. In order for the ribozyme to form in the correct confirmation, complementary sections of the ribozyme

must base-pair together to form short double-stranded sections which comprise the framework of the ribozyme. These sections of the ribozyme must be preserved. The selectable marker must therefore be inserted into a loop portion of the ribozyme. In other words, the marker should be inserted into a region of the ribozyme which is a single-stranded portion between the double-stranded portions of the ribozyme, but in a fashion such that it does not interfere with the self-splicing catalytic mechanism. This is shown schematically in the bottom part of Figure 2.

The product thus obtained will be a mutated target nucleic acid. In Figure 2, a single point mutation is illustrated, however the same approach can be applied to mutate any of the base pairs that fall within the replacement sequence created between the homology region and the self-splicing ribozyme. For example, mutations may comprise substitutions, insertions or deletions. Following the excision of the region between the two splice sites, here defined by the components of a self-splicing ribozyme, the transcript now contains only the natural sequence and the introduced mutation, and no extraneous sequences from the self-splicing ribozyme, or the selectable marker encoded within it. In Figure 2, following excision of the self-splicing ribozyme, the mRNA is translated into a protein which incorporates the mutation in the replacement sequence of the introduced nucleic acid fragment.

Insertion

A further embodiment of the method of the invention provides a simple variation to achieve small insertions and replacements. All steps are the same as illustrated above except the inserted nucleic acid fragment is differently configured to suit the experimental purpose.

According to this embodiment, the sequence to be inserted or replaced may be positioned adjacent to the self-splicing ribozyme in the target nucleic acid molecule. In an alternative, the sequence to be inserted or replaced is positioned within one or both of the homology regions. This embodiment may also be envisaged using Figure 2 as a guide. In this case, the asterisk denotes an insertion of one or more base pairs. In this embodiment, the nucleic acid fragment that is introduced in step a) of the method includes, at either end, within or between the homology arms, a replacement nucleic acid sequence that includes an insertion relative to the original target nucleic acid. In this embodiment of the method, the introduced nucleic acid is selected for using the selectable marker, which remains in the target nucleic acid. Upon transcription of the target nucleic acid, preferably a self-splicing ribozyme is used to catalyse its own excision from the transcript, but

variants on this strategy are possible, as discussed above. In this manner, excision of the self-splicing ribozyme, and the selectable marker within it, leaves behind the inserted sequence. In this manner, insertion of the sequence into the transcript is achieved without leaving any scar or residual heterologous sequence.

The inserted sequence can be any sequence of any length, ranging from single or multiple base pair insertions through tens of bps, hundreds, thousands (kpbs), and even tens of thousands (mbps). Often insertions will be of less than 1000 bps length, for example, about 2, 5, 10, 20, 50, 100, 200 or 500 bps. The obtained product will be a target nucleic acid containing the inserted sequence in any chosen place including the self-splicing ribozyme and selectable marker. The inserted sequence can be any sequence that can be introduced to the left (5'), or right (3'), or both (5' and 3'), of the self-splicing intron. Short insertions may be incorporated in synthetic oligonucleotides that are then used to amplify a fragment encoding an intron PCR template, as shown in Figure 2. Such insertions are only limited by the size of synthetic oligonucleotides. Longer insertions may be constructed by the insertion of sequences in both the forward and reverse primers which are then incorporated into the sequence of the PCR product. Such sequences will be joined together in the transcript following the excision of the ribozyme. Design of sequence in the primer in order to put into practice this embodiment of the invention will be of no burden to the skilled artisan following the teachings herein. Alternatively, larger insertions can be constructed in episomes, for example, a plasmid vector. Following cleavage from the vector by restriction nucleases, the excised fragment may then be used in the method described above.

Usefully, the inserted sequence can be one, or more than one, codon(s) (that can replace codon(s) in the target nucleic acid molecule. Also possible are larger in-frame insertions which can be accomplished with the insertion of larger lengths of nucleic acid. It will generally be preferred that the inserted length of nucleic acid contains a number of nucleotides which is divisible by three, such that the reading frame is maintained. In one preferred embodiment, the inserted sequence is 3 bp long and in-frame within a coding region. Hence this method encompasses codon-based mutagenesis, which is superior to single nucleotide mutagenesis for the mutation of protein coding regions.

Furthermore, by including in the introduced nucleic acid fragment, sequence homology arms that span regions of non-identical sequence compared to the target nucleic acid molecule, further mutations such as substitutions, (for example, point mutations),

insertions and/or deletions may be simultaneously introduced into the target nucleic acid molecule.

The inserted sequence can be a loxP site or any other site-specific recombinase site; it can also be a short tag or a restriction site. Other examples will be clear to those of skill in the art.

Large insertions

A further embodiment of the invention provides a method suitable for larger insertions. In this method, the introduced nucleic acid fragment may be comprised of two or more components. "Triple recombination" is a term coined to describe an embodiment of the method of this aspect of the invention where the introduced nucleic acid is itself made from two fragments. "Quadruple recombination" is used to describe an embodiment of the method of this aspect of the invention where the introduced nucleic acid is itself made from three fragments. The use of triple and quadruple recombination is particularly applicable to the current invention.

In one preferred embodiment of the method of the invention, the introduced nucleic acid fragment is made from two separate fragments *in vivo*. In triple recombination, where two components are involved each component includes a region of homology to the target nucleic acid molecule. Additionally, both components include a region of mutual homology that allows the components to undergo homologous recombination to knit together to form a single nucleic acid fragment for introduction into the target nucleic acid molecule.

In one embodiment, one of the components includes both 5' and 3' splice sites and the selectable marker. This is the most convenient arrangement of the components. Yet, as will be evident to the skilled addressee, this embodiment may also be put into practice wherein the regions of homology used to knit together the components to form the introduced nucleic acid fragment are part of the sequence of the self-splicing ribozyme or the selectable marker.

This method is thus the same as the more simple methods described above except that the introduced fragment is the product of a triple recombination process. This process is described in co-pending International application PCT/IB2009/000488, entitled "Method of nucleic acid recombination". Although not as efficient as the simple recombination applications illustrated above, this triple intermediate recombination works with acceptable efficiency because the internal recombination product must occur for the

primary recombination step to integrate the selectable marker. This represents a convenient way to introduce larger sequences such as coding regions for fluorescent proteins. Examples of lengths suitable for insertion in accordance with this embodiment of the invention range from a few hundred base pairs to many hundreds, thousands, tens of thousands or even hundreds of thousands of base pairs.

In one preferred embodiment of the method of the first aspect of the invention, the introduced nucleic acid fragment is made from three separate fragments *in vivo*. A first fragment comprises both 5' and 3' splice sites, encompassing the selectable marker, flanked on either side by two annealing regions which are not capable of annealing to the target nucleic acid. A second fragment comprises a first region of homology capable of annealing with a first sequence on the target nucleic acid, and a second region of homology capable of annealing with a first annealing region on the first fragment. In this embodiment a third fragment comprises a first region of homology capable of annealing with a second sequence of the target nucleic acid, and a second region of homology capable of annealing with a second annealing region on the first fragment.

In one embodiment the second or third fragment may comprise the coding sequence of a gene. In a further embodiment, the second and/or third fragments may comprise a partial coding sequence of a gene. In a further embodiment the second and/or third fragments may comprise a promoter element. In a further embodiment the second and third fragments may comprise partial coding sequences of the same gene such that following excision of the self-splicing ribozyme the coding sequence of the gene is reconstituted in the transcript.

Embodiments of the invention that exploit triple and quadruple recombination may incorporate methods of terminal adaptation, as detailed in co-pending application PCT/IB2009/000488, entitled "Method of nucleic acid recombination". According to this methodology, in adapting the fragments the appropriate strands of each fragment can be preferentially degraded such that the efficiency of the recombineering step is maximised. Preferably, the preferentially degraded strands are the strands of the first and third fragments that are not capable of annealing to the lagging strand of the target nucleic acid at a replication fork. It is preferable that the retained strand of second nucleic fragment is the strand which can anneal to the sequence of the preferentially degraded first and third fragments.

Deletion

The method of the invention is also particularly useful for the generation of seamless deletions within transcribed nucleic acids. All steps are the same as those illustrated above except the inserted nucleic acid fragment is differently configured to suit the experimental purpose. By appropriate design of the primers employed to generate the nucleic acid fragment introduced by the method of the invention, the method can effect a deletion from the target nucleic acid. In order to effect a deletion, the regions of homology must be designed so that the sequence within or between the regions of homology on the introduced nucleic acid fragment, excluding those nucleotides which encode the self-splicing ribozyme and selectable marker, lacks nucleotides which are present in the corresponding region between the regions of homology on the target nucleic acid molecule.

The deleted sequence can be any sequence of any length although most often it will be less than 1000 bps for convenience, for example, about 2, 3, 5, 10, 20, 50, 100, 200 or 500 bps. Deletions of much larger lengths of DNA are also achievable using the method of the invention, for example, 1000 bps, 2000 bps, 5000bps, or many kilobasepairs, for example one megabase, or larger than one megabase. Deletions are preferably 3bp, such that a single amino acid residue is deleted from a coding nucleic acid, or larger, for example deletion of a domain in a protein, an entire gene, or multiple genes in an operon. The obtained product will be the altered target nucleic acid deleted for a portion of sequence (determined by the regions of homology of the introduced nucleic acid), wherein the portion of sequence has been replaced by the intron template and selectable marker in its place.

This embodiment of the method of the invention is also suitable for applications involving triple and quadruple recombination. A fragment comprising the 5' and 3' splice sites and selectable marker may be co-transformed into a cell with oligonucleotides, wherein the oligonucleotides contain sequence homology to both the fragment and the target nucleic acid. In this way, by simply varying the supplied oligonucleotides which direct the homologous recombination of the intron into the target nucleic acid, panels of deletion mutants incorporating a selectable maker can be created easily. Such a strategy will be of great use in high throughput codon, domain, gene or operon targeted deletion studies. As detailed above regarding insertions, this embodiment is particularly suited to method of terminal adaptation (see co-pending application PCT/IB2009/000488, entitled "Method of nucleic acid recombination").

General considerations

There is no restriction to the type of alteration event to which the present application is applied, although immediately-apparent applications include those which are extremely difficult or time-consuming using approaches that are currently available. Particularly the alteration may be one which is not amenable to high-throughput methodologies using current techniques. Examples include the precise modification of endogenous nucleic acid molecules in any species, such as yeast chromosomes, mouse embryonic stem cell chromosomes, *C. elegans* chromosomes, Arabidopsis and Drosophila chromosomes, human cell lines, viruses and parasites, or exogenous molecules such as plasmids, yeast artificial chromosomes (YACs) and human artificial chromosomes (HACs).

In all the aspects of the invention described herein, the introduced nucleic acid fragments may be circular or linear, but are preferably linear DNA or RNA molecules, either double-stranded or single-stranded. DNA is generally preferred. Preferred nucleic acids thus include single-stranded DNA or RNA, in either orientation, 5' or 3'. Annealed oligonucleotides may also be used, either with blunt ends, or possessing 5' or 3' overhangs. In one embodiment, single-stranded oligonucleotides are used. In another embodiment, single-stranded deoxyribonucleotides are used. Introduced nucleic acid molecules carrying a synthetic modification can also be used.

It should be noted that the introduced nucleic acid fragments do not necessarily represent a single species of nucleic acid molecule. For example, it is possible to use a heterogeneous population of nucleic acid molecules, for example, to generate a DNA library, such as a genomic or cDNA library.

A number of different types of target nucleic acid molecule may be used in the method of the invention. Accordingly, intact circular double-stranded nucleic acid molecules (DNA and RNA), such as plasmids, and other extrachromosomal DNA molecules based on cosmid, P1, BAC or PAC vector technology may be used as the target nucleic acid molecule according to the invention described above. Examples of such vectors are described, for example, by Sambrook and Russell (Molecular Cloning, Third Edition (2000), Cold Spring Harbor Laboratory Press) and Ioannou et al. (Nature Genet. 6 (1994), 84-89) and the references cited therein.

The target nucleic acid molecule may also be a host cell chromosome, such as, for example, the *E. coli* chromosome. Alternatively, a eukaryotic host cell chromosome (for example, from yeast, *C. elegans*, Drosophila, mouse or human) or eukaryotic extrachromosomal DNA molecule such as a plasmid, YAC and HAC can be used.

Alternatively, the target nucleic acid molecule need not be circular, but may be linear. Preferably, the target nucleic acid molecule is a double-stranded nucleic acid molecule, more preferably, a double-stranded DNA molecule.

The method of the invention may be effected, in whole or in part, in a host. Suitable hosts include cells of many species, such prokaryotes and eukaryotes, and also including viruses and parasites, although bacteria, such as gram negative bacteria are a preferred host. More preferably, the host cell is an enterobacterial cell, such as a *Salmonella*, *Klebsiella*, *Photobabdus*, *Psuedomonas*, *Neisseria* or *Escherichia coli* cell (the method of the invention works effectively in all strains of *E. coli* that have been tested so far). It should be noted, however, that the method of the present invention is also suitable for use in gram positive hosts such as *Bacillus* and eukaryotic cells or organisms, such as fungi, plant or animal cells, as well as viral and parasitic cells and organisms. The system has been demonstrated to function well in ES cells, specifically mouse ES cells, and there is no reason to suppose that it will not also be functional in other eukaryotic cells.

In embodiments in which the method is effected, in whole or in part, in a host cell, the host cell is typically an isolated host cell, although the use of non-isolated host cells is also envisaged.

The method of the invention may comprise the contacting of the introduced and target nucleic acid molecules *in vivo*. In one embodiment, the introduced nucleic acid molecule may be transformed into a host cell that already harbours the target nucleic acid molecule. In a different embodiment, the introduced and target nucleic acid molecules may be mixed together *in vitro* before their co-transformation into the host cell. Of course, one or both of the species of nucleic acid molecule may be introduced into the host cell by any means, such as by transfection, transduction, transformation, electroporation and so on. For bacterial cells, a preferred method of transformation or cotransformation is electroporation.

In some embodiments, the method is effected, in whole or in part, *ex vivo*. In some embodiments, methods for treatment of the human or animal body by surgery or therapy are excluded from the scope of the invention.

In one embodiment the homologous recombination of the method is initiated entirely *in vitro*, without the participation of host cells or the cellular recombination machinery. Phage annealing proteins such as RecT are able to form complexes *in vitro* between the protein itself, an oligonucleotide molecule and a double-stranded nucleic acid molecule (Noirot and Kolodner, J Biol Chem 273 (1998), 12274-12280). One example of such a

complex is that formed between RecT, a ssDNA oligonucleotide and an intact circular plasmid. Such complexes lead to the formation of complexes that are herein termed "joint molecules" (consisting, in this example, of the plasmid and the ssDNA oligonucleotide). Such joint molecules have been found to be stable after removal of the phage annealing protein. The formation of stable joint molecules has been found to be dependent on the existence of shared homology regions between the ssDNA oligonucleotide and the plasmid.

The potential of RecA to make joint molecules *in vitro* has already been exploited to allow the isolation of desired DNA strategies from a pool, for example in RecA-assisted cloning (Ferrin and Camerini-Otero, Proc Natl Acad Sci USA 95 (1998), 2152-2157, for review see Ferrin, Methods Mol Biol 152 (2000), 135-147), Li MZ, Elledge SJ. MAGIC, an *in vivo* genetic method for the rapid construction of recombinant DNA molecules. Nat Genet. 2005 37:311-9 and in RecA-mediated affinity capture (Zhumabayeva et al., Biotechniques 27 (1999), 834-840). It is proposed herein that so-called "joint molecules" as described above may be used directly to mediate recombination in a host cell, where the host cell does not need to express any phage annealing protein whatsoever.

As detailed herein, the methods of the invention rely on recombination events that involve the replacement of a section of target nucleic acid for an equivalent section of introduced nucleic acid, to which the introduced fragment is directed through the existence of shared regions of sequence homology between the two molecule types. As with conventional homologous recombination events, the introduced nucleic acid becomes covalently attached to the target nucleic acid. In this manner, the sequence information in the introduced nucleic acid molecule becomes integrated into the target nucleic acid molecule in a precise and specific manner, and with a high degree of fidelity. The efficiency of this step, when coupled with a selection step, is high, and allows the simple manipulation of sequences.

The nucleic acid molecule fragments used to replace target sequence may be single-stranded. This single-stranded nucleic acid may be generated *in vivo* or *in vitro*. In other words the single-stranded nucleic acid may be generated in a host cell. The generation of the single-stranded replacement nucleic acid from the double-stranded nucleic acid substrate prior to recombination may be mediated by any suitable means. The double-stranded nucleic acid substrate may be adapted such that one strand is preferentially degraded entirely to leave the other strand as the single-stranded replacement nucleic acid (see co-pending International application PCT/IB2010/000893 filed on 20th February

2009 and entitled "Method of nucleic acid recombination"). The degradation is preferably mediated by an exonuclease. The exonuclease may be a 3' to 5' exonuclease but is preferably a 5' to 3' exonuclease. Preferably, the 5' to 3' exonuclease is Red alpha (Kovall, R. and Matthews, B.W. Science, 1997, 277, 1824-1827; Carter, D.M. and Radding, C.M., 1971, J. Biol. Chem. 246, 2502-2512; Little, J.W. 1967, J. Biol. Chem., 242, 679-686) or a functional equivalent thereof. In an alternative embodiment, the exonuclease is RecBCD. Alternatively and/or additionally, the single-stranded replacement nucleic acid is generated from the double-stranded nucleic acid substrate by a helicase. The helicase separates the dsDNA substrate into two single-stranded nucleic acids, one of which is the single-stranded replacement nucleic acid. The helicase may be either a 5'-3' or 3'-5' helicase. Preferably the helicase is RecBCD whilst it is inhibited by Red gamma. In other preferred embodiments, the helicase is any helicase of the RecQ, RecG or DnaB classes. In some embodiments, the single-stranded replacement nucleic acid generated by the helicase is preferentially stabilised relative to the other single-stranded nucleic acid generated by the helicase.

In preferred embodiments, the step of generating the single-stranded replacement nucleic acid from the double-stranded nucleic acid substrate is carried out in a host cell in which the recombination occurs. Alternatively, the step of generating the single-stranded replacement nucleic acid may be carried out in a separate host cell from the host cell in which the recombination occurs and may then be transferred to the host cell in which recombination occurs by any suitable means, for example, by transduction, transfection or electroporation. Alternatively, the step of generating the single-stranded nucleic acid from the double-stranded nucleic acid substrate may be carried out *in vitro*. Thus, the requirement in the host cell in which recombination takes place for Red alpha or an alternative enzyme that preferentially degrades one strand of the double-stranded nucleic acid substrate, or which separates the two strands, may be bypassed by providing the single-stranded nucleic acid to the host cell.

Advantageously, adapting one or both 5' ends of the double-stranded nucleic acid increases the yield of the single-stranded nucleic acid. Preferably, this increase in yield is due to the effect of adapting the 5' end(s) on the enzymes that act to generate the single-stranded nucleic acid.

Preferably, the double-stranded nucleic acid substrate is adapted so that it is asymmetric at its 5' ends. The asymmetry preferably causes one strand to be preferentially degraded. This preferably results in the other strand being maintained and so the production of a

single-stranded nucleic acid is favoured, thereby improving the yield of the single-stranded nucleic acid.

By preparing a double-stranded nucleic acid substrate with asymmetric 5' ends and bringing this into contact with a target nucleic acid in the presence of Red beta and a suitable degradation/separation enzyme (preferably Red alpha or a helicase), it is possible to increase engineering efficiencies to levels greater than any other configuration yet described for recombineering methodologies. Therefore, the method of the invention preferably utilises a double-stranded nucleic acid substrate having asymmetry at its 5' ends wherein the method is conducted in the presence of Red alpha and/or a helicase and in the presence of Red beta. Red gamma is preferably also present as Red gamma inhibits RecBCD, which degrades double-stranded DNA. Another efficient way to engineer DNA using Red-mediated homologous recombination employs a double-stranded nucleic acid substrate that is adapted to have asymmetric 5' ends in the presence of Red beta and Red gamma, without Red alpha. A less efficient but still operable way to engineer DNA using Red-mediated homologous recombination employs a double-stranded nucleic acid substrate that is adapted to have asymmetric 5' ends in the presence of Red beta, without Red gamma (or a functional equivalent thereof) and without Red alpha (or a functional equivalent thereof). Such a method is also encompassed within the scope of the invention.

Any suitable method of making a double-stranded nucleic acid substrate asymmetric such that one strand is preferentially degraded whilst the other is maintained is envisaged by the present invention. The asymmetry may be conferred, for example, by one or more features present in only one strand of the double-stranded nucleic acid substrate or by one or more features present in both strands of the double-stranded nucleic acid substrate, wherein different features are present in different strands.

Preferably, the asymmetry is present at or in close proximity to the 5' ends of the two strands of the double-stranded nucleic acid substrate, most preferably at the 5' ends. For example, the asymmetry is preferably present at the 5' end of the 5' homology regions of the double-stranded nucleic acid substrate, or may be present in a region 5' of the 5' identity regions. The "homology regions" of the double-stranded nucleic acid substrate correspond to the regions of the single-stranded nucleic acid that are identical to sequence on the target nucleic acid, or are complementary thereto. For example, the double-stranded nucleic acid substrate may have one or more features at or in close proximity to the 5' end of one of its strands but not at or in close proximity to the 5' end of the other strand which make it asymmetric.

Preferably, the asymmetry is conferred by a modification to the nucleic acid sequence. Preferably, the modification affects the progression of exonuclease, preferably a 5'-3' exonuclease, preferably Red alpha exonuclease, on one strand but does not affect the progression of the exonuclease on the other strand. For example, the modification may inhibit the progression of exonuclease on one strand such that the exonuclease preferentially degrades the other strand. By "inhibit" the progression of exonuclease is meant that the modification inhibits the progression of the exonuclease on that strand relative to the other strand, for example, by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, most preferably 100%. For example, the modification may be the inclusion of a blocking DNA sequence, such as the Red alpha exonuclease pause sequence, more preferably, the Red alpha pentanucleotide pause sequence GGCGA, more preferably GGCGATTCT, more preferably, the left lambda cohesive end, also called the cos site (Perkins TT, Dalal RV, Mitsis PG, Block SM Sequence-dependent pausing of single lambda exonuclease molecules. Science 301:1914-8). The Red alpha exonuclease pause site may, for example, be placed at or in close proximity to the 5' end of one strand but not at or in close proximity to the 5' end of the other strand.

In a further preferred embodiment, the modification prevents the exonuclease from binding to one strand of the double-stranded nucleic acid substrate such that only the other strand is degraded. In a further preferred embodiment, the modification does not prevent the exonuclease from binding but blocks it from degrading one strand or both of the double-stranded nucleic acid substrate such that the strand that will anneal to the lagging strand template is stabilized upon separation from the dsDNA substrate by a helicase. In an alternative embodiment, the modification may promote the progression of exonuclease, preferably of 5'-3' exonuclease, more preferably Red alpha exonuclease, on one strand such that the exonuclease preferentially degrades that strand relative to the other strand. By "promote" the progression of exonuclease is meant that the modification promotes the progression of exonuclease activity on that strand relative to the other strand, for example, by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 100%, at least 200%, at least 300%, or at least 400%. In embodiments in which the two strands of the double-stranded nucleic acid substrate are separated by a helicase, the modification may serve to preferentially stabilise one strand, for example, by preventing an exonuclease or endonuclease from binding to that strand. In another embodiment, the

modification prevents exonuclease degradation of both strands such that one strand is protected and can be released from the other by the action of a helicase.

In a preferred embodiment, the modification is one or more covalent modifications. Preferably, the covalent modification is present at or in close proximity to the 5' end of one strand but is not present at or in close proximity to the 5' end of the other strand. More preferably, the covalent modification is present at the 5' end of one strand but is not present at the 5' end of the other strand.

Preferred covalent modifications are the presence of a replacement nucleotide, such as the presence of a hydroxyl group or a phosphothioester bond. Such covalent modifications disfavour the action of exonucleases.

For example, in embodiments in which it is desired to protect the 5' end of the strand to be maintained, the covalent modification is preferably selected from one or more of the following:

- one or more phosphothioates in place of one or more phosphodiester bonds. Preferably, the phosphothioate(s) is present in place of the 5'-most bond in the 5' identity region, or are present in place of the first two bonds, or are present in place of up to each of the first six (e.g. 3, 4, 5, 6) or more bonds in the 5' identity region;
- one or more phosphoacetates in place of one or more phosphodiester bonds. Preferably, the phosphoacetate(s) is present in place of the 5'-most bond in the 5' identity region, or are present in place of the first two bonds, or are present in place of up to each of the first six (e.g. 3, 4, 5, 6) or more bonds in the 5' identity region;
- one or more locked nucleotides (preferably LNA; 2'-O and/or 4'-C-Methylen-beta-D-ribofuranosyl) in place of one or more nucleotides. Preferably, the one or more locked nucleotides are present in place of the first nucleotide in the 5' identity region, or are present in place of the first two nucleotides, or in place of up to the first six (e.g. 3, 4, 5 or 6) nucleotides;
- a hydroxyl group. Preferably, the 5' most nucleotide of the substrate is also the 5' most nucleotide of the region that is identical to sequence on the target nucleic acid and the hydroxyl group is at the 5' end of this region of sequence identity;
- a 5' protruding end. For example, the covalent modification may be 2 or more protruding nucleotides, 4 or more protruding nucleotides, 6 or more protruding nucleotides, preferably 11 or more protruding nucleotides, preferably a 5' end containing the Red alpha pause sequence, preferably the left lambda cohesive end known as cos;

- any other covalent adduct that renders resistance to 5'-3' exonucleases. For example, the 5' end may be modified to contain an attached adduct such as biotin deoxygenin or fluorophore such as FITC.

In embodiments in which it is desired to render one strand of the double-stranded nucleic acid substrate sensitive to 5'-3' exonucleases such that the other strand is the strand to be maintained, the covalent modification is preferably selected from one or more of the following:

- a 5' phosphate group;
- a 5' end that is either flush or recessed with respect to the adjacent 3' end;
- a 5' end that carries a stretch of DNA sequence that is not identical to the target DNA. The stretch of DNA sequence may be, for example, 1-29 bps in length, more preferably 30-99 bps in length, more preferably 100-999 bps in length, even more preferably more than 1 kb in length;
- a 5' end that includes deoxy uridine nucleotides in place of deoxy thymidine nucleotides in the DNA strand;
- any other covalent adduct that conveys sensitivity to 5'-3' exonucleases.

Also encompassed within the scope of the invention are methods which use a double-stranded nucleic acid substrate for the production of a single-stranded nucleic acid that contains one or more covalent modifications that protect the 5' end of the strand to be maintained and also one or more covalent modifications that render the other strand of the double-stranded nucleic acid substrate sensitive to 5'-3' exonucleases. For example, the double-stranded nucleic acid substrate may lack the 5' phosphate (i.e. presence of hydroxyl) on one strand whilst the other strand comprises the 5' phosphate.

Preferably, the double-stranded nucleic acid substrate is adapted such that it comprises a 5' phosphothioate at one of its 5' ends but not at the other 5' end.

Any other chemical modification at or near the 5' end which inhibits or promotes exonuclease progression or blocks exonuclease binding is also encompassed within the scope of the invention.

As mentioned above, the asymmetry may be caused by the double-stranded nucleic acid substrate having different extensions of single-strandedness; that is different combinations of 5' protruding, blunt (or "flush") or 3' protruding ends. For example, the double-stranded nucleic acid substrate may have only one 5' protruding end, only one 3'

protruding end and/or only one blunt end. The asymmetry may be created by restriction cleavage to create different ends on the nucleic acid substrate. Restriction enzymes leave either 5' protruding, blunt or 3' protruding ends. The 5' protruding ends are least favoured for Red alpha digestion. Thus, in one embodiment, the double-stranded nucleic acid substrate preferably has only one 5' protruding end. In embodiments in which the asymmetry is generated by different extensions of single-strandedness, it is preferred that each strand of the double-stranded nucleic acid substrate is a continuous nucleic acid strand.

The asymmetry may alternatively be caused by the double-stranded nucleic acid substrate having different extensions of double-strandedness. For example, one end may have no additional nucleic acid sequence beyond the end of the identity region, and the other may have additional non-identical sequences. The additional non-identical sequences may be as short as 4 base pairs, however, preferably will be longer than 10 base pairs, and more preferably longer than 100 base pairs.

As mentioned above, it has also been found that homologous recombination may also occur in the absence of the Red α exonuclease when a double-stranded nucleic acid substrate is exposed to a target nucleic acid under conditions suitable for recombination to occur. It is hypothesised that a helicase acts to separate the two strands of the double-stranded nucleic acid substrate and that the strand that is the single-stranded nucleic acid is then available for use in homologous recombination. Surprisingly, it has been found that adapting both of the 5' ends of the double-stranded nucleic acid substrate leads to improved efficiencies of homologous recombination in such systems compared to systems in which the 5' ends are not adapted. Thus, in an alternative embodiment, the double-stranded nucleic acid is symmetrically adapted at both of its 5' ends. In a preferred embodiment, the double-stranded nucleic acid substrate is covalently modified at both of its 5' ends. Particularly preferred is the use of a double-stranded nucleic acid substrate in which both 5' ends are covalently modified with a biotin molecule, or more preferably, with a phosphothioate. Preferably, in such embodiments, the recombination is carried out in the absence of Red α . Alternatively, the invention also envisages using a helicase to generate the single-stranded nucleic acid from a double-stranded nucleic acid substrate that has 5' asymmetric ends, as described above.

The skilled person will understand the techniques required to adapt the double-stranded nucleic acid substrate to make it asymmetric. For example, following cleavage by a restriction enzyme, the substrate may be dephosphorylated with alkaline phosphatase, and

then cleaved with a second restriction enzyme. As restriction enzymes usually leave phosphates on the 5' end, this will generate an asymmetrically phosphorylated substrate.

Two oligonucleotides may be designed for use as the terminal identity regions as is usual for a recombineering exercise. These oligonucleotides may be chemically synthesized so that their 5' ends are different with respect to the presence of a replacement nucleotide at or in close proximity to the 5' end. These oligonucleotides can be used, for example, for oligonucleotide-directed mutagenesis after annealing, or PCR on templates to create the asymmetrically ended double-stranded nucleic acid substrate or mixed with standard double-stranded nucleic acid cassettes and co-introduced into a host for 'quadruple' recombination.

The double-stranded nucleic acid substrate may be made by any suitable method. For example, it may be generated by PCR techniques or may be made from two single-stranded nucleic acids that anneal to each other. The double-stranded nucleic acid substrate may in particular be generated by long range PCR. Long range PCR has been used in the art to generate double-stranded fragments, for example of up to 50kb (Cheng et al. (1994) Proc Natl Acad Sci 91: 5695-5699). The 5' ends of one or both of the primers used in this long range PCR may be adapted so that the PCR product is suitable for use as the double-stranded nucleic acid substrate in the methods of the invention.

A preferred embodiment for the invention is to perform the homologous recombination in a host cell mutated for exonucleases, specifically an *E. coli* host mutated for *sbcB*. Thus, the invention also provides a method comprising performing the homologous recombination in a host cell in which the activity of its endogenous *sbcB* exonuclease, or the orthologue or functional equivalent thereof, has been inactivated or reduced. Also provided is a host cell in which the activity of its endogenous *sbcB* exonuclease, or the orthologue thereof, has been reduced or inactivated relative to its wild-type counterpart; such a host cell forms an aspect of the present invention. Preferably, the host cell is *E. coli*. *SbcB* or its orthologue or functional equivalent may be inactivated or the activity thereof may be reduced by way of a mutation. Preferably, the mutation inactivates the *SbcB* or its orthologue. Any suitable mutation is envisaged, for example, a deletion, insertion or substitution. For example, the entire gene encoding the exonuclease may be deleted or one or more point mutations may be used to inactivate the *SbcB* or its orthologue. The exonuclease may be inactivated in any other appropriate way, for example, by gene silencing techniques, by the use of exonuclease-specific antagonists or by degradation of the exonuclease.

Methods which utilise the mutant of the SbcB/orthologue/functional equivalent described above, may be a method according to the present invention. Also provided is the use of the SbcB mutants (and corresponding orthologues/functional equivalents) in broader aspects of homologous recombination technology. Thus, there is provided a method of altering the sequence of a target nucleic acid comprising (a) bringing a first nucleic acid molecule into contact with a target nucleic acid molecule in the presence of a phage annealing protein, or a functional equivalent or fragment thereof, wherein said first nucleic acid molecule comprises at least two regions of shared sequence homology with the target nucleic acid molecule, under conditions suitable for repair recombination to occur between said first and second nucleic acid molecules and wherein the functional equivalent or fragment retains the ability to mediate recombination and wherein the activity of the host's endogenous sbcB exonuclease or orthologue or functional equivalent thereof has been inactivated or reduced; and (b) selecting a target nucleic acid molecule whose sequence has been altered so as to include sequence from said first nucleic acid molecule. Preferably, the phage annealing protein is Red beta or a functional equivalent thereof. The method may be carried out in the absence or presence of one or both of Red alpha and/or Red gamma or their functional equivalents.

Preferably, the target nucleic acid is the lagging strand template of a DNA replication fork and the inserted nucleic acid has 5' and 3' homology regions that can anneal to the lagging strand template of the target DNA when it is replicating. The term "lagging strand", as used herein, refers to the strand that is formed during discontinuous synthesis of a dsDNA molecule during DNA replication. The single-stranded replacement nucleic acid anneals through its 5' and 3' identity regions to the lagging strand template of the target nucleic acid and promotes Okazaki-like synthesis and is thereby incorporated into the lagging strand. The direction of replication for plasmids, BACs and chromosomes is known, and so it is possible to design the double-stranded nucleic acid substrate so that the maintained strand is the one that will anneal to the lagging strand template.

In some embodiments, the double-stranded nucleic acid substrate is made from two or more double-stranded nucleic acids or from one or more double-stranded nucleic acids together with one or more single-stranded oligonucleotides. The use of two double-stranded nucleic acids to make the double-stranded nucleic acid substrate is referred to herein as 'triple' recombination because there are two double-stranded nucleic acid molecules which are used to make the double-stranded nucleic acid substrate and there is one target nucleic acid. The use of three nucleic acids to make the double-stranded

nucleic acid substrate is referred to herein as 'quadruple' recombination because there are three nucleic acids which are used to make the double-stranded nucleic acid substrate and there is one target nucleic acid. Any number of single-stranded and/or double-stranded nucleic acids may be used to make the double-stranded nucleic acid substrate provided that the resulting double-stranded nucleic acid substrate is adapted at one or both of its 5' ends such that preferential degradation of one strand and/or strand separation generates the single-stranded nucleic acid.

In all cases where more than one nucleic acid is used to make the double-stranded nucleic acid substrate, a part of each of the more than one nucleic acids must be able to anneal with a part of its neighbouring nucleic acid. For example, for triple recombination, one end of each double-stranded nucleic acid that is used to make up the double-stranded nucleic acid substrate must be able to anneal to the target, whereas the other ends of each double-stranded nucleic acid that is used to make up the double-stranded nucleic acid substrate must be able to anneal to each other. The two double-stranded nucleic acids that are used to make up the double-stranded nucleic acid substrate are adapted such that one strand of each double-stranded nucleic acid is preferentially maintained. Methods for adaptation that lead to preferential degradation are discussed above. Following degradation of one strand of each of the two double-stranded nucleic acids, the remaining single strands anneal with each other to form the double-stranded nucleic acid substrate of the invention.

It is a great strength of the methods of the invention that no complex selection steps are necessary to select for the altered molecules. The advantages of the methods stem from the use of physical markers in both selection steps, namely selection for a selectable marker in step b), which removes any need for labour intensive screening for clones containing the mutation. Moreover, this selectable marker can be subsequently used to apply a selective pressure for maintenance of the introduced nucleic acid within the altered target molecule, subsequent to its introduction. A further strength of the method of the invention, further engendering the method to high-throughput analysis is its simplicity. Only one step is required for selectable, seamless mutagenesis. Other methods require multiple sequential steps for incorporation and subsequent removal of a selectable marker. In comparison, the present invention requires only insertion, the excision is a spontaneous event which requires no further experimental manipulation to effect. As a result of these features, the invention will find an easy and convenient

application in: protein engineering, synthetic biology, bacteria genome evolution and other directed mutagenesis applications.

The invention will now be illustrated by representative examples. It will be appreciated that modification of detail may be made without departing from the scope of the invention.

Brief description of the figures

Figure 1: Schematic representation of the Group I self-splicing intron mechanism;

Figure 2: Schematic representation of the *in vivo* site directed mutagenesis by intron recombineering;

Figure 3: Schematic illustration of repair of a point mutation in a kanamycin resistance cassette by *in vivo* site directed mutagenesis by intron recombineering;

Figure 4: Restoration of kanamycin resistance following *in vivo* site directed mutagenesis by intron recombineering: Panel A shows recovered colonies on an agar plate, Panel B is a graph showing colony counts from the plates in panel A.

EXAMPLES

Example 1: Site directed mutagenesis

An antibiotic selection marker coding for the Blasticidin resistance gene under the bacterial promoter EM7 was inserted into the Group I Td Intron sequence. This intron containing cassette (IR) was then PCR amplified using specific primers. The primers carried homologous sequences to the specific DNA region where insertion of the sequence mutation was desired. Upon PCR amplification the cassette contained the mutation and the flanking homology arms to mediate the recombination (see Figure 2 for schematic representation).

The modified DNA sequence can be transcribed in all three kind of RNA (mRNA, rRNA, tRNA) and in all the cases the intron sequence will formed a looping structure able to self-catalyze *in vivo* or *in vitro* the splicing of the RNA precursor.

The self-splicing proceeded by two consecutive transesterification reactions. The first of these was initiated by an external guanosine at the 5' splice site (Figure 1). The reaction did not require any external proteins and occurred with an efficiency between 40% and 90% depending on the chosen intron sequence.

In the case of messenger RNA, the spliced RNA was translated to a protein without any scars except the chosen mutation. Due to the high fidelity of recombineering and the use of an antibiotic selection, the approach was easily applied in a high-throughput pipeline in liquid-culture.

Example 2: Corrected of a frame shift mutation in the neomycin gene by IRN

Here, a frame shift mutation in the neomycin gene (Neo*) in a BAC was corrected to restore the expression of kanamycin resistance (Figure 3).

HS996 *E. Coli* cells were electroporated with the Neo* BAC clone containing the mutated kanamycin resistant gene (Neo). This mutation was shown by the absence of growth in the presence of the Kanamycin antibiotic (Figure 4, panel A). The 5' primer used to amplify the IR cassette (in this case a Group I Td Intron sequence from phage T4 encompassing a blastacidin resistance marker) carried a sequence that, upon recombination, restored the frame of the Neo resistant gene.

Following electroporation of the IR cassette cells were selected by their resistance to blastacidin. The colonies were also shown to be resistance to kanamycin (see Figure 4, panel B).

Upon self-slicing the intron sequence was precisely removed from the RNA precursor and leads to a mutated RNA without any scars or additional not designed nucleotide, demonstrated by the restoration of kanamycin resistance.

CLAIMS

1. A method for altering the sequence of a target nucleic acid molecule to incorporate a replacement nucleic acid sequence, said method comprising the steps of:
 - a) introducing a nucleic acid fragment into the target nucleic acid molecule by homologous recombination, wherein the nucleic acid fragment comprises:
 - i) a first region of homology to the target nucleic acid molecule;
 - ii) a 5' splice site;
 - iii) a selectable marker;
 - iv) a 3' splice site;
 - v) a second region of homology to the target nucleic acid molecule;wherein components i) to v) are ordered from 5' to 3'; and
the replacement nucleic acid sequence is positioned within or between the first and second regions of homology but external of the splice sites;
 - b) selecting for the introduction of the nucleic acid fragment using the selectable marker;
 - c) incubating the product of step b) under conditions suitable for a splicing reaction to occur such that the selectable marker is excised from the target nucleic acid, and thereby generating the desired altered target nucleic acid sequence.
2. The method of claim 1, wherein the 5' and 3' splice sites are components of a self-splicing ribozyme.
3. The method of claim 2 wherein the self-splicing ribozyme is non-motile.
4. The method of claim 2 or claim 3, wherein the self-splicing ribozyme is a self-splicing intron.
5. The method of claim 4 wherein the self-splicing intron is a Group I self-splicing intron.
6. The method of any one of the preceding claims, wherein the step of altering the sequence of a target nucleic acid molecule comprises one or more insertions, deletions, or substitutions.

7. The method of any one of the preceding claims wherein the nucleic acid fragment introduced in step a) is comprised of two or more components.
8. The method of any one of the preceding claims, wherein the step of homologous recombination uses combinations of proteins comprising the Red β protein, for example the Red α , Red β and Red γ proteins.
9. The method of any one of the preceding claims, wherein the step of homologous recombination uses combinations of proteins related to the Red β protein, for example the RecT, Erf or Plu β proteins.
10. The method of any one of the preceding claims, wherein the nucleic acid introduced in step a) is a single-stranded nucleic acid.
11. The method of claim 10 wherein the single-stranded nucleic acid is formed from a double-stranded nucleic acid.
12. The method of claim 11 wherein the double-stranded nucleic acid is adapted so that it is asymmetric at its 5' ends, wherein the asymmetry causes one strand to be preferentially degraded.
13. The method of any one of claims 10-12 wherein the single-stranded replacement nucleic acid anneals to the lagging strand template of a DNA replication fork.
14. The method of claim 1 wherein the selectable marker is encoded within a loop region between the 5' and 3' splice sites.
15. The method of claim 1 wherein the selectable marker is encoded within a loop region between the 5' and 3' splice sites.
16. The method of any one of the preceding claims, wherein steps a) and b) are conducted in a host cell.
17. The method of any one of the preceding claims, wherein step c) is conducted in a host cell.
18. The method of claim 14 or claim 15 wherein the host cell is a prokaryotic host cell.
19. The method of claim 14 or claim 15 wherein the host cell is a eukaryotic host cell.

Figure 1

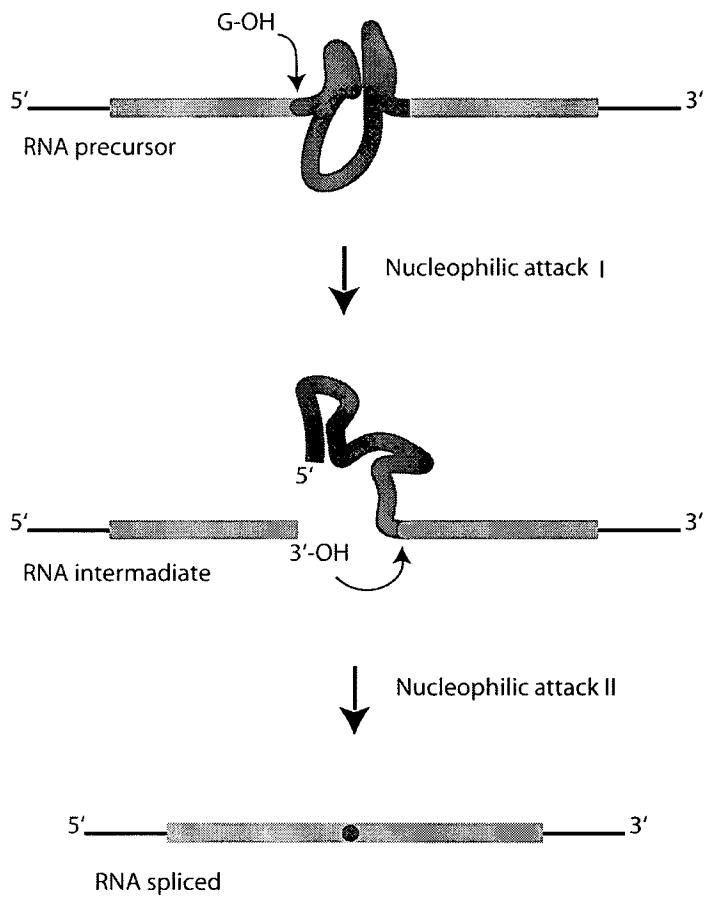


Figure 2

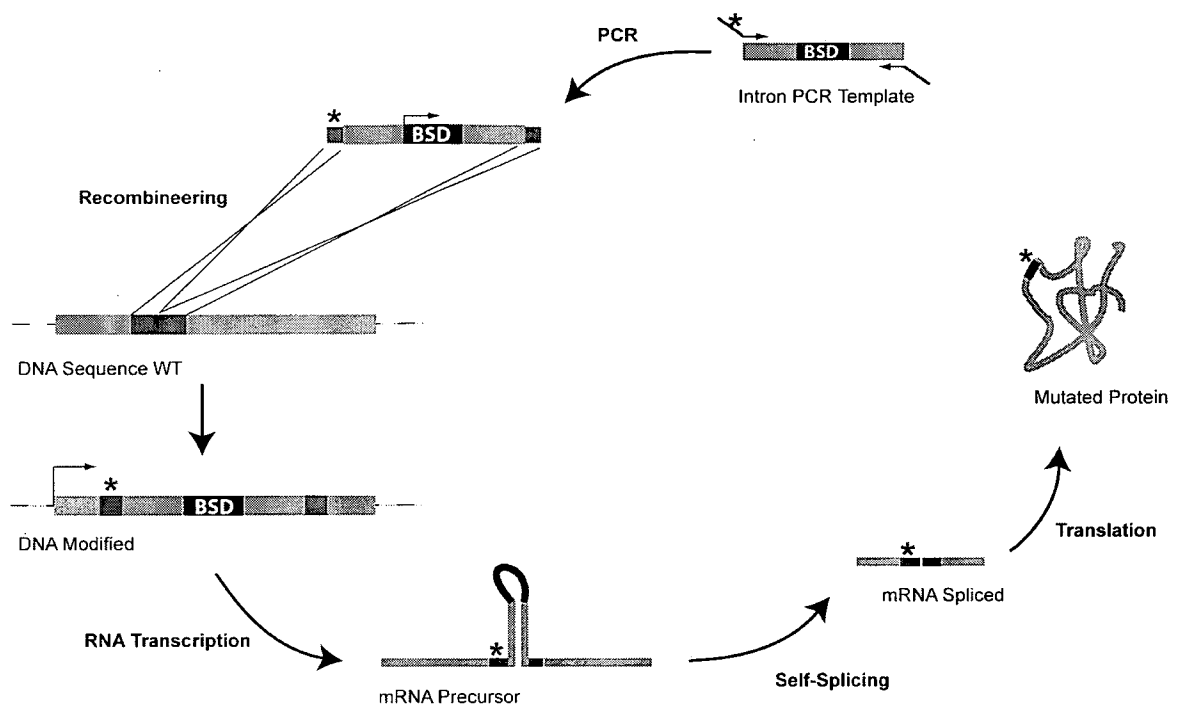


Figure 3

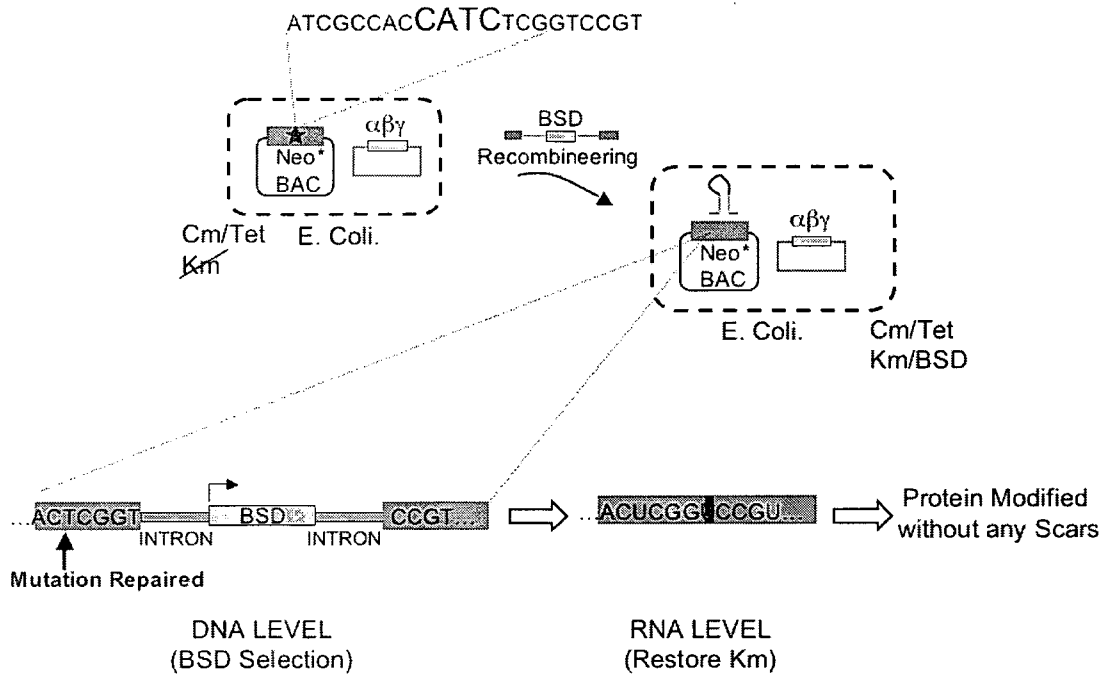


Figure 4

