Conditional Gene Trapping Construct for the Disruption of Genes

The present invention relates to a gene trapping construct which causes conditional mutations in genes, and the use of this gene trapping construct to mutationally inactivate all cellular genes. In addition the invention relates to a cell, preferably a mammalian cell which contains the above mentioned construct. Moreover, the invention relates to the use of said cell for identification and/or isolation of genes and for the creation of transgenic organisms to study gene function at various developmental stages, including the adult. In conclusion, the present invention provides a process which enables a temporally and/or spatially restricted inactivation of all genes that constitute a living organism.

Background of the Invention

For some years targeted mutagenesis in totipotent mouse embryonic stem (ES) cells has been used to inactivate genes for which cloned sequences were available (Capecchi, The new mouse genetics: altering the genome by gene targeting, Trends in Genetics, 5, 70 - 76 (1989)). Since ES cells can pass mutations induced in vitro to transgenic offspring in vivo it is possible to analyze the consequences of gene disruption in the context of the entire organism. Thus, numerous mouse strains with functionally inactivated genes ("knock out mice") have been created by this technology and utilized to study the biological function of a variety of genes. However, targeted mutagenesis requires detailed knowledge of gene structure and organization as well as its physical isolation in a cloning vector. Moreover, the targeting vector containing the gene of interest needs to be transduced into ES cells to isolate mutant ES clones in which a normal allele is replaced by a mutant allele by homologous recombination (Capecchi 1989). Overall, the generation of mutant mouse strains by this procedure is time consuming, labor intensive, expensive and inefficient because it can handle only one gene at the time (Koller and Smithies, Altering genes in animals by gene targeting. Ann. Rev. Immunol. 10, 705 - 730 (1992)). More importantly, mouse mutants created by this procedure (also known as "conventional, complete or classical mutants"), contain the inactivated gene in all cells and tissues throughout life.

CONFIRMATION COPY

A refined method of targeted mutagenesis, referred to as conditional mutagenesis, employs a site specific recombination system (e.g. Cre/loxP or Flp/frt – Sauer and Henderson, Site-specific recombination in mammalian cells by the Cre recombinase of bacteriophage P1, Proc. Natl. Acad. Sci. USA *85*, 5166-5170 (1988); Senecoff et al., DNA recognition by the FLP recombinase of the yeast 2m plasmid: a mutational analysis of the FLP binding site , J. Mol. Biol.,201, 405 - 421 (1988)) which enables a temporally and/or spatially restricted inactivation of target genes (Rajewsky et al., Conditional gene targeting, J. Clin. Invest., 98, 600 - 603 (1996)). The creation of conditional mouse mutants requires the generation of two mouse strains, i.e. the recombinase recognition strain and the recombinase expressing strain. The recombinase recognition strain is generated by homologous recombination in ES cells as described above except that the targeted exon(s) is (are) flanked by two recombinase recognition sequences (hereinafter "RRS"), e.g. loxP or frt. Since the RRS reside in introns they do not interfere with gene expression. The recombinase expressing strain contains a recombinase transgene (e.g. Cre, Flp) whose expression is either restricted to certain cells and tissues or is inducible by external agents. Crossing of the recombinase recognition strain with the recombinase expressing strain deletes the RRS-flanked exons from the doubly transgenic offspring in a prespecified temporally and/or spatially restricted manner. Thus, the method allows the temporal analysis of gene function in particular cells and tissues of otherwise widely expressed genes. Moreover, it enables the analysis of gene function in the adult organism by circumventing embryonic lethality which is frequently the consequence of gene inactivation. For pharmaceutical research, aiming to validate the utility of genes and their products as targets for drug development, inducible mutations provide an excellent genetic tool. However, as is the case for conventional mutants, the generation of conditional mutants is a time consuming and labor intensive procedure which must be performed individually for every gene.

To address some of the problems encountered with targeted mutagenesis, several types of vectors referred to as "gene traps" have been developed that insert a promoterless reporter gene into mostly random chromosomal sites, including transcriptionally active regions (reviewed in Hill and Wurst, Mutagenic strategies to dissect mammalian development Curr. Topics. Dev,. Biol. *28*, 181-206 (1993); Gossler and Zachgo, Gene and enhancer trap screens in ES cell chimaeras, in: Gene targeting – a practical approach (A.L. Joyner, Ed.), Oxford University Press, Oxford (1993)). Integrations into the exons (exon traps) or

introns (intron traps) of a cellular gene dirsupt the gene and generally lead to its functional inactivation. Moreover, the gene trap provides a molecular tag to clone any gene linked to a specific function. Thus, unlike homologous recombination, gene trapping is not restricted to genes for which cloned sequences are available and can be used to disrupt any gene that resides in the genome. In most cases, gene traps are transduced as retroviruses, although electroporated DNA is also used. Retroviruses have the advantage that they integrate by a precise mechanism into the genome and cause no damage to the neighbouring DNA. Moreover, unlike electroporated DNA, retroviruses almost never integrate in tandem. Since gene trap integrations are distributed randomly throughout the genome, it is possible to create mutations in a large number of genes within a a limited number of experiments (Zambrowicz et al., Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells, Nature *392*, 608-611 (1998)). In principle, this enables the mutational analysis of all genes within a genome by simply generating an integration library in which every single gene is disrupted by a gene trap. In most cases murine embryonic stem (ES) cells are used as targets for gene trap integrations since mutant clones can be easily obtained in vitro and passaged to transgenic offspring in vivo (Gossler et al., Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes, Science, 244, 463 - 465 (1989); Friedrich and Soriano, Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmentally genes in mice, Genes & Development, 5, 1513 - 1523 (1991); Skarnes et al., A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice, Genes Dev. *6*, 903-918 (1992); von Melchner et al., Selective disruption of genes expressed in totipotent embryonal stem cells, Genes & Dev *6*, 919-927 (1992)). Mouse mutants obtained with this technology often develop a "loss of function phenotype" which sometimes will disclose the biological significance of a particular gene (Evans et al., Gene trapping and functional genomics, Trends in Genetics, 13, 370 - 374 (1997)). However, all the above gene traps induce an irreversible inactivation of their target. Consequently, a mouse mutant generated from an ES cell clone, is not different from a knock-out mouse obtained by conventional gene targetting.

Recently WO 99/50426 described conditional mutagenesis by utilizing a gene trap cassette capable of producing mutations (genome rearrangements) that can be switched on and off temporarily. These gene trap cassettes comprise suitably arranged frt or loxP recombinase sites, which – when exposed to Flp or Cre,

respectively - are removing or „flipping" (i.e., inverting) the gene trap cassette and therewith cause the genome rearrangements. However, conventional site specific recombinases, as Flp and Cre utilized in the above documents, normally recombine reversibly between identical recognition sites (like two loxP or FRT sites). In a conditional gene trap setting where the recombinase effects a „flipping" of the gene trap cassette as described in the above documents this would mean that equal amounts of sense and antisense products would be generated whereby one half of the target alleles carrying the gene trap would be inactivated whereas the other half would still have a functional configuration. Since most genes of the mammalian genome are recessive, inactivation of one allele is normally insufficient to produce an observable phenotype. Therefore, it is important to shift the equilibrium of the recombinase reaction towards the gene trap inversion that causes gene inactivation.

With regard to the above mentioned Cre/loxP recombination system, it is known that the recombination between mutant loxP recognition sites (e.g., single mutant recognition sites lox66 and lox71; Albert et al., The Plant Journal, 7, 649 - 659 (1995)) generates a double mutant- and a wildtype- loxP site (see SEQ ID NOs:4 and 5), each on one side of the inverted DNA. Since the latter combination of loxP sites is less efficiently recognised by the Cre-recombinase, the inversion is mostly unidirectional. The above lox66/lox71 system is utilized by Araki et al., Cell. Mol. Biol. 45(5), 737-750 (July 1999) in a two step gene trap system. In a first step a gene trap vector (having a functional gene flanked by lox sites in the same direction) is integrated into an endogenous gene. In a second step the functional gene is replaces with a second functional gene by Cre-mediated recombination/integration.

In addition to the above Cre/mutant loxP system it was assumed in the references below that recombinases which normally recognize nonidentical RRSs, such as the phage derived integrases of the integrase (Nunes-Düby et al., Nucleic Acid Res. 26, 391 - 406 (1998)) or resolvase (Thorpe et al., Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)) family of recombinases, may be used for unidirectional inversion.

Among the phage derived integrases, the phage λ-encoded integrase (λ-Int) is the most extensively characterised system (for reviews, see: Craig, Annu. Rev. Genet., 22, 77-105 (1988); Landy, Annu. Rev. Biochem, 58, 913-949 (1989); Landy, Curr. Op. Genet. Dev., 3, 699-707 (1993); Sadowsky, FASEB J., 7, 9-17

(1993); Hallet and Sherratt, FEMS Microb. Rev., 21, 157-178 (1997)). Integrative recombination occurs between an attP attachement site present in the phage genome and an attB attachement site in the bacterial genome. If both sites are placed on the same DNA molecule in the same orientation as used for integration, the intervening DNA segement is deleted from the parental molecule. In contrast to e.g. the loxP/Cre system, attP and attB sites are nonidentical but share some sequence identity in the λ-Int binding core region. A varying number of recognition sequences flanking the core region serves as binding sites for the accessory proteins IHF, Fis or Xis. Upon integrative recombination between an attP and attB site two new sites, termed attR and attL are generated. Since these sites are only recombined again back to an attP/attB pair for excisive recombination if the Xis protein is present, the λ-Int system is unidirectional in its absence.

That the λ-Int system requires the IHF accessory protein also for integrative recombination complicates its biotechnological application in mammalian cells; however, the use of IHF independent mutant λ-Int forms has been recently described in human cells (Lorbach et al., J. Mol. Biol. 296, 1175-1181 (2000)).

In contrast to λ-Int, the Integrase of the phage ΦC31 (hereinafter C31-Int) has been characterised to recombine a pair of the respective attP/attB sites without the need for accessory proteins (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)). C31-Int AttP/attB recombination leads to an integrative reaction if the two sites are placed on different molecules or to a deletion of the intervening DNA if the sites are present on the same DNA molecule in the same orientation as used for integration. Since the product of the recombination, a pair of attL/attR sites, is not recombined any more by C31-Int alone (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)), the reaction is unidirectional. C31-Int has been used in human cells for the integration and deletion of DNA molecules in human cells (Groth et al., Proc. Natl. Acad. Sci. 97, 5995-6000 (2000)).

Summary of the Invention

The object to be solved by the invention of the present application is the provision of a gene trap system that does not show the above drawbacks and

allow selective, unidirectional genome rearrangements. It was found that the above object can be solved by a gene trap construct using specific recombinase/recombinase recognition site systems (e.g., systems comprising mutant recombinase recognition sites) which allow an unidirectional inversion, preferably such systems recombine (i.e., invert) only once.

The present invention combines conditional gene targeting with gene trap mutagenesis. A conditional gene trap vector typically integrates into an intron without interfering with gene function. However, activation of the gene trap by a site specific recombinase induces gene disruption and loss of its function. Unlike conditional gene targeting, the strategy is not limited to known genes and can be used to disrupt all genes in a temporally and/or spatially restricted manner.

The present invention thus provides:

(1) a gene trapping construct capable of causing conditional mutations in genes which comprises a functional DNA segment inserted in a functional or non-functional manner, in sense or antisense direction relative to the gene to be trapped, said functional DNA segment being flanked by two recombinase recognition sequences (hereinafter RRSs) which are specific to a site specific recombinase capable of unidirectional inversion of a double stranded DNA segment;

(2) a preferred embodiment of the gene trapping construct defined in (1) above, which comprises two functional DNA segments,

(a) a first DNA segment (disruption cassette) being inserted in antisense orientation relative to the transcriptional orientation of the gene to be trapped and being flanked by two RRSs which are specific to a site specific recombinase capable of inverting a double stranded DNA segment, and

(b) a second segment (selection cassette) being positioned in sense direction relative to the transcriptional orientation of the gene to be trapped and being flanked by two RRSs in the same orientation;

(3) a cell comprising the gene trapping construct as defined in (1) or (2) above;

(4) the use of the cell of (3) above for the identification and/or isolation of genes;

(5) a process for the generation of conditional mutations in all genes of an organism comprising

(i) installation of a gene trapping construct as defined in (2) above in a suitable cell,

(ii) selection of cells in which the construct is incorporated in a gene,

(iii) identification and/or isolation of the gene in which the construct is incorporated,

(iv) deletion of the selection cassette from the trapped gene,

(v) induction of a mutation in the trapped gene by inversion of the functional DNA segment;

(6) a transgenic organism obtainable by the method of (5) above; and

(7) the use of the transgenic organism according to (6) above to study gene function at various developmental stages.

Short Description of Figures

Fig. 1: Conditional gene trap vector to modify genes expressed in the target cells.

Fig. 2: Conditional retroviral gene trap vector to modify expressed genes.

Fig. 3: Conditional gene trap vector to modifiy genes independently of their expression level in the target cells.

Fig. 4: Conditional retroviral gene trap vector to modify genes independently of their expression.

Fig. 5 shows the conditional gene trap vector pRK57SA-β.

Fig. 6 shows a Southern blot analysis of genomic DNA from the gene trap ES clone GT2.

Fig. 7 shows the test vector for C31-Int mediated inversion, pRK73, and the product of inversion, pRK73-Inv.

Fig. 8 shows the structure of the C31-Int expression vector pRK65.

Fig. 9 shows the PCR products generated with the primers P64-1 and P64-3 and a sequence comparison of the PCR products.

Fig. 10 shows the recombination substrate vectors used for Cre mediated inversion.

Detailed Description of the Invention

The gene trapping constructs (1) and (2) are hereinafter described in more detail:

The present invention provides a site specific recombination system which enables an unidirectional inversion of sequences flanked by RRS1 sites. For example, in the Cre/loxP recombination system, it is known that the recombination between mutant loxP recognition sites (e.g., single mutant recognition sites lox66 and lox71; Albert et al., The Plant Journal, 7, 649 - 659 (1995); see SEQ ID NOs:2 and 3) generates a double mutant- and a wildtype-loxP site (see SEQ ID NOs:4 and 5), each on one side of the inverted DNA. Since this combination of loxP sites is less efficiently recognised by the Cre-recombinase, the inversion is "unidirectional".

A recombination system is "unidirectional" in accordance with the present invention if a reaction of a site specific recombinase corresponding to the RRSs within the construct produces the inverted double stranded DNA segment in an amount of greater 50%, preferably greater 75%, more preferably greater 90%, relative to the starting double stranded DNA segment.

Inversion of a functional DNA segment is achieved through flanking RRSs sites being oriented in opposite orientation. Suitable RRSs in accordance with the present invention include, but are not limited to, constructs wherein the two RRSs are selected from mutant loxP sites, mutant frt sites and AttP/AttB sites, and the site specific recombinase is Cre, Flp, ΦC31-Int or λ-Int, respectively. Preferred RRSs are constructs wherein the two RRSs are single mutant loxP sites, preferably a lox66 and a lox71 site (SEQ ID NOs:2 and 3) or an AttP and an AttB site of Φ C31 as shown in Fig. 7, and the site specific recombinase is Cre or Φ C31, respectively.

In a preferred embodiment of the invention the functional DNA segment of the contsruct (1) further comprises one or more of the following functional elements: splice acceptor, splice donor, internal ribosomal entry site (IRES),

polyadenylation sequence, a gene coding for a reporter protein, a resistance gene and a gene coding for a further site specific recombinase. Suitable splice aacceptors include, but are not limited to, the adenovirus type 2 splice acceptor of exxon 2 shown in SEQ ID NO:1; suitable donors include, but are not limited to, the adenovirus exxon 1 splice donor; suitable IRES include, but are not limited to, that of the ECM virus shown in SEQ ID NO:8; and suitable polyadenylation sequences are the polyadenylation site sequence of the bovine growth hormone (bpA) or that shown in SEQ ID NO:6.

Suitable reporter genes include, but are not limited to, E. coli ß-galactosidase (see SEQ ID NO:10), fire fly luciferase, fluorescent porteins (e.g., GFP) and human placental alkaline phosphatase (PLAP). Suitable resistance genes include, but are not limited to, neomicin phosphotransferase, purimicin (see SEQ ID NO:9) and hygromicin. Suitable further (second) site specific recombinase are all recombinases which do not interfere with the RRSs of the gene trapping construct.

In a further preferred embodiment of the present invention the construct (1) further comprises a selection DNA segment suitable for selecting for genes having an incorporated gene trapping construct, said selection DNA segment comprising a reporter or resistance gene and flanking recombinase recognition sites in same orientation. Suitable resistance genes are those mentioned above, provided, however, that they do not interfere with the resistance gene of the functional DNA segment. Suitable recombinase recognition sites in same orientation include, but are not limited to, loxP and mutants thereof (see SEQ ID NOs:5, 2 and 3), fRT and mutants thereof (see SEQ ID NO:7), provided, however, that these RRSs do not interfere with the RRSs of the functional segment.

The activation mechanism of a conditional gene trap is based on the ability of a site specific recombinase (e.g. Cre or Flp) to invert a double stranded DNA segment flanked by two RRSs (e.g. loxP or frt) inserted in opposite orientations (Abremski et al., Cell, 32, 1301 – 1311 (1983); Hamilton et al., J. Mol. Biol., 178, 481 – 486 (1984); Albert et al., The Plant Journal, 7, 649 - 659 (1995); Vetter et

al., Proc. Natl. Acad. Sci. USA, 80, 7284 – 7288 (1983)). If this is used in the context of a gene trap vector, sequence elements that interfere with gene expression (e.g. splice sites, polyadenylation signals) can be first inserted into an intron in an antisense direction relative to the gene. By residing on the non-transcribed DNA strand, these elements will not interfere with gene transcription. Whenever or wherever a gene disruption is desired, inversion is induced by expressing the appropriate site specific recombinase.

The preferred types of operation with respect to the invention are plasmid (Figures 1 and 3) or retroviral vectors (Figures 2 and 4) which can select for integrations into introns. Selection is based on the expression of a reporter gene that requires either an active cellular promoter (expressed genes) (Figures 1 and 2) or the acquisition of a endogenous transcriptional termination (polyadenylation) signal (Figures 3 and 4).

The principal elements of a conditional gene trap vector of embodiments (1) and (2) of the invention that selects for integrations into expressed genes are (i) a conditional gene disruption cassette, containing a 3′ splice site (splice acceptor; SA; see SEQ ID NO:1) and a polyadenylation sequence (polyA; see SEQ ID NO:6) flanked by two RRSs (RRS1) of a site specific recombinase (rec1) in opposite orientation, and (ii) a selection cassette containing a reporter or selectable marker gene flanked by an upstream SA- and a downstream polyA-site. The selection cassette is flanked by two RRSs (RRS2) in same orientation which are recognized by second recombinase (rec2) and is in opposite orientation to the gene disruption cassette (Figures 1 and 3). Selection for gene expression yields recombinants in which the reporter gene is fused to the regulatory elements of an endogenous gene. Transcripts generated by these fusions encode for a truncated cellular protein which has lost its normal function. Since selection for a gene trap event relies on the expression of the selection cassette which is by itself mutagenic, it needs to be removed to recreate gene function. This is achieved by expressing rec2 in recombinants selected for gene trap integrations. In a favoured operational process the conditional gene trap is transduced into ES cells. After selecting for integrations into the introns of expressed genes, rec2 is transiently expressed in individual clones to delete the selection cassette to

restore gene function. The resulting clones containing only the gene disruption cassette are used to create transgenic mouse strains. Such mouse strains are crossed to rec1 expressing mouse strains to obtain doubly transgenic offspring where the gene disruption cassette is reversed to its mutagenic (sense) orientation.

As a further preferred embodiment the invention provides a conditional gene trap vector that selects for integrations into all genes. This is achieved by replacing the elements of the original selection cassette by a reporter gene that is fused to an upstream constitutive promoter and to a downstream 5'splice site (splice donor) (Zambrowicz, et al., 1998) (Figures 3 and 4). Expression of this gene trap is dependent on the acquisition of an endogenous polyadenylation sequence which occurs by splicing of the selection cassette to the donwstream exons of the target gene. Since the process is driven by a constitutive promoter, selection for gene trap integrations is independent of target gene expression. As with the other conditional gene trap, a favoured operational process is its transduction into ES cells and the generation of mutant mouse strains.

Thus, as a preferred embodiment the present invention provides for a conditional gene trapping construct selecting for integrations into expressed genes. In said construct the functional DNA segment preferably comprises a splice acceptor and a polyadenylation sequence flanked by two RRSs of a first site specific recombinase capable of of unidirectional inversion of a double stranded DNA segment as defined above. Said construct further comprises a selection DNA segment (selection cassette) which comprises a reporter or selectable marker gene flanked by an upstream splice acceptor sequence and a downstream polyadenylation sequence, said selection cassette being flanked by two conventional RRSs of a second site specific recombinase in same orientation.
As a further preferred embodiment the present invention provides for a conditional gene trapping construct selecting for integrations into all genes. In said construct the functional DNA segment comprises a splice acceptor and a polyadenylation sequence flanked by two RRSs of a first site specific recombinase capable of of unidirectional inversion of a double stranded DNA segment as defined above. Said constuct further comprises a selection DNA segment

(selection cassette) which comprises a reporter or selectable marker gene fused to an upstream constitutive promoter and a downstream splice donor site, said selection cassette being flanked by two RRSs of a second site specific recombinase in same orientation.

In said preferred embodiments the RRSs of the second site specific recombinase may be any conventional RRS provided , however, that tey do not interfere with the RRSs of the first recombinase.

The cell of embodiment (3) of the invention can be prepared by conventional methods including electroporation or retroviral infection.

In a preferred embodiment of the process (5) of the invention the mutation in step (v) is effected by a site specific recombinase capable of unidiretional inversion of the functional DNA of the gene trapping construct. The process (5) is suitable for temporarily and/or spatially restricted inactivation of all genes that constitute a living organism and for preparing transgenic mammals, wherein in step (i) the gene trapping construct is installed in an ES cell.

The above process possesses the following advantages over current technology:

(i) mutations are inducible in prespecified cells and tissues and during prespecified time intervals;

(ii) mutations can be induced in all genes, including those for which cloned sequences are not available;

(i) the functional analysis of the mutant genes in appropriate organisms is relatively fast and cheap.

The appended figures further explain the present invention:

Figure 1 shows a favoured operational form of the process relating to a conditional gene trap transduced by electroporation to capture expressed genes. A: Target gene after insertion of a conditional gene trap vector into an intron of an expressed gene. In transcripts initiating at the target gene's promoter the LacZ- and the puromycin resistance genes (selection cassette) are fused to the

upstream exons by a splice acceptor site. The fusion disrupts and inactivates the target gene.

B: Modified allele after FLP mediated deletion of the selection cassette. Excision of sequences flanked by the frt recognition sites which include the splice acceptor signal restores wild type transcription and gene function.

C: Target gene after Cre mediated inversion of the loxP flanked gene trap elements (gene disruption cassette). Gene trap inversion positions the splice acceptor- and the poly A- sites into the sense orientation which is mutagenic.

SA: splice acceptor signal sequence; SD: splice donor signal sequence; pA: polyadenylation sequence; lacZ: beta-Galactosidase; puro: puromycin resistance gene; filled rectangles: exons; filled triangles: loxP recognition sites, loxP: wildtype loxP site, lox66: single mutant lox site, lox71: single mutant lox site; lox 66/71: double mutant lox site; shaded triangles: FRT recognition sites (frt); thick continuous lines: mRNA transcripts; thin continuous lines: intron; thin stippled lines: recombinase mediated deletion or inversion between lox or FRT sites. Cre: Cre recombinase; FLP: FLP recombinase. Ires: Internal ribosomal entry site.

Figure 2 shows a favoured operational form of the process relating to a conditional gene trap transduced by retroviruses to capture expressed genes.

A: Retroviral plasmid vector.

B. Target gene after insertion of a conditional gene trap provirus into an intron of an expressed gene. Virus replication and LTR mediated duplication places the gene disruption cassette in U3 between mutant loxP sites in both 5' and 3' proviral LTRs. Moreover, it places the selection cassette inserted into the body of the virus between two frt sites. Transcripts initiating in the cellular promoter are spliced to the selection cassette by an upstream splice acceptor site. This fusion disrupts and inactivates the target gene.

C: Modified allele after FLP mediated excision of all proviral sequences placed between the frt sites. This restores gene function and leaves only one copy of gene disruption cassette behind.

D: Target gene after Cre mediated inversion of the loxP flanked gene trap elements. LTR inversion positions the gene disruption cassette into a mutagenic sense orientation.

LTR = long terminal repeat, U3 = 3' unique region of the retroviral LTR, R = repetitive region of the retroviral LTR, U5 = 5' unique region of the retroviral LTR, SA = splice acceptor signal sequence; SD = splice donor sequence; pA = polyadenylation sequence; Puro = puromycin resistance gene; lx = loxP recognition sites; frt = Flp recognition sites; thin continuous lines = intron; thin stippled lines = sequences excised by mRNA processing; arrows indicate transcriptional orientation.

Figure 3 shows a favoured operational form of the process relating to a conditional gene trap transduced by electroporation to capture all genes.
A: Target gene after insertion of a conditional gene trap vector into an intron of a gene. Transcripts initiating in the independent pgk promoter express the puromycin resistance gene after splicing into downstream exons to acquire a polyadenylation site. Splicing is mediated by a splice donor sequence inserted downstream of pgkpuro. The fusion of the selection cassette to downstream exons inactivates the target gene.
B: Modified allele after FLP mediated deletion of the selection cassette which restores gene function.
C: Target gene after Cre mediated inversion of the gene disruption cassette which causes the mutation.
SA: splice acceptor signal sequence; SD: splice donor signal sequence; pA: polyadenylation sequence; lacZ: beta-Galactosidase; puro: puromycin resistance gene; filled rectangles: exons; filled triangles: loxP recognition sites, loxP: wildtype loxP site, lox66: single mutant lox site, lox71: single mutant lox site; lox 66/71: double mutant lox site; shaded triangles: FRT recognition sites (frt); thick continuous lines: mRNA transcripts; thin continuous lines: intron; thin stippled lines: recombinase mediated deletion or inversion between lox or FRT sites. Cre: Cre recombinase; FLP: FLP recombinase. Ires: Internal ribosomal entry site; pgk: phosphoglyceratekinase promoter.

Figure 4 shows a favoured operational form of the process relating to a conditional gene trap transduced by retroviruses to capture all genes.
A: Retroviral plasmid vector

B. Target gene after insertion of a conditional gene trap provirus into an intron of a gene.

Virus replication and LTR mediated duplication places the gene disruption cassette in U3 between mutant loxP sites in both 5' and 3' proviral LTRs. Moreover, it places the selection cassette inserted into the body of the virus between two frt sites. Transcripts initiating in the independent pgk promoter express the puromycin resistance gene after splicing into downstream exons to acquire a polyadenylation site. Splicing is mediated by a splice donor sequence inserted downstream of pgkpuro. The fusion of the selection cassette to the downstream exons inactivates the target gene.

C: Modified allele after FLP mediated excision of all proviral sequences placed between the frt sites. This restores gene function and leaves only one copy of gene disruption cassette behind.

D: Target gene after Cre mediated inversion of the loxP flanked gene trap elements. LTR inversion positions the gene disruption cassette into a mutagenic sense orientation.

LTR = long terminal repeat, U3 = 3' unique region of the retroviral LTR, R = repetitive region of the retroviral LTR, U5 = 5' unique region of the retroviral LTR, SA = splice acceptor signal sequence; SD = splice donor sequence; pA = polyadenylation sequence; Puro = puromycin resistance gene; lx = loxP recognition sites; frt = Flp recognition sites; thin continuous lines = intron; thin stippled lines = sequences excised by mRNA processing; pgk = phosphoglyceratekinase promoter; arrows indicate the direction of transcript elongation.

Figure 5 shows the conditional gene trap vector pRK57SA-β

A: Original configuration of the conditional gene trap vector after insertion into an intron of an expressed gene. In transcripts initiating at the target gene's promoter the β-geo gene is fused to the upstream exons by a splice acceptor site. The fusion disrupts and inactivates the target gene. The splice acceptor – β-geo cassette is flanked by two mutant lox sites in opposite orientation. The arrow indicates the trancriptional direction of the β-geo gene.

B: Conditional gene trap vector after Cre mediated inversion of the lox flanked gene trap elements (gene disruption cassette). Gene trap inversion positions the

splice acceptor- and the poly A- sites into the antisense orientation which is non-mutagenic. The recombination of lox66 and lox71 sites generates one loxP wildtype and one double mutant lox 66/71 site.

SA: splice acceptor signal sequence; pA: polyadenylation sequence; β-geo: fusion gene of β-galactosidase and neomycinphosphotransferase; filled rectangle: exon; triangles: lox recognition sites, loxP: wildtype loxP site, lox66: single mutant lox site, lox71: single mutant lox site; lox 66/71: double mutant lox site; thick continuous line: mRNA transcript; thin stippled lines: recombinase mediated inversion between lox sites. Cre: Cre recombinase.

Figure 6 shows a Southern blot analysis of genomic DNA from the gene trap ES clone GT2. Upon digestion with BamHI and XmnI, and a segment of the neomycin phosphotransferase gene as a probe, the integrated pRK57SA-β vector gives rise to a band of 3.06 kb, upon Cre mediated inversion the band is reduced to a size of 1.17 kb. Shown are samples of GT2 cells at 6 days after infection with the Cre expressing retrovirus pBABE-pgkCre, the control virus pBABE-Srf, or which were not treated with virus. The band indicating inversion of the gene trap vector appears only in the sample treated with the Cre expressing retrovirus.

Figure 7 shows the test vector for C31-Int mediated inversion, pRK73, and the product of inversion, pRK73-Inv.

A: Plasmid pRK73 contains the 1.1 kb cassette of the coding region of the puromycin resistance gene and a polyadenylation signal, which is flanked 5` by the 84 bp attB and 3´ by the 84 bp attP recognition site of C31-Int. These attB and attP sites are oriented in opposite to each other (thick black arrows), as compared to the orientation of attB and attP before the integration of the phiC31 phage into the bacterial genome. In addition, the cassette is flanked by two Cre recombinase recognition (loxP) sites in opposite orientation to each other. For better orientation the half sites of the att sequences are labelled by a direction (thin arrow) and numbered 1-4. The 3 bp sequence within the att sites at which recombination occurs is framed by a box. The positions at which the PCR primers P64-1 and P64-3 hybridise to the pRK73 vector are indicated by an arrow, pointing into the 3` direction of both oligonucleotides.

B: The inversion product pRK73-Inv expected from the C31-Int mediated recombination between the att sites of pRK73. The recombination between a pair of attB/attP sites generates a pair of attR/attL sites flanking the inverted puromycin cassette. In this configuration the primers P64-1 and P64-3 will amplify a PCR product of 608 bp.

Figure 8 shows the structure of the C31-Int expression vector pRK65.
The 1.85 kb C31-Int coding region is expressed by the CMV promoter for expression in mammalian cells. A 300 bp hybrid intron between promoter and coding region provides a splice donor and acceptor site; the coding region is followed by a synthetic polyadenylation signal sequence. The expected transcript is depicted as a thin line.

Figure 9 shows the PCR products generated with the primers P64-1 and P64-3 and a sequence comparison of the PCR products.
A: Analysis of PCR products on an agarose gel from PCR reactions using the Primers P64-1 and P64-3 on DNA extracted from MEF5-5 cells transfected 2 days before with plasmid pRK73 alone (lane 4), with pRK73 + CMV-Cre (lane 3), with pRK73 + pRK65 (lane 2), and from nontransfected cells (lane 1). The products with an apparent size around 650 bp, as compared to the size marker used, from lane 2 were excised from the agarose gel and purified. The PCR products were cloned into a sequencing plasmid vector and gave rise to the plasmids pRK80b and pRK80c. The insert of these plasmids was sequenced using reverse primer (seq80b and seq80c) and compared to the predicted sequence of the pRK73 vector after C31-Int mediated inversion of the att flanked cassette, pRK73-Inv. Both PCR products show a 100% identity with the predicted attR sequence after inversion. The generated attR site is shown in a box, with the same sequence designation used in Figure 7B. The nucleotide positions of the compared sequences pRK73, seq80b and seq80c are indicated.

Figure 10 shows the recombination substrate vectors used for Cre mediated inversion.
Plasmid pRK74 contains a CMV promoter for expression in mammalian cells followed by cassette containing the E. coli β-galactosidase gene and a

polyadenylation sequence in inverse orientation to the CMV promoter. This cassette is flanked by the lox66 and lox71 single mutant lox sites (Albert et al., The Plant Journal, 7, 649 - 659 (1995)) in opposite orientation. Plasmid pRK76 is structured like pRK74 except that the inverse β-galactosidase cassette is flanked by a lox66/71 double mutant lox site and a wildtype loxP site. The straight arrow indicates the transcriptional orientation of the β-galactosidase expression cassette.

The present invention is further illustrated by the following Examples which are, however not to be construed as to limit the invention.

Examples

Example 1

A plasmid vector containing a splice acceptor sequence followed by a fusion gene of ß-galactosidase and neomycin phosphotransferase (SA-ß-geo), flanked by two different mutant lox sites in opposite orientation, was constructed and used as a gene trap in ES cells. Cell lines exhibiting ß-galactosidase activity were infected with a retroviral vector for expression of Cre to investigate the inactivation of the gene trap vector by recombinase mediated inversion of the lox flanked DNA segment.

A. Construction of the gene trap vector pRK57SA-ß: As the first step of vector construction, a synthetic 74 bp DNA fragment with protruding XbaI ends containing a lox66 and a lox71 Cre recombinase recognition site in opposite orientation (Albert et al., The Plant Journal, 7, 649 - 659 (1995)), separated by a Xho site, was ligated into the XbaI site of the plasmid pNEB193 (New England Biolabs) resulting in the vector pRK57. The synthetic DNA fragment was obtained by annealing of the two synthetic oligonucleotides lox3 (5`-CTAGATAACT TCGTATAGCA TACATTATAC GAACGGTACT CGAGATAACT TCGTATAATG TATGCTATAC GAACGGTA-3`; SEQ ID NO:11) and lox4 (5`-CTAGATACCG TTCGTATAGC ATACATTATA CGAAGTTATC TCGAGTACCG TTCGTATAAT

GTATGCTATA CGAAGTTAT-3`; SEQ ID NO:12). Next, plasmid pRK57 was cut with the restriction enzyme XhoI, dephosphorylated with calf intestinal phosphatase and the 5' protruding ends were filled in with Desoxynucleotide-triphosphates using the Klenow fragment of E.coli DNA polymerase I. This vector was ligated to a 4.3 kb DNA fragment isolated from the plasmid ROSAbetageo (Friedrich et al., Genes&Development 5, 1513-1523 (1991)), which was obtained by digestion with the restriction enzyme XhoI, and filling in the 5' protruding ends as described. The ligation product, vector pRK57SA-ß (SEQ ID NO:13), contains a cassette with the adenovirus type 2 splice acceptor from exon2 of the major late region (SA, pos. 525 to 628; SEQ ID NO:1), a fusion gene of E.coli ß-galactosidase and neomycin-phosphotransferase (ß-geo, pos. 667 to 4548) and the transcription termination and polyadenylation signal from the bovine growth hormone gene (pA, pos. 4561 to 4843). This SA-ß-geo cassette is flanked by one lox66 site 5' to the splice acceptor (pos. 446 to 479) and one lox71 site 3' to the bovine growth hormone gene sequences (pos. 4869 to 4902). The two lox sites are oriented in opposite orientation to each other. The vector pRK57SA-ß contains BamHI sites at the positions 427, 503, 673 and 3723, a single XmnI site at position 6788, and a single KpnI site at position 413.

B. Construction of retroviral vectors: A 1.1.kb NheI fragment carrying the coding sequence for Cre-recombinase (Cre) with an N-terminal nuclear localization signal derived from the simian virus 40 large T-antigen was inserted into the plasmid pBluescript II KS(+) (Stratagene), opened with the restriction enzyme XbaI and dephosphorylated with calf intestinal phosphatase. The resulting plasmid was cut with the restriction enzyme SpeI, dephosphorylated with calf intestinal phosphatase and the 5' protruding ends were filled in with Desoxynucleotide-triphosphates using the Klenow fragment of E.coli DNA polymerase I. This vector was ligated to a 0.6 kb blunt-ended fragment containing the promoter region of the mouse phosphoglycerate kinase (PGK) 1a gene. The ligation product contains a cassette with the PGK promoter and the cre coding sequence in the backbone of pBluescript II KS(+). This cassette was excised as a 1.6 kb NotI fragment, filled in as described and inserted into the filled in SnaBI site of a derivative of the mouse moloney leukemia virus based vector pBABEpuro (Morgenstern and Land, Nucleic Acids Res.18, 3587-3596

(1990)) creating the retroviral cre expression vector pBABE-pgkCre (SEQ ID NO:14). A control vector pBABE-Srf (SEQ ID NO:15) was produced, which contains a short polylinker region instead of the pgkCre cassette. The plasmid forms of these retroviral vectors contain the following functional elements in a plasmid backbone identical to pBABEpuro:

pBABE-pgkCre:	position
partial MMLV U3	8-335
MMLV R	336-402
MMLV U5	403-480
MMLV primer binding site and extended packaging signal (incl. mutated splice donor and 5' end of gag coding sequence with mutated ATG start codon)	481-1374
mouse PGK promoter	1417-1921
cre coding sequence	1972-3024
promoter/enhancer deleted MMLV U3	3088-3168
MMLV R	3187-3253
MMLV U5	3254-3332

pBABE-Srf:	
partial MMLV U3	8-335
MMLV R	336-402
MMLV U5	403-480
MMLV primer binding site and extended packaging signal (incl. mutated splice donor and 5' end of gag coding sequence with mutated ATG start codon)	481-1374
promoter/enhancer deleted MMLV U3	1451-1531
MMLV R	1549-1616
MMLV U5	1617-1695
MMLV U5 cw	1617-1695

Infectious retroviral particles were produced by transfecting $2x10^7$ Phoenix eco packaging cells (http://www.stanford.edu/group/nolan/phoenix_info.html) with a total of 100 µg of the plasmid forms of the retroviral vectors using the calcium-phosphate coprecipitation technique. Two days after transfection the tissue

culture supernatants were harvested, filtered through a 0.45 µm filter and mixed with polybrene to a final concentration of 4 µg/ml. This virus supernatant was then used to infect ES cell clones.

To determine the titer of the pBABE-pgkCre virus serial dilutions of the virus containing supernatant were used to infect 10000 NIH-3T3 cells carrying the plasmid pSVpaX1 (Buchholz et al., Nucleic Acids Res. 24, 4256-4262 (1996)) as a stable integrate. Successful infection with the Cre-expression virus leads to excision of the pac/SV40polyA-cassette from the pSVpaX1 construct and concomitantly to expression of the beta galactosidase gene. Expression of this marker was monitored two days after infection by fixing the cells and staining with the chromogenic substrate X-Gal (5-Bromo-4-chloro-3-indolyl-beta-D-galactopyranoside). The number of beta-galactosidase positive cells was used to estimate the titer of the original virus suspensions used for infection. The titer of the pBABE-pgkCre virus supernatant was determined to contain 5000 infectious particles per milliliter of supernatant.

C. Establishment of ß-galactosidase expressing ES cell clones: 100 µg of vector pRK57SA-ß were linearised with KpnI and electroporated into 3 x 10^7 cells of the embryonic stem (ES) cell E14, cultured as described in Kühn et al., Science 254, 707-710 (1991). Two days after electroporation the cells were grown for 7 days in medium containing 275 mg/ml of G418 (Gibco BRL) to select for clones which integrated the gene trap vector into an expressed gene. The concomitant expression of the ß-geo fusion gene allows such clones to grow in G418 containing medium due to the expression of the neomycin phosphotransferase. At day 9 after electroporation 60 G418 resistant clones were obtained and further expanded. Aliquots of all clones were tested for expression of ß-galactosidase activity by X-Gal staining, resulting in 23 positive clones showing 1% - 75% positive cells. These clones were analysed further by Southern blot hybridisation for the number of integrated vector copies. 11 of the 23 clones proved to obtain multicopy integrates and were excluded from further analysis. The two single copy integrants, GT2 and GT6, which showed the highest proportion of ß-galactosidase positive cells were chosen for further analysis.

D. Infection of ES cell clones with retroviral vectors: To test for the induced inversion of the SA-ß-geo cassette within the integrated pRK57SA-ß vector in the clones GT2 and GT6, cells of both clones were infected either with the Cre recombinase expressing retrovirus pBABE-pgkCre or with the control retrovirus pBABE-Srf lacking the Cre gene. For the later analysis of ß-galactosidase expression 5000 cells/well of clones GT2 and GT6 were plated into wells of 48-well culture plates (Falcon) one day before virus infection. For infection the culture medium was replaced with 300 ml of virus containing supernatants (containing about 1500 infectious particles; i.e. an infection rate of 30%) and the cells were further cultured for 4 hours. After 4 hours the wells were filled up with 300 ml culture medium and incubated overnight. At the following day the supernatants were completely replaced with fresh culture medium and grown for further 2 or 6 days until the analysis of ß-galactosidase activity. For the later analysis of the genomic DNA of clone GT2, cells were essentially treated as described above, except that 30000 cells were plated into 24-well culture plates and treated with 600 ml of virus containing supernatants (containing about 3000 infectious particles; i.e. an infection rate of 10%). Genomic DNA was prepared from these samples at day 6 post infection.

E. Histochemical detection of ß-galactosidase activity: To quantitate ß-galactosidase expression the ES cells were plated one day before analysis at low density to be able to easily observe single cells. At the following day the culture medium was removed from the wells, the wells were washed once with phosphate buffered saline (PBS), and the cells were fixed for 5 minutes at room temperature in a solution of 2% formaldehyde and 1% glutaraldehyde in PBS. Next, the cells were washed twice with PBS and finally incubated in staining solution for 24 hours at 37°C (staining solution: 5 mM K3(Fe(CN)6), 5mM K4(Fe(CN)6), 2mM MgCl2, 1mg/ml X-Gal (BioMol) in PBS). Blue stained, ß-galactosidase positive cells were detected and distinguished from negative (transparent) cells in a cell culture binocular microscope under 200x magnification. For each determination a minimum of 200 cells was counted.

F. Southern blot hybridization: Genomic DNA was prepared from retrovirus infected ES cell clones following standard methods (Sambrook, Fritsch and

Maniatis, Molecular Cloning. A laboratory manual. ColdSpring Harbour Laboratory Press (1989)). Approximately 5 mg of purified DNA were digested with the restriction enzymes BamHI and XmnI to detect the inversion of the pRK57SA-ß genetrap vector. Digested DNA was separated in a 0.8% agarose® gel and transferred to nylon membranes (GeneScreen Plus, NEN DuPont) under alkaline conditions for 16 hours. The filter was dried and hybridised for 16 hours at 65°C with a probe representing the 3`part of the neomycin phosphotransferase gene (~600 bp PstI - XbaI fragment of plasmid pgk-neo (Soriano et al., Cell 64,693-702 (1991)). The probe was radiolabeled with P32-marked α-dCTP (Amersham) using the Megaprime Kit (Amersham). Hybridisation was performed in a buffer consisting of 10% Dextranesulfate, 1% SDS, 50mM Tris and 100mM NaCL, pH7.5). After hybridisation the filter was washed with 2x SSC and exposed to BioMax MS1 X-ray films (Kodak) at – 80°C for 3 days.

G. Results: The plasmid vector pRK57SA-ß which serves as a conditional gene trap was constructed by placing a cassette, consisting of a splice acceptor sequence followed by the ß-geo gene (SA-ß-geo), between two mutant loxP sites, lox 66 and lox71. These recognition sites of Cre recombinase are in opposite orientation to each other and allow to invert the mutagenic splice acceptor-ß-geo cassette within the vector upon expression of Cre (Fig.5). Upon recombination of a lox66 and a lox71 site one wildtype loxP and one double mutant lox66/71 site is generated (Albert et al., The Plant Journal, 7, 649 - 659 (1995)). As the combination of loxP and lox66/71 sites is much less efficiently recombined by Cre than a pair of lox66 and lox71 sites the outcome of the inversion reaction is shifted towards the reaction products, representing 80-90% of all recombined molecules (Albert et al., The Plant Journal, 7, 649 - 659 (1995)). In contrast, the use of two oppositely oriented wildtype loxP sites for Cre mediated inversion would lead to an equilibrium with a 50% proportion for each of the two inversion products (Albert et al., The Plant Journal, 7, 649 - 659 (1995)).

The vector pRK57SA-ß was linearised 5`to the lox66 site and introduced into murine embryonic stem (ES) cells by electroporation. Stable transfected clones which integrated the conditional gene trap vector into genes which are actively

transcribed in ES cells were obtained by selection in G418 containing cell culture medium. Surviving clones need to express the promoterless ß-geo fusion gene through a fusion transcript between the splice donor sequence of a mRNA of an endogenous gene and the splice acceptor sequence of the gene trap vector located in an intron.

After the establishment of G418 resistant clones the expression of ß-galactosidase activity was tested by histochemical staining. As expected only a fraction of the G418 resistant clones exhibited some ß-galactosidase activity, varying between 1%-75% of the cells of individual clones. In addition, these clones were analysed for the copy number of the integrated conditional gene trap vector by Southern blot hybridisation. Two ES cell clones, GT2 and GT6, which showed a single copy integration of the vector and a high proportion of ß-galactosidase positive cells were choosen for further analysis.

These clones were either infected with the retroviral vector pBABE-pgkCre, designed for expression of Cre recombinase, or with the retroviral control vector pBABE-Srf which lacks the Cre expression cassette. The inversion of the lox-flanked SA-ß-geo cassette was tested two and six days after infection by histochemical X-Gal staining for ß-galactosidase activity and by Southern blot analysis of genomic DNA. As shown below in Table 1 the proportion of ß-galactosidase expressing cells, exhibiting blue staining, among both ES cell clones was significantly reduced by 14 - 27% at both time points in samples infected with the Cre expression retrovirus as compared to samples treated with control virus.

Table 1:

	2 days post infection			6 days post infection		
	pBABE-Srf % blue cells	pBABE-pgkCre % blue cells	decrease	pBABE-Srf % blue cells	pBABE-pgkCre % blue cells	% decrease
Clone GT2	73%	59%	14%	76%	49%	27%
Clone GT6	67%	48%	19%	67%	47%	20%

Since of the clone GT2 only 76% and in clone GT6 only 67% of the cells expressed ß-galactosidase in the control sample, the reduction of blue stained cells by 27% in clone GT2 and by 20% in clone GT6 at day 6 after infection with the Cre expressing virus would correspond to an inversion rate of 36% in clone GT2 and of 30% in clone GT6. These numbers are in close correspondence to the calculated infection rate of 30% of the cells with the pBABE-pgkCre virus supernatant under the conditions used (see above); this correspondence suggests that most of the cells infected with the Cre virus underwent an (almost unidirectional) inversion of the conditional gene trap vector.

The observed increase in non-stained, ß-galactosidase negative cells is the expected outcome of Cre mediated inversion of the lox flanked SA-ß-geo cassette within its genomic integration site: the SA-ß-geo cassette in the clones GT2 and GT6 is initially located in the same transcriptional orientation as the endogenous gene leading to the expression of ß-galactosidase from a fusion transcript. Upon inversion of this cassette the transcript of the endogenous gene is not any more interrupted by splicing to the SA-ß-geo cassette which, in the inverse orientation, is not any more translated into ß-galactosidase protein.

In addition, the inversion reaction of the SA-ß-geo cassette upon treatment with Cre expressing retrovirus was directly tested by Southern blot analysis of genomic DNA of the clone GT2. Upon digestion of genomic DNA with BamHI and XmnI the gene trap vector gives rise to a band of 3.06 kb in its original configuration using a segment of the ß-geo gene as a probe. Upon inversion of the lox flanked SA-ß-geo cassette the fragment detected by the probe is reduced to 1.17 kb (Fig.5). As shown in Figure 6 the expected band demonstrating the recombinase mediated inversion reaction is specifically detected in the sample of the gene trap ES clone GT2 treated with Cre expressing retrovirus but not in the controls. The relatively weak strength of the band indicating inversion corresponds to the inefficient infection of the GT2 cells with Cre virus (10% infection rate) under the conditions used for DNA preparation.

We conclude from this experiment that it is possible by the use of mutant lox sites to invert the DNA segment of a gene trap vector, which is initially mutagenic for the trapped endogenous gene, preferentially (>50%) into a configuration which is not any more mutagenic and allow its wildtype expression.

Likewise, if the Sa-ß-geo cassette would be initially integrated into an endogenous gene in its non-mutagenic orientation, Cre mediated inversion would allow to switch its orientation preferentially to the mutagenic form.

Example 2

A plasmid vector was constructed which allows to test for the inversion of a DNA segment flanked by the attB and attP recognition sequences of the phiC31 phage derived integrase (C31-Int). This vector was transfected together with an expression vector for phiC31 integrase into a murine cell line and recombinase mediated inversion was detected by a specific PCR reaction. The amplified PCR product was cloned and its sequence determined. The obtained sequence confirms that integrase mediated inversion of the test vector occurs in a mammalian cell line and that the recombination occurs at the known breakpoints within the attB and attP sites.

A. Plasmid constructions:

Construction of the recombination test vector pRK73 (SEQ ID NO:16): First, an attB site (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)), generated by the annealing of the two synthetic oligonucleotides C31-4 (5`-CGTGA CGG TCT CGA AGC CGC GGT GCG GGT GCC AGG GCG TGC CCT TGG GCT CCC CGG GCG CGT ACT CCA CCT CAC CCA TCT GGT CCA; SEQ ID NO:17) and C31-5 (5`-CG TGG ACC AGA TGG GTG AGG TGG AGT ACG CGC CCG GGG AGC CCA AGG GCA CGC CCT GGC CCA CGC ACC GCG GCT TCG AGA CCG TCA; SEQ ID NO:18) was ligated into the BstBI restriction site of the vector PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)), 5`of its puromycin resistance gene and loxP site, giving rise to plasmid pRK52. The sequence and orientation of the cloned attB site was confirmed by DNA sequence analysis. Next, an attP site, generated by the annealing of the two synthetic oligonucleotides C31-6-2 (GATCCAGAAG CGGTTTTCGG GAGTAGTGCC CCAACTGGGG TAACCTTTGAG TTCTCTCAGTT GGGGGCGTAG GGTCGCCGAC ATGACACG; SEQ ID NO:19) and C31-7-2 (GATCCGTGTC ATGTCGGCGA CCCTACGCCC CCAACTGAGA GAACTCAAAG GTTACCCCAG TTGGGGCACT ACTCCCGAAA ACCGCTTCTG; SEQ ID NO:20), was ligated into the BamHI

restriction site of the plasmid pRK52, downstream of the puromycin resistance gene and loxP site, giving rise to plasmid pRK64-dir. In this plasmid the newly cloned attB and attP sites are in the same orientation flanking the puromycin resistance gene. Next, a 260 bp HindII + NruI fragment, containing the attP and a loxP site, was isolated from pRK64-dir, and ligated into plasmid pRK52, which was opened before with BamHI + NruI (the protruding BamHI end was filled with Klenow enzyme and nucleotides) such that the fragment containing the downstream loxP site was removed. The plasmid generated, pRK73, bears a puromycin resistance gene flanked by an attB and attP site in opposite orientation, as well as a pairs of loxP sites in opposite orientation (Fig7A). The sequence and orientation of the newly cloned attP site in pRK73 was confirmed by DNA sequence analysis.

Construction of the C31-Int expression vector pRK65 (SEQ ID NO:21): First, the C31-Int gene of phage phiC31 was amplified by PCR from phage DNA (# DSM 49156; German collection of microorganisms and cell cultures, Mascheroder Weg 16, 38124 Braunschweig, Germany) using the primers PC31-1 (5`-ATAAGAAT GCGGCCGC CCGAT ATG ACA CAA GGG GTT GTG ACC GGG-3`; SEQ ID NO:22) and PC31-3 (5`-ATAAGAAT GCGGCCGC ATC CGC CGC TAC GTC TTC CGT GCC-3`; SEQ ID NO:23). The ends of the PCR product were digested with NotI and the product was ligated into plasmid pBluescript II KS, opened with NotI, giving rise to plasmid pRK40. The DNA sequence of the 1.85 kb insert was determined and found to be identical to the published C31-Int gene (Kuhstoss and Rao, J. Mol. Biol. 222, 897-908 (1991)), except for an error in the stop codon. This error was repaired by PCR amplification of 300 bp fragment from plasmid pRK40 using the primers PC31-8 (5`-CCCGTTGGCA GGAAGCACTT CCGG-3`; SEQ ID NO:24) and PC31-9 (5`- GGATCCTCGA GCCGCGGGCG GCCGCCTACG CCGCTACGTC TTCCGTGCCG TCCTG-3`; SEQ ID NO:25), which provide a corrected Stop codon. The ends of this PCR fragment were digested with Eco47III and XhoI and the fragment ligated into plasmid pRK40, opened with Eco47III and XhoI to remove the defetive stop codon. The resulting plasmid, pRK55, contains the correct C31-Int gene as confirmed by DNA sequence analysis. The gene was isolated from pRK55 as 1.85 kb fragment by digestion with NotI and XhoI and ligated into the plasmid pRK50, opened with NotI and XhoI, giving rise to the C31-Int expression

vector pRK65. PRK65 contains a 700 bp Cytomegalovirus immediated early gene promoter (position 1 - 700) and a 270 bp hybrid intron (position 700 - 970) upstream of the NotI site and a 189 bp synthetic polyadenylation sequence (position 2831 - 3020) downstream of the XhoI site (all elements derived from pRK50). The C31-Int gene occupies position 978 - 2819 of pRK65. To generate the Cre expression plasmid CMV-Cre, the coding sequence of Cre recombinase was cloned into the NotI and XhoI sites of plasmid pRK50.

B.Transfection of Cells and PCR amplification: MEF5-5 mouse fibroblasts (Schwenk et al., Nucleic Acids Research, 26, 1427-1432 (1998)) (20000 cells per well of a 12 well plate (Falcon)) were transfected with 0.5 µg pRK73 alone or together with 125 ng pRK65 or CMV-Cre using the FuGene transfection reagent (Roche Diagnostics). After 2 days DNA was extracted from these cells and used for PCR amplification with Primers P64-1 (5`-TCA GCA ACC AGG CTC CCC AGC AGG C-3`; SEQ ID NO:26) and P64-3 (5`-TAG AGG ATC ATA AAT CAG CCA TAC CAC-3`; SEQ ID NO:27) using the Expand High Fidelity PCR kit (Roche Diagnostics).
PCR products were separated on a 0.8% agarose® gel, extracted with QuiaEx® (Quiagen) and cloned into the pCR2.1 vector using the TA cloning kit (Invitrogen) resulting in plasmids pRK80b and pRK80c. The sequence of the inserts - determined using the reverse sequencing primer and standard methods (MWG Biotech) - are shown in SEQ ID NOs:28 and 29, respectively.

C. Results: As a test vector for C31-Int mediated DNA inversion the plasmid pRK73 was constructed as described in A. above. PRK73 contains a DNA segment of 1.1 kb size (the coding region of the puromycin resistance gene) which is flanked 5` by the 84 bp attB and 3´ by the 84 bp attP recognition site of C31-Int (Fig. 7A). These attB and attP sites are oriented in opposite to each other, allowing inversion of the flanked DNA segment (Fig.7B), as compared to the natural orientation of attB and attP in the phiC31 phage and the bacterial genome, which allows integration or deletion of the phage genome. As a control,

vector pRK73 contains in addition two Cre recombinase recognition (loxP) sites in opposite orientation next to the att sites.

The predicted product of C31-Int mediated inversion of pRK73 is shown in pRK73-inv (SEQ ID NO:30).

As the second component of the desired test system we constructed a mammalian expression vector for C31-Int, plasmid pRK65, which contains the CMV-IE-Promoter upstream of the C31-Int coding region followed by a synthetic polyadenylation signal (Fig.8).

The recombination substrate vector pRK73 was transfected into the murine fibroblast cell line MEF5-5 either alone, or together with the C31-Int expression vector pRK65, or together with an expression vector for Cre recombinase, CMV-Cre. Two days after transfection DNA was prepared from the three samples and tested for the occurrence of the expected Cre or C31-Int generated inversion product by a specific polymerase chain reaction (PCR). Using primers P64-1 and P64-3 for amplification a 608 bp product can be only obtained after recombinase mediated inversion between the lox or att sites (Fig.7B). This product represents the junction 5` of the puromycin resistance gene which is occupied by an attB site in the vector pRK73; after C31-Int mediated inversion this site would be recombined into an attR site.

As shown in Figure 9A such an amplification product was found only in the samples cotransfected with the Cre or C31-Int recombinase expression vectors. To prove that the PCR product found after cotransfection of plasmids pRK73 and pRK65 represents indeed the inversion product of C31-Int mediated recombination, this DNA fragment was cloned into a plasmid vector and its DNA sequence determined. Two independent clones, pRK80b and pRK80c, were analysed and showed exactly the sequence of an attR site as expected from C31-Int mediated recombination (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)), which leads to the inversion of the attB/attP flanked puromycin resistance gene (Fig.9B). Although not formally proven here, it follows from the knowledge on C31-Int mediated recombination (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)) that the junction located 3` of the inverted puromycin resistance gene represents an attL site.

In conclusion, this experiment demonstrates that it possible to invert in a mammalian cell line a DNA segment flanked by a pair of C31-Int attB/attP recognition sites, generating a pair of attR/attL sites. As the combination of attR/attL is not recombined by C31-Int alone (Thorpe and Smith, Proc. Natl. Acad. Sci. USA 95, 5505 - 5510 (1998)) the reaction described above represents a system for the unidirectional inversion of DNA segments.

Example 3

Plasmid vectors were constructed which allow to test for the efficiency of the Cre recombinase mediated inversion of a DNA segment flanked by different single mutant lox sites compared to inversion of the same DNA segment flanked by a wildtype loxP site and a double mutant lox site. These vectors were transfected together with a Cre expression vector into a murine cell line and the relative inversion efficiency was determined through the expression of ß-galactosidase as a reporter gene.

A. Vector construction:

Construction of pRK74 (SEQ ID NO:31): starting from plasmid pRK62, which is identical to plasmid pRK57 (see example 1), except that a pair of (irrelevant) FRT sites has been inserted into the PmeI site, a 4.4 kb XhoI – NotI (protruding ends filled up) cassette was ligated into the NotI site of pRK62 (ends filled up). The 4.4 kb cassette contains a 120 bp splice acceptor sequence from from exon2 of the major late region of adenovirus type 2, followed by the 550 bp internal ribosomal entry site of ECMV virus and the coding region of ß-galactosidase (AgeI-BamHI fragment of plasmid pCH110; Pharmacia) and a polyadenylation sequence from plasmid PSV-Pax1 (Buchholz et al., Nucleic Acids Res., 24, 4256-4262 (1996)). The resulting plasmid pRK72 contains the ß-galactosidase cassette between a pair of lox66 and lox71 sites in opposite orientation. Next, plasmid pRK72 was opened with PacI and ligated with a 850 bp PmeI-HindIII fragment (ends blunted) from plasmid pRK50 providing a CMV promoter and a splice donor site (Choi et al., Mol. Cell. Biol. 11, 3070-3074 (1991)), resulting in plasmid pRK74. To derive pRK76, plasmid pRK72 was opened with SalI (ends filled) and ligated with the 850 bp PmeI-HindIII fragment (ends blunted) from plasmid

pRK50, resulting in plasmid pRK75. PRK75 was co-transformed into E.coli DH5 cells with the Cre expression vector p705-Cre (Buchholz et al., Nucleic Acids Research, 24, 3118-3119 (1996)) to invert the ß-galactosidase cassette in bacteria. One of the inverted subclones was designated pRK76 (SEQ ID NO:32). To generate the Cre expression plasmid CMV-Cre, the coding sequence of Cre recombinase was cloned into the NotI and XhoI sites of plasmid pRK50 (SEQ ID NO:33).

B. Transfection and measurement of ß-galactosidase activity: MEF5-5 mouse fibroblasts (Schwenk et al., Nucleic Acids Research, 26, 1427-1432 (1998)) (10000 cells per well of a 24 well plate (Falcon)) were transfected with 1 µg pRK74 or pRK76 alone or together with 500 ng CMV-Cre plasmid using the FuGene transfection reagent (Roche Diagnostics). After 2 days the cells were lysed and the ß-galactosidase activities were determined with the ß-galactosidase reporter gene assay (Roche Diagnostics) according to the manufacturers guidelines using a Lumat LB 9507 luminometer (Berthold).

C. Results: The efficiency of Cre mediated inversion of a DNA segment flanked by the single mutant lox66 and lox71 sites was compared to the inversion of the same DNA segment flanked by a wildtype loxP site and a double mutant lox site using the recombination substrate vectors pRK74 and pRK76 (Fig.10). Plasmid pRK74 contains a CMV promoter for expression in mammalian cells followed by cassette containing the E. coli ß-galactosidase gene and a polyadenylation sequence in inverse orientation to the CMV promoter. This cassette is flanked by the lox66 and lox71 single mutant lox sites (Albert et al., The Plant Journal, 7, 649 - 659 (1995)) in opposite orientation. Upon expression of Cre recombinase this cassette can be inverted and allows the production of ß-galactosidase as a measure of the recombination efficiency. Plasmid pRK76 is structured like pRK74 except that the inverse ß-galactosidase cassette is flanked by a lox66/71 double mutant lox site and a wildtype loxP site. Since the lox66 and lox71 sites are recombined into a pair of lox66/71 (named lox72 in Albert et al., The Plant Journal, 7, 649 - 659 (1995)) and loxP sites (Albert et al., The Plant Journal, 7, 649 - 659 (1995)), the comparison of the ß-galactosidase activities generated

from plasmids pRK74 or pRK76 after co-expression of Cre recombinase allows to compare the relative efficiency at which a pair of lox66//lox71 or lox66/71//loxP sites are recombined.

PRK74 and pRK76 were transiently transfected into a murine fibroblast cell line either alone or with the Cre expression vector CMV-Cre. After two days the cells were lysed and the levels of ß-galactosidase activity in the lysates was determined by chemoluminescence. As shown in Table 2 below, the co-expression of Cre with plasmid pRK74 resulted in an increase of 20024 RLU while co-expression of Cre with plasmid pRK76 showed an reduced activity increase of 7554 RLU. If the latter value is defined as an inversion efficiency of 1, plasmid pRK74 is recombined with a 2.64-fold higher efficiency (Table2).

Table 2

Plasmids	ß-galactosidase activity (relative light units(RLU)	activity increase (RLU)	relative increase
pRK74	1923		
pRK74 + CMV-Cre	21947	20024	2.64
pRK76	5252		
pRK76 + CMV-Cre	12826	7574	1
No DNA	1660		

From these results we conclude that a pair of lox66/71//loxP sites (in pRK76) can be recombined (by inversion) by Cre in mammalian cells, but that this type of recognition sites is about 3-fold less efficiently recognised than a pair of lox66/lox71 sites (in pRK74). Thus, under equilibrium conditions, providing a saturating level of Cre recombinase, a plasmid containing one of the two lox site combinations, should be recombined such that about 75% of the molecules contain the lox66/71//loxP conformation and about 25% of the molecules contain the lox66/lox71 conformation. This result confirms the similar observation obtained by Albert et al., The Plant Journal, 7, 649 - 659 (1995) in plant cells and confirms that unequal recombination between mutant lox sites also occurs in a murine cell line.

In conclusion, this reaction can be used to shift the equilibrium of a Cre mediated inversion reaction to the side of the product containing the pair of lox66/71//loxP sites.

* C L M S *

Claims

1. A gene trapping construct capable of causing conditional mutations in genes which comprises a functional DNA segment inserted in a functional or non-functional manner, in sense or antisense direction relative to the gene to be trapped, said functional DNA segment being flanked by two recombinase recognition sequences (RRSs) which are specific to a site specific recombinase capable of unidirectional inversion of a double stranded DNA segment.

2. The construct according to claim 1, wherein the RRSs
(i) are present in opposite orientation, and/or
(ii) are specific to a site specific recombinase producing the inverted double stranded DNA segment in an amount of greater 50%, preferably greater 75%, more preferably greater 90%, relative to the starting double stranded DNA segment.

3. The construct according to claim 1 or 2, wherein the two RRSs are selected from mutant loxP sites mutant frt sites and AttP/AttB sites, and the site specific recombinase is selected from Cre, Flp, ΦC31-Int and λ-Int.

4. The construct according to claim 3, wherein the two RRSs are single mutant loxP sites, preferably a lox66 and a lox71 site (SEQ ID NOs:2 and 3) or an AttP and an AttB site of Φ C31, and the site specific recombinase is Cre or Φ C31, respectively.

5. The construct according to any one of claims 1 to 4, wherein the functional DNA segment further comprises one or more of the following: splice acceptor, splice donor, internal ribosomal entry site, polyadenylation sequence, a gene coding for a reporter protein, a resistance gene and a gene coding for a further site specific recombinase.

6. The construct according to any one of claims 1 to 5, which further comprises a selection DNA segment suitable for selecting for genes having an incorporated

This Sheet Has Been Replaced

gene trapping construct, said selection DNA segment comprising a reporter or resistance gene and flanking recombinase recognition sites in same orientation.

7. The construct according to any one of claims 1 to 4 which comprises two functional DNA segments,

(a) a first DNA segment (disruption cassette) being oriented to be inserted in antisense orientation relative to the transcriptional orientation of the gene to be trapped and being flanked by two RRSs which are specific to a first site specific recombinase capable of of unidirectional inversion of a double stranded DNA segment, and

(b) a second segment (selection cassette) being positioned in sense direction relative to the transcriptional orientation of the gene to be trapped and being flanked by two RRSs of a second site specific recombinase in the same orientation.

8. The construct according to claim 7, wherein the disruption cassette comprises a splice acceptor and a polyadenylation sequence flanked by said two RRSs of the first site specific recombinase, and the selection cassette comprises a reporter or selectable marker gene flanked by an upstream splice acceptor sequence and a downstream polyadenylation sequence, said construct being a conditional gene trapping construct selecting for integrations into expressed genes.

9. The construct according to claim 7, wherein the disruption cassette comprises a splice acceptor and a polyadenylation sequence flanked by said two RRSs of the first site specific recombinase, and the selection cassette comprises a reporter or selectable marker gene fused to an upstream constitutive promoter and a downstream splice donor site, said construct being a conditional gene trapping construct selecting for integrations into all genes.

10. A cell comprising the gene trapping construct as defined in claims 1 to 9.

11. Use of the cell of claim 10 for the identification and/or isolation of genes.

This Sheet Has Been Replaced

12. A process for the generation of conditional mutations in all genes of an organism comprising

(i) installation of a gene trapping construct as defined in claims 1 to 9 in a suitable cell,

(ii) selection of cells in which the construct is incorporated in a gene,

(iii) identification and/or isolation of the gene in which the construct is incorporated,

(iv) deletion of the selection cassette from the trapped gene,

(v) induction of a mutation in the trapped gene by inversion of the functional DNA segment.

13. The process according to claim 12, wherein the mutation in step (v) is effected by a site specific recombinase capable of unidiretional inversion of the functional DNA of the gene trapping construct.

14. The process according to claim 12 or 13, which is suitable for temporarily and/or spatially restricted inactivation of all genes that constitute a living organism.

15. The process according to any one of claims 12 to 14 being a process for preparing transgenic mammals, wherein in step (i) the gene trapping construct is installed in an ES cell.

16. A transgenic organism obtainable by the method of claims 12 to 15.

17. Use of the transgenic organism according to claim 16 to study gene function at various developmental stages.

This Sheet Has Been Replaced