ACADEMIC PRESS

Biochemical and Biophysical Research Communications 297 (2002) 1–5

BBRC

www.academicpress.com

Breakthroughs and Views

Gene targeting by homologous recombination: a powerful addition to the genetic arsenal for *Drosophila* geneticists

Yikang S. Rong*

Laboratory of Molecular Cell Biology National Cancer Institute, NIH, 37 Convent Dr., Bethesda, MD 20892, USA

Received 31 July 2002

Abstract

A series of recent publications have firmly established the notion that *Drosophila* researchers now have a general method to subject genes for targeted modification by homologous recombination (HR) [Science 288 (2000) 2013; Genetics 157 (3) (2001) 1307; Genes Dev. 16 (12) (2002) 1568; Genetics 161 (2002) 1125–1136]. This method allows one to knockout essentially any gene starting with the DNA sequence of the gene. It has greatly enhanced studies of gene function as demonstrated by over 20 years of gene targeting practice in yeast and mouse. Here, I discuss the basic targeting methodology for eukaryotic organisms. I compare the *Drosophila* method with the traditional targeting scheme in yeast and mouse mainly to show that the targeting mechanism as well as many aspects of the experimental design remain unchanged, and that the *Drosophila* scheme differs only in the way in which the donor molecule for targeting is generated. I propose that the *Drosophila* method can be readily adapted in other organisms without culturable stem cells, since the mechanism for in vivo donor generation in *Drosophila* is likely to be functional in a variety of different organisms.

The better understanding of gene functions can be greatly facilitated by analyses of mutants in which normal gene function has been disrupted. For example, one could deduce the specific developmental process that a gene controls by characterizing the abnormal development caused by mutations in that gene. Before a DNA sequence-based knockout method was developed, mutant lines were routinely isolated in genetic screens based on phenotypic changes. The gene was then identified by molecular characterization of the responsible mutation. Experiences have taught us that only a minority of genes could be identified in traditional screens (about one-third in *Drosophila*, [5]). This number would be even smaller for organisms whose genomes have multiple copies of certain regions due to, for example, extensive genome duplication through evolution. As more genes are identified by genomic sequencing, a knockout method becomes increasingly important as it provides the shortest link between knowledge of gene sequence and the decipherment of its function.

Gene targeting is the most widely used knockout method. It allows one to permanently modify virtually any endogenous gene. Gene targeting was invented in yeast ([6], for review see [7]). Mouse targeting was accomplished not long after ([8,9], for a recent review see [10]). The success in mouse was partly due to the establishment of an embryo-derived stem cell system (ES cells) [11,12]. Cultured ES cells can be reintroduced into the embryo and develop into part of the adult germline [13]. Until recently, a generally applicable gene targeting scheme was prominently missing for *Drosophila*, an organism long analyzed in genetic studies. So far, the Rong and Golic method has led to the targeting of more 20 different loci [1–4,14], most of which were not previously identified in screens.

Donor components: what was learned from over 20 years of targeting?

Gene targeting is based on the universal process of HR in which a cell's enzymatic machinery carries out physical exchanges between two pieces of DNA sharing regions with maximal sequence similarity. In practice,

* Fax: 1-301-435-3697.
E-mail address: rongy@mail.nih.gov

0006-291X/02/$ - see front matter
PII: S0006-291X(02)02066-1

one of the two DNA pieces is the chromosomal DNA carrying the wildtype gene one wishes to knockout, hereby referred to as the "target." The other piece is referred to as the "donor," which is exogenously supplied and carries the desired sequence modification.

In a gene targeting experiment, the greatest effort is devoted to making the best donor molecule, which optimally should have the following components. (1) A DNA double strand break (DSB). The presence of a DSB on the donor increases targeting efficiency [1,15–17]. (2) Sequence homology shared with the target locus. Extended donor-target homology increases targeting efficiency [3,18,19]. In yeast, as little as 40 bp of homology can sufficiently direct targeting [20], while in mouse ES cells, a homology of 2 kb is needed to achieve practical targeting efficiencies [18]. In *Drosophila*, the smallest amount of homology that has allowed targeting is about 2 kb [2,14]. (3) A marker for the recovery of targeting events. Selectable markers for viability greatly facilitate the recovery of targeting events. Currently, a selectable marker is not necessary in *Drosophila* due to reasonably high targeting frequencies [1–3]. Instead, targeting events have been recovered by screening for an eye color marker.

Targeting in single cell systems: what are the advantages?

Before the success in *Drosophila*, gene targeting has only been routinely performed in either single cell organisms (e.g., yeast, Dictyostelium [21], etc.), or multicellular organisms in which germ cells can be cultured as single cells in vitro (e.g., moss [17] and mouse). Several technical advantages have made the single cell systems easily adaptable to targeting manipulations. First, efficient DNA delivery methods (e.g., electroporation or chemical treatments) have allowed the introduction of donor molecules to virtually unlimited number of cells. Second, and more importantly, selections (e.g., for drug resistance) can be imposed on a large population of cells so that rare events could be easily recovered. *Drosophila*, like most other multicellular organisms, does not have a culturable stem cell system. Flies are routinely transformed with randomly inserted *P* transposable elements by DNA microinjection into early embryos [22]. There has been no report of successful targeted donor integration by direct embryo injection. On the other hand, rare targeting events have been recovered by selection in a *Drosophila* somatic cell line [23].

Drosophila targeting (I): how to generate the donor?

It would be virtually impossible to directly deliver exogenous donor molecules into *Drosophila* germ cells within an intact animal that have gone through the early stages of embryonic development. Instead, the donor has to be generated in vivo. This was accomplished by first introducing the donor into the fly genome as a precursor *P* element. Subsequently, by the concerted actions of two yeast enzymes expressed in *Drosophila* from heat inducible transgenes, the final donor for targeting was liberated from the chromosomal *P* element.

The first of the two yeast enzymes is the FLP site-specific recombinase [24]. FLP catalyzes exchanges between its 34 bp targets called *FRT*. An *FRT* has directionality. For a piece of chromosomal DNA flanked by two directly repeated *FRT*s, FLP can excise this piece of DNA from the chromosome as a circular molecule. This FLP-catalyzed excision reaction can be induced in the *Drosophila* germline at a rate close to 100% [25]. The FLP/*FRT* site-specific recombination system thus offers an efficient way to produce a piece of circular DNA not constrained by the chromosomes. To turn this circle into a recombinogenic donor for targeting, a DSB was made to the circle by the second yeast enzyme, the I-SceI site-specific endonuclease. I-SceI inflicts a DSB specifically at its 18 bp recognition site [26], which is not present in the *Drosophila* genome. It has been used to generate site-specific DSBs in many heterologous systems including *Drosophila* [27]. Greater than 90% site-specific cutting by I-SceI has been achieved in the *Drosophila* germline [1].

The donor molecule for *Drosophila* targeting has several extra features to accommodate the uses of both FLP and I-SceI (Fig. 1). First, the entire donor is placed in a *P* element vector. Second, the marker gene and the target-homologous DNA fragment are flanked by two direct *FRT*s. Lastly, an I-SceI cut site is placed within the targeting homology. After FLP and I-SceI actions (with no requirement on a particular sequence of reactions), the liberated *Drosophila* donor is almost identical to a traditional donor, except that it carries an *FRT* and the DNA ends are capped by part of an I-SceI cut site (Fig. 1).

In a targeting experiment, tripartite animals are generated by crossing. They each carry the three components of the targeting system in three different *P* elements (Fig. 1). Upon heat-shock induction of both FLP and I-SceI, a donor is generated in the germ cells of these animals, and allowed to pair and recombine with the target locus. To recover targeting events, these heat-shocked animals are mated to animals that are null for the marker gene. Most progeny of this cross would be marker-less. This is due to very frequent loss of the FLP-excised donor molecule. If a targeting event occurs, an offspring will manifest the marker phenotype. Moreover, the genetic linkage of the marker in these progeny will change from the donor chromosome to the target chromosome as a result of targeting (Fig. 1). This change in chromosome linkage has been the basis for the recovery of targeting events in *Drosophila* [1]. Subsequent molecular analyses can be used to distinguish targeted *vs.* random donor integration events. The genetic linkage screen has now been replaced by a more

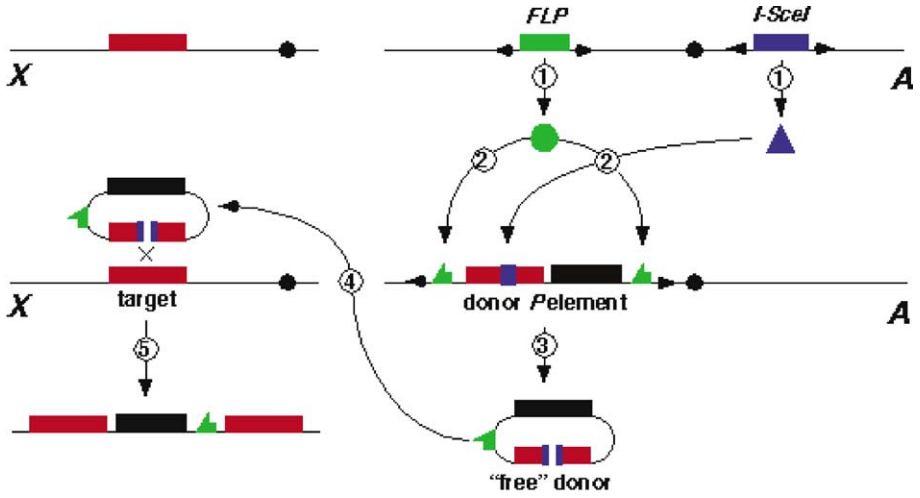

Fig. 1. An inside-the-nucleus view of the *Drosophila* targeting process. Flies with the three *P* elements are made by crossing. Only the *X* chromosome (*X*) and two autosomes (*A*) are shown. Centromeres are shown as black circles. Each *X* chromosome carries a target locus (red box). The top autosome carries transgenes for both FLP and I-SceI in two different *P* elements, with *P* element ends indicated as arrowheads. The lower autosome carries the precursor donor *P* element with the targeting homology shown in red, the marker in black, the I-SceI cut site in blue, and the *FRT*s in green half arrows. Events are numbered for the targeting process. (1) Induction of FLP and I-SceI syntheses. The green circle represents nuclear localized FLP protein. The blue triangle represents I-SceI protein. (2) The enzymes acting on respective recognition sites. (3) Generation of the extrachromosomal donor. The donor carries a DSB that is capped by I-SceI half sites. (4) The donor pairing with either one of the target loci. (5) Integration of the donor into one of the targets by HR. The I-SceI site is converted to target sequences during HR. Note that a targeted event changes the marker linkage from autosomal to *X*-linked.

convenient screening method, which is based on the facts that the marker gene is flanked by two *FRT*s in the donor so that it can be subjected to FLP-induced excision and loss, whereas only one *FRT* is left at the targeted locus so that the marker gene is stable in the presence of FLP (Fig. 1) [2,3].

Interestingly, targeting in *Drosophila* was much more efficient in the female than in the male germline by an unknown mechanism [1–3]. For the six loci targeted in the Golic laboratory, frequencies range from one event in 500–30,000 female gametes. The frequency can be improved with increased donor-target homology. The efficiency is locus-dependent, but not correlated with the distance from the target locus to the closest telomere [28]. Unlike traditional targeting in cell culture in which a large number of donor molecules are delivered into a cell, a maximum of two donor molecules can exist per cell in *Drosophila* (if the donor is produced in the G2 phase of the cell cycle). Yet, targeting in *Drosophila* has a comparable frequency. This suggests that an *in vivo* generated donor, which is most likely protein-coated, may be a better targeting substrate than "naked" DNA.

Drosophila targeting (II): how to disrupt a gene by the current scheme?

The current targeting scheme leads to a tandem duplication of the target gene. Clever designs in yeast and other organisms have allowed the generation of mutants, even in the presence of such a duplication [7]. Two methods have been successfully implemented in *Drosophila* to make null alleles of the *pugilist* gene [29]. In the first case, a "doubly truncated" donor was used, which contained an internal fragment of *pugilist* rather than the entire gene (Fig. 2A). Gene targeting resulted in a null allele with two copies of the gene each missing a different portion of the coding region [2]. This method would be especially suitable for disrupting large genes, since one does not have to limit the amount of targeting homology, even when using an internal piece of the gene. In the second case, a *pugilist* null was made by targeting with a donor carrying two point mutations on either side of the I-SceI-induced DSB [3]. In this way, both target genes were disrupted by a point mutation (Fig. 2B). It has been shown that the targeting reaction can sometimes be imprecise so that one of the target gene copies has an insertion or deletion of sequences. This interesting feature of the targeting process has been taken advantage by various groups to generate different alleles of their genes [1,3,4,14].

Despite the successful implementation of these variations in *Drosophila*, it is still desirable, in many cases, to have a single copy of the target gene carrying the engineered mutation. This can be done by a two-step allelic replacement scheme in which a reduction step will be carried out following the first step of targeting. In the reduction step, HR between the two target repeats gives rise to a single mutated target gene (Fig. 2C). In yeast and mouse, reduction events can be recovered spontaneously by selecting for the loss of the marker gene. In *Drosophila*, spontaneous reduction events have not been

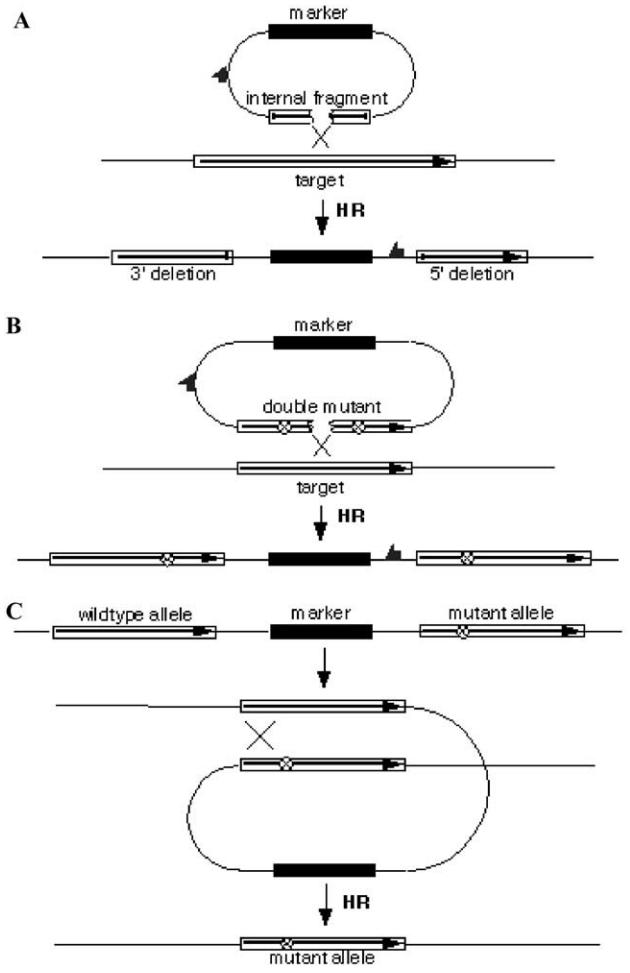

Fig. 2. Three ways to generate mutants by the current targeting scheme. In both (A) and (B), the I-SceI-cut and FLP-excised donor is shown at the top, the target locus in the middle, and the targeted locus at the bottom. The arrow represents direction of transcription. (A) A fragment missing both the 5′ and the 3′ of the target gene is used as targeting homology. A targeting event gives rise to two defective copies of the gene. (B) A fragment carrying two point mutations (stop signs) is used as targeting homology. The two mutations are on either side of the DSB. The result of a targeting event is two mutant genes each carrying a different mutation. (C) Step two of the two-step allelic replacement scheme. Step one is a targeting event similar to that shown in (B). It results in a tandem duplication of the target locus (top). The right copy carries a point mutation. Recombination occurs between the two repeats (middle), which leads to the loss of the marker (not shown) and the retention of only one target gene carrying the mutation (bottom).

recovered. However, a similar two-step replacement scheme has been developed in *Drosophila*, in which reduction events were induced by a site-specific DSB between the two target repeats [3].

Future directions: does your favorite organism have a targeting system?

It has been proposed that the targeting scheme in *Drosophila* ought to be applicable to other organisms [1]. This was based on the following considerations. First, the fundamental properties of the donor DNA should be the same for targeting in any organism. Second, the concept of generating the donor in vivo by combining the uses of a site-specific recombination system and a site-specific DSB system should not be *Drosophila* specific, since neither of the two systems requires any host factors. There is ample evidence showing that both systems are functional in many heterologous systems [30–33].

An interesting and surprising discovery was made that targeting in *Drosophila* was much more efficient in the female germ cells. The mechanism may be related to the fact that *Drosophila* males do not undergo meiotic recombination. This serves to remind us to take into account the biology of different organisms when choosing the cells in which to induce targeting. One possibility is to perform gene targeting in meiotic cells. There are at least two advantages for a meiotic targeting system. First, the machinery for HR should be highly active in meiotic cells in which homologous chromosomes normally recombine. Second, since each meiotic cell could give rise to an independent "targeted" progeny, one could greatly increase the number of targeting trials in a single individual. Targeting events in *Drosophila* were recovered in pre-meiotic germ cells that were mitotically dividing [1]. The number of trials per animal was limited to the small number of germ cells at the time of heat shocking. It is possible that *Drosophila* targeting can be made more efficient in female meiotic cells.

Acknowledgments

I thank Dr. Kent Golic at the University of Utah in whose laboratory the *Drosophila* targeting method was developed. I thank Dr. Kent Golic and Drs' Claude Klee, Bruce Paterson, and Carl Wu at NCI for comments on the manuscript.

References

[1] Y.S. Rong, K.G. Golic, Science 288 (2000) 2013–2018.
[2] Y.S. Rong, K.G. Golic, Genetics 157 (3) (2001) 1307–1312.
[3] Y.S. Rong, S.W. Titen, H.B. Xie, M.M. Golic, M. Bastiani, P. Bandyopadhyay, B.M. Olivera, M. Brodsky, G.M. Rubin, K.G. Golic, Genes Dev. 16 (12) (2002) 1568–1581.
[4] C. Seum, D. Pauli, M. Delattre, Y. Jaquet, A. Spierer, P. Spierer, Genetics 161 (2002) 1125–1136.
[5] M. Ashburner et al., Genetics 153 (1999) 179–219.
[6] A. Hinnen, J.B. Hicks, G.R. Fink, Proc. Natl. Acad. Sci. USA 75 (1978) 1929–1933.
[7] R. Rothstein, Methods Enzymol. 194 (1991) 281–301.
[8] T. Doetschman, R.G. Gregg, N. Maeda, M.L. Hooper, D.W. Melton, S. Thompson, O. Smithies, Nature 330 (1987) 576–578.
[9] K.R. Thomas, K.R. Folger, M.R. Capecchi, Cell 44 (1986) 419–428.
[10] U. Muller, Mech. Dev. 82 (1–2) (1999) 3–21.

[11] M.J. Evans, M.H. Kaufman, Nature 292 (5819) (1981) 54–56.
[12] G.R. Martin, Proc. Natl. Acad. Sci. USA 78 (1981) 7634–7638.
[13] A. Bradley, M. Evans, M.H. Kaufman, E. Robertson, Nature 309 (5965) (1984) 255–256.
[14] Personal communications from various labs.
[15] T.L. Orr-Weaver, J.W. Szostak, R.J. Rothstein, Proc. Natl. Acad. Sci. USA 78 (10) (1981) 6354–6358.
[16] M. Jasin, P. Berg, Genes Dev. 2 (11) (1988) 1353–1363.
[17] D.G. Schaefer, J.P. Zryd, Plant J. 11 (6) (1997) 1195–1206.
[18] C. Deng, M.R. Capecchi, Mol. Cell. Biol. 12 (1992) 3365–3371.
[19] B. Papadopoulou, C. Dumas, Nucleic Acids Res. 25 (1997) 4278–4286.
[20] A. Baudin, O. Ozier-Kalogeropoulos, A. Denouel, F. Lacroute, C. Cullin, Nucleic Acids Res. 21 (1993) 3329–3330.
[21] A. De Lozanne, J.A. Spudich, Science 236 (4805) (1987) 1086–1091.
[22] G.M. Rubin, A.C. Spradling, Science 218 (4570) (1982) 348–353.
[23] L. Cherbas, P. Cherbas, Genetics 145 (1997) 349–358.
[24] P.D. Sadowski, Prog. Nucleic Acid Res. Mol. Biol. 51 (1995) 53–91.
[25] K.G. Golic, S. Lindquist, Cell 59 (1989) 499–509.
[26] L. Colleaux, L. d'Auriol, M. Betermier, G. Cottarel, A. Jacquier, F. Galibert, B. Dujon, Cell 44 (1986) 521–533.
[27] Y. Bellaiche, V. Mogila, N. Perrimon, Genetics 152 (3) (1999) 1037–1044.
[28] W.R. Engels, Science 288 (5473) (2000) 1973–1975.
[29] Y.S. Rong, K.G. Golic, Genetics 150 (1998) 1551–1566.
[30] J. Odell, P. Caimi, B. Sauer, S. Russell, Mol. Gen. Genet. 223 (1990) 369–378.
[31] H. Puchta, B. Dujon, B. Hohn, Nucleic Acids Res. 21 (1993) 5034–5040.
[32] A.M. Lloyd, R.W. Davis, Mol. Gen. Genet. 242 (1994) 653–657.
[33] P. Rouet, F. Smih, M. Jasin, Proc. Natl. Acad. Sci. USA 91 (1994) 6064–6068.