Targeted homologous recombination in mammalian cells

Alan S. Waldman

Walther Oncology Center and Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, U.S.A.

(Accepted 10 October 1991)

Contents

I. Abstract ................................................................. 50

II. Introduction ........................................................... 50

III. First attempts at gene targeting ...................................... 51

A. Targeting artificial selectable loci ................................... 51

B. Direction of flow of genetic information during gene targeting ........................................................................................................... 51

C. What is the rate-limiting step of gene targeting? ................. 52

IV. Refinements in gene targeting .......................................... 52

A. Targeting natural loci ................................................... 52

B. The use of mouse embryonic stem cells ............................... 53

C. Targeting the hprt gene: sequence replacement vs. sequence insertion ................................................................. 53

V. Homology dependence of gene targeting ............................ 54

VI. Fidelity of gene targeting ............................................... 54

VII. Methods for detecting gene targeting ................................ 55

A. Use of the polymerase chain reaction (PCR) ....................... 55

B. Positive-negative selection ............................................... 56

C. Other enrichment methods .............................................. 58

VIII. Knocking-out both alleles of a gene by gene targeting .......... 59

IX. Additional approaches ................................................. 60

A. Direct injection into mouse eggs ...................................... 60

B. Introducing subtle changes by gene targeting .................... 60

1. An 'in-out' targeting procedure ....................................... 60

2. Selectable subtleties ..................................................... 61

X. Summary and future perspectives ...................................... 61

References ................................................................. 63

Alan S. Waldman received a Ph.D. in Biochemistry from the Johns Hopkins University in Baltimore, MD. He completed a postdoctoral fellowship at the Yale University School of Medicine in New Haven, CT. Dr. Waldman is currently an Assistant Member of the Walther Oncology Center and an Assistant Professor of Biochemistry and Molecular Biology at the Indiana University School of Medicine in Indianapolis, IN.

Correspondence: Alan S. Waldman, Walther Oncology Center, Indiana University School of Medicine, 975 West Walnut Street, Indianapolis, IN 46202, U.S.A.
I. Abstract

This article presents a review of recent progress in the field of targeted homologous recombination in mammalian cells. Beginning with an introduction of basic terminology and why 'gene targeting' is potentially such a powerful genetic tool, the article explores some of the obstacles that must be overcome in order for targeting to be generally useful. In particular, the different ways in which investigators have been able to work around the great inefficiency of gene targeting is covered in some detail. When possible, insights into the mechanism(s) of gene targeting are extracted from the published literature. The use of targeted gene 'knockout' in mouse embryonic stem cells to create animal disease models is discussed. The need for systematic studies into the mechanism(s) of targeting to make gene targeting useful for human gene therapy is recognized, and some suggestions are made.

II. Introduction

In the purest sense, 'homologous recombination' can be defined as any process in which two similar DNA sequences interact and undergo an exchange of genetic information. The outcome of homologous recombination is the production of new 'recombinant' molecules containing sequence information from each of the two parental molecules. With such a broad definition as a starting point, one might well imagine that there exists quite a variety of cellular processes that might be termed homologous recombination. The ambiguity of the term is indeed sometimes a source of confusion. For convenience, we may divide homologous recombination into three general categories: extrachromosomal recombination (occurring between two extrachromosomal sequences), intrachromosomal recombination (occurring between two chromosomal sequences), and targeted recombination or 'gene targeting' (occurring between one extrachromosomal and one chromosomal sequence).

This latter form of recombination has received much attention in the past several years and the term 'homologous recombination' is now commonly (albeit incorrectly) used to refer specifically to gene targeting. This review will in fact focus primarily on this particular subclass of homologous recombination events. The reader is also referred to additional reviews on gene targeting [1–5]. Other reviews [1,6–11] offer coverage of other aspects of homologous recombination in mammalian cells in more detail.

Why is gene targeting exciting? The ability of a cell to swap genetic information between an extrachromosomal DNA molecule and a chromosomal 'target' sequence opens the door for the deliberate introduction of defined alterations to a genome. Such a feat can theoretically be accomplished by first engineering the altered sequence in vitro and then introducing the new genetic information into a cell by any of a variety of transfection techniques. If a transfected molecule successfully interacts with the chromosomal target via homologous recombination, the desired alteration will be transferred to the target.

By such methodology, there are two basic types of alterations that can be made to a chromosomal sequence: disruption of genetic function or restoration of function. Disruption-of-function alterations are commonly referred to as targeted 'gene knockouts'. By disabling a gene, it becomes feasible to learn about gene structure and function and ultimately to create animal models for genetic diseases. On the other hand, correcting genetic defects by targeting holds the promise for an approach for effective human gene therapy (see Verma [12] and Belmont and Caskey [13] for reviews on gene therapy).

Currently used approaches for gene therapy make use primarily of retrovirus vectors which are capable of efficiently delivering a functional gene to a human genome, thus complementing a given genetic defect by gene augmentation. In such a procedure the functional gene, usually along with the vector, integrates into random loci. Such 'illegitimate' integrations pose several potential problems: (i) the integration or expression might be unstable and the therapy thus short-lived; (ii) the integrated gene might be weakly or inappropriately expressed at its unnatural locus; (iii) the integration event itself might disrupt an essential gene or activate a quiescent oncogene; (iv) such approaches are useless for alleviating genetic problems involving overexpression of a protein or expression of a deleterious protein. Because of such problems intrinsic to approaches involving random integration, it is clear that the best way to perform gene therapy would be to precisely replace the damaged gene with a wild type gene, leaving the remainder of the genome unchanged (that is, gene therapy by gene targeting).

Before the full potential of gene targeting can be realized, however, there are obstacles that must be overcome. A major obstacle is the inefficiency of the targeting process. When DNA is transfected into mammalian cells, the vast majority of interactions between the transfected molecule and the genomic DNA will result in a random integration ('illegitimate recombination') of the input DNA molecule rather than gene targeting. The ratio of illegitimate events to gene targeting has been reported to be anywhere between 30:1 to more than 100,000:1, with a ratio of 1000:1 about 'average' [1–5]. This is in striking contrast to what occurs when DNA
is introduced into lower eukaryotes such as yeast or fungi, in which targeting is the rule and illegitimate recombination is the exception (see Roth and Wilson [14] for a review). It is the rarity and potential power of targeting in higher eukaryotes that makes gene targeting in mammalian cells an intriguing process that presents molecular geneticists with a worthwhile challenge. This review documents how some investigators have taken up the challenge to manipulate the genome of mammalian cells through homologous recombination.

(This review will focus on advances in methodology and insights gained into the mechanisms of the gene targeting process itself rather than what has been learned about specific genetic loci through gene targeting approaches. The reader is encouraged to refer to the primary literature cited in this review for further information.)

III. First attempts at gene targeting

III-A. Targeting artificial selectable loci

The mammalian genome is enormous (upwards of $10^9$ bp) in comparison to the small size of a target sequence (typically on the order of $10^3$ bp). In addition, mammalian cells are quite efficient at performing random integration of transfected DNA [14]. Armed with this knowledge, the expectation of investigators attempting the initial gene targeting trials was that gene targeting events would be very rare if they occurred at all. The first attempts at gene targeting, therefore, involved reconstruction of selectable targets to aid in the recovery of the rare events. Most early work made use of artificially introduced loci rather than natural loci. The general scheme was to introduce a defective marker gene into a cell so that it became stably integrated into a chromosome. Such a cell line was then transfected with a second copy of the marker gene harboring a different mutation. Selection for reconstruction of a functional marker thereby served as a direct selection for gene targeting. Markers used included the neo [15-17] and Herpes thymidine kinase (tk) [18] genes. These early attempts met with some success; the ratio of gene targeting to random integration varied from $10^{-2}$ to $10^{-5}$.

The initial studies illustrated several important points. First, and perhaps foremost, it became evident that gene targeting in mammalian cells was a real possibility. Secondly, a theme emerged that the frequency of targeting may vary over several orders of magnitude depending on the precise nature of the targeting event studied. In studies by Lin et al. [18] involving targeted correction of a defective Herpes simplex virus tk gene target in mouse Ltk− fibroblasts, only one out of ten cell lines constructed was capable of supporting a measurable level of gene targeting. Line-to-line variation was seen in other studies using artificial loci [17]. It may be speculated that perhaps the site of integration of the artificial tk locus in the various cell lines might have influenced the targeting rate observed by Lin et al. [18]. This suggests that local chromatin structure and function might influence gene targeting in general. In the single cell line that did exhibit targeting capability, the absolute targeting frequency was about one event per $10^6$ cells transfected and the ratio of random integration to targeting was estimated at 100,000:1 [18].

Studies by Thomas et al. [16] were somewhat more successful. In these latter studies, mouse Ltk− cells were again used. In this work, a chromosomal target consisting of a defective neo gene was corrected via targeting with a second defective neo gene that was microinjected directly into the nuclei of the cells. The events recovered were predominantly gene conversions (nonreciprocal exchanges of genetic information). Thomas et al. recorded a ratio of random integration to targeting of 100:1 (with an overall efficiency of 1 targeting event per 1000 cells injected.) Why these investigators met with such success compared to Lin et al. was not clear, although it was speculated that the method of transfection, CaPO₄ coprecipitation (used by Lin et al.) versus microinjection (used by Thomas et al.), might have been a factor [16]. Following CaPO₄ coprecipitation, DNA forms large concatemers [19]. Furthermore, DNA is damaged during a variety of transfection procedures [20,21]. Both of these occurrences, neither of which occurs at appreciable levels following nuclear microinjection, may hinder the targeting process. Thomas et al. also observed no line-to-line variation of targeting efficiency. This could have been a reflection of the sites of integration of the neo target or a property intrinsic to the neo sequence. It was also noted that linearization of input DNA greatly stimulated targeting.

III-B. Direction of flow of genetic information during gene targeting

Some of the G418' colonies that arose in the work of Thomas et al. [16] were due to the incoming defective neo gene being corrected by the target sequence and the corrected transfected molecule subsequently integrating into a random genomic locus. This 'reverse' flow of genetic information from chromosomal target to transfected DNA has been occasionally observed in other works [22-24] and is clearly not desirable for most intended gene targeting applications. There was a suggestion made by Thomas et al. [16] that the nature of the two neo mutations used in their work might have dictat-
ed the direction of flow of genetic information. Sequence information appeared to flow predominantly from a neo gene with an amber mutation to a neo gene carrying a deletion. The authors indicated that such observations might have relevance for planned targeted disruptions of endogenous loci. They suggested that it might be wise to try to introduce an insertion rather than a deletion mutation into a genomic sequence so that information would likely flow to rather than from the genomic target. The accuracy of this speculation has not been sufficiently tested.

III-C. What is the rate-limiting step of gene targeting?

Thomas et al. [16] made an intriguing observation that directly addresses the question as to what limits the efficiency of gene targeting in mammalian cells. They determined that the efficiency of targeting did not improve as the amount of DNA microinjected into the nuclei of mouse cells was increased over a 20-fold range, from 5 copies to 100 copies per cell. A similar observation was made in another study by Rommerskirch et al. [25] involving targeting of SV40 sequences in COS7 cells. From a slightly different perspective, Zheng and Wilson [26] later reported that the frequency of targeting of the dihydrofolate reductase gene in Chinese hamster ovary (CHO) cells was not enhanced by amplifying the copy number of the target gene by a factor of 400. The general interpretation that has been applied to this collective set of data is that the initial interaction between a transfected DNA molecule and a homologous chromosomal target sequence is not the rate-limiting step of gene targeting. This interpretation implies that random integration does not compete with targeting and that the low frequency of targeting in mammals relative to that observed in yeast is not due to dilution of the target sequence in a genome that is 200-fold larger than that of yeast [26].

One must be cautioned, however, that such an interpretation may not be valid due to numerous complexities in what may seem to be straightforward experiments. For example, as one introduces increasing amounts of DNA into a mammalian cell, the frequency of encounters between these transfected molecules will likely increase as will the occurrence of extrachromosomal recombination among them. It was demonstrated [27] that when large amounts of DNA are introduced into cells, long tandem arrays of the transfected molecules form due to homologous recombination among the transfected molecules. (In general, as the amount of input DNA is increased, the length of the tandem arrays integrated into the genome increases but the number of genomic sites of integration per cell does not increase dramatically [27].) If homologous recombination among transfected molecules competes with gene targeting, then increasing the number of transfected molecules might not improve targeting efficiency simply because the frequency of a competing event is also increased. This may be true irrespective of whether or not the initial encounter between transfected DNA and the target sequence is a rate-limiting step in gene targeting. Furthermore, in studies in which a target sequence was amplified 400-fold with no effect on targeting frequency [26], interpreting the results may be complicated by the possibility that a tandem array of 400 target sequences may be different than 400 unlinked targets. A large tandem repeat of genomic sequences, as studied by Zheng and Wilson [26], may have special structural features due to homologous interactions among the repeats, and this may have important effects on the ‘targetability’ of the locus.

It is a curious and rather counter-intuitive notion that the rate of targeted homologous recombination might in fact not be dependent on the concentration of either of the two molecules taking part in the event. To add to this curiosity is the report that when targeting takes place in cell lines with 800 copies of a target, only a single targeting event typically occurs [26] (seemingly indicating that a targeting event excludes additional targeting events in the same cell). In comparison, the rate of homologous recombination between transfected DNA molecules is DNA concentration dependent [28]. The true rate-limiting step of targeting must be elucidated if we are to improve the efficiency of the process.

IV. Refinements in gene targeting

IV-A. Targeting natural loci

Initial studies of targeting at artificial loci were useful and such models may still serve as valuable experimental tools, but it was recognized early on that gene targeting would truly be useful only if natural loci could be targeted. The first example of targeting at a natural locus was reported by Smithies et al. [29]. The locus they targeted was the human β-globin gene. The complicated scheme used relied on detecting the linkage of a supF marker (originally contained in the targeting construct) to the target locus and was designed to enable the recovery of an unselected targeting event. Targeted insertion of exogenous DNA sequences into the β-globin locus occurred in about 1 in 1000 successfully transfected cells, similar to ratios seen at artificial loci. Importantly, targeting of the β-globin gene occurred at equal frequency in cells in which the globin locus was expressed or unexpressed. The concept of targeting unexpressed genes will be discussed below.
The studies by Smithies et al. opened the door for other targeting experiments involving natural loci. One fruitful approach was to study targeting of natural selectable loci. Adair et al. [30] were able to correct a defective adenine phosphoribosyltransferase (aprt) gene in an aprt-deficient CHO cell line. The aprt+ phenotype was easily selected [30]. Transfection with a plasmid containing a fragment of the wild-type aprt gene yielded aprt+ recombinants at a frequency of about $4 \times 10^{-7}$, with a ratio of targeting to random integrations equal to approx. 1:4000. Many studies have involved targeting the hypoxanthine phosphoribosyltransferase (hprt) gene, another selectable locus. Some of these studies will be discussed below.

Other investigators [31,32] have used established hybridoma cell lines to correct a defective immunoglobulin gene by gene targeting. These experiments involved a screen for plaque-forming cells, rather than a direct selection, in order to isolate cells in which the immunoglobulin function had been restored. In these reports the overall frequency of targeting was about $1 \times 10^{-7}$ and the ratio of targeting to random integration was about 1:1000. Additional reports have appeared on the production of a mouse-human chimeric antibody resulting from gene targeting at the heavy chain locus of a hybridoma cell line [33].

**IV-B. The use of mouse embryonic stem cells**

An important advance in the field of targeting was the introduction of the use of mouse embryo-derived stem (ES) cells. ES cells are derived from a 4.5 day preimplantation embryo (the blastocyst) and can be grown in vitro in the presence of a feeder layer of cells that provide an attachment matrix as well as factors that inhibit differentiation of the ES cells [34,35]. (Recently, investigators have developed methodology to grow ES cells in the absence of feeder layers by supplementing the growth medium with leukemia inhibitory factor [36,37], which greatly simplifies experimental manipulations.) If handled properly, the ES cells retain pluripotency and, when reintroduced into a blastocyst, the ES cells contribute to the formation of all tissues in a chimeric mouse [4]. Thus if gene targeting is performed in ES cells in vitro, the resulting change to the mouse genome can be incorporated into a whole animal. This provides a powerful genetic tool and allows the development of animal models for disease.

**IV-C. Targeting the hprt gene: sequence replacement vs. sequence insertion**

The hprt gene has provided a particularly useful model system for several groups studying targeting in ES cells. Features that make the hprt gene useful are: (i) the hprt gene is on the X chromosome and so there is only one copy of the target in male ES cells; (ii) there are strong genetic selections both for and against hprt expression.

Thomas and Capecchi [38] were the first to report gene targeting in ES cells in work involving the disruption of the hprt locus by insertion of a neo gene into the hprt coding region. Cells were electroporated with constructs containing a portion of the hprt gene disrupted by a neo gene. Cells in which targeting had likely occurred were isolated by selecting for G418' (which selected for any integration of the neo gene into the genome) and resistance to 6-thioguanine (6-TG) (which directly selected for the hprt- phenotype). The frequency of targeting was about $10^{-7}$ per electroporated cell while the ratio of targeting to random integration in the most successful experiments was about 1:1000.

The work by Thomas and Capecchi [38] helped to illustrate that there are two general modes by which targeting can occur: sequence 'replacement' or sequence 'insertion'. In sequence replacement, target sequences are precisely replaced by sequences on the transfected molecule by a gene conversion or double crossover. There is no net change in the number of nucleotides in the genome. (This is presumably the type of event desired in gene therapy.) In a sequence insertion, the entire transfected molecule inserts into the target locus bringing about a net increase in the number of nucleotides in the genomic sequence. In an insertion event, nonhomologous sequences on the transfected molecule are brought into the target locus; the target sequence itself is duplicated and flanks the inserted vector sequences.

Thomas and Capecchi [38] determined that it is possible to experimentally manipulate the relative frequencies of replacement versus insertion by a very simple means. It was observed that if the transfected molecule were cleaved prior to transfection so that the vector hprt sequences were colinear with the endogenous sequence, then, among the targeted events recovered, all were sequence replacements. If, on the other hand, the transfected molecule were cleaved within the hprt gene so that the ends of the linearized molecule were adjacent to one another on the hprt map, then the majority of the targeting events recovered were sequence insertion events. Thomas and Capecchi also reported that replacement and insertion events can occur with nearly equal facility.

Several groups have reported on the successful 'correction', rather than disruption, of a mutant hprt gene in ES cells via targeted recombination [39-41]. Work by Thompson et al. [40] on correction of hprt in ES cells had the added significance of being the first report of the
successful transmission of a targeted genomic alteration through the germline of chimeric mice. Koller et al. [41] have subsequently also reported on the successful transmission of a corrected mutant hprt gene into the germ-line of chimeric mice. There have been additional reports of the germ line transmission of targeted disruptions of the c-abl [42], β2-microglobulin [43], int-1 [44] and c-src [45] loci in mice derived from ES cells.

V. Homology dependence of gene targeting

An important observation made by Thomas and Capecchi [38] in their work describing the targeted knockout of the hprt gene in ES cells was that targeting exhibited an apparent strong homology dependency. Increasing the amount of homology shared between transfected molecule and target from 4 kb to 9.1 kb increased the absolute frequency of targeting over 10-fold and the ratio of gene targeting to random integration was increased 40-fold. In studies involving targeting at an immunoglobulin gene locus in hybridoma cell lines, Shulman et al. [46] also recorded an increase in targeting frequency as homology between transfected DNA and target was increased from 1.2 to 9.5 kb. The largest effect, a 25-fold increase in targeting, was seen as homology was increased over the range from 2.5 to 9.5 kb. Between 1.2 kb and 2.5 kb of homology, there was no discernable effect on targeting.

Doetschman et al. [47], in work involving disruption of the hprt gene in ES cells by a promoterless neo gene, reported a higher absolute frequency of targeting than did Thomas and Capecchi [38] (0.02–0.3 × 10⁻⁶ vs. 0.4–1.6 × 10⁻⁶). Interestingly, the transfected DNA in the work of Doetschman et al. shared considerably less homology (1.3 kb) with the hprt target than did the constructs used by Thomas and Capecchi (4–9 kb). It has been suggested [1,47] that perhaps the different regions of the hprt gene that were targeted in the two works might have influenced the targeting frequencies. In the final analysis, it is presently difficult to make an assessment of any generalized effect that the extent of homology alone might have on targeting efficiency.

The above works did not address the effect that sequence heterologies may have on the frequency of targeting. Studies by Letsou and Liskay [48] demonstrated that the frequency of intrachromosomal gene conversion in mouse fibroblasts was inversely proportional to the length of a block of nonhomologous sequences converted in the recipient DNA sequence. This has led to the assumption that the frequency of gene targeting in mammalian cells would also be inversely proportional to the length of a block of nonhomologous sequence transferred to the chromosome [1,2,38]. It was recently demonstrated by Mansour et al. [49] that this is not the case. The frequency of targeting was not affected as the length of nonhomologous sequences transferred to the hprt locus of ES cells was varied from 8 bp to as much as 12 000 bp [49]. In these experiments, it is important to note that substantial blocks of contiguous homology (> 6 kb) were still shared between the exogenous sequences and the target.

Work by Waldman and Liskay [50] indicated that initiation of intrachromosomal gene conversion in cultured mouse cells is very sensitive to base pair mismatches, and that even single nucleotide heterologies can have an impact on the frequency of homologous recombination. Efficient initiation of intrachromosomal recombination in mouse cells requires that two sequences share more than 130 bp of contiguous, uninterrupted homology. Of relevance to this discussion, Waldman and Liskay [50] have shown that once intrachromosomal homologous recombination is initiated within a region of homology, the event may propagate through nearby heterologous sequences. Whether gene targeting in mammalian cells exhibits a similar homology dependence for ‘initiation’ and a similar heterology insensitivity for ‘propagation’ remains to be determined.

VI. Fidelity of gene targeting

Gene targeting ideally involves the precise alteration of the target sequence with no other change to the genome. What is the likelihood of such a scenario? Unfortunately, there is growing evidence that targeted homologous recombination may be a mutagenic process. Unlike intrachromosomal homologous recombination, which appears to operate with great fidelity in mammalian cells [51], targeting events are often associated with aberrant alterations of the input DNA, the target, or adjoining sequences [43,47,52–55]. In work by Thomas and Capecchi [52], as much as 50% of the events initially recovered and presumed to have arisen from targeted correction of a defective neo target were subsequently found to be due to the introduction of a second compensating mutation into the chromosomal target by interaction with the transfected DNA molecule without any correction of the original mutation. The process responsible for this mutagenesis was termed ‘heteroduplex-induced mutagenesis’. This type of event has yet to be reported elsewhere probably because these events were directly selectable in the system of Thomas and Capecchi and because the very nature of the neo sequences allowed for the strand-slipage mechanism proposed for this phenomenon [52].

Although no other groups have reported on hetero-
duplex-induced mutagenesis, there have been several reports of various mutations and illegitimate rearrangements recovered along with a desired targeting event. These unwanted occurrences included small insertions [52], deletions [47], point mutations [54] and other mutations or abnormal rearrangements [43,53,55] of target and flanking sequences. The precise mechanisms for the generation of these unwanted rearrangements is not presently clear. It should be noted that aberrant rearrangements are also commonly seen at the integration site of illegitimate genomic insertions of transfected DNA (see Roth and Wilson [14] for a review). This raises the possibility that the insertion of exogenous sequences into a mammalian genome might always be associated with additional processing of nearby genomic DNA.

An issue related to the observations of unwanted sequence changes at the target locus itself is the question as to whether illegitimate integrations of additional copies of transfected DNA molecules occur elsewhere in the genome of a cell in which gene targeting has occurred. Although it has been reported [1] that random integration of transfected DNA is generally rare in cells in which gene targeting has occurred, there are indeed numerous reports in which random integrations have been observed along with targeting (for example, see Thomas et al. [16], Lin et al. [18], Adair et al. [30], Sedivy and Sharp [53], and le Mouellic et al. [56]). In some reports, multiple random integrations were recovered along with the targeting event. The lack of observation of random integrations in targeted cells in certain studies is almost certainly a reflection of the rarity of the simultaneous occurrence of two rare events rather than a mutual exclusion of random and homologous integrations within a single cell. For the purposes of gene therapy, it is clearly desirable, if not imperative, to eliminate random integration as well as mutagenesis at the target locus.

VII. Methods for detecting gene targeting

The low frequency of gene targeting in mammalian cells presents an obvious difficulty in the recovery of such events. There are potentially two ways to deal with this problem – increase the frequency of gene targeting or devise clever means of isolating the rare events. To this day, largely due to the lack of any significant gains in our understanding of targeting mechanisms, investigators have focused on the latter approach (with great success, I might add). When disruption or activation of the target sequence itself provides a powerful genetic selection, as in the case of the hprt locus, recovering gene targeting events is reasonably straightforward. When there is no direct genetic selection for or against expression of the target locus, other means must be sought.

VII-A. Use of the polymerase chain reaction (PCR)

The advent of the development of the polymerase chain reaction (PCR) has provided a means of screening for gene targeting. Kim and Smithies [57] have demonstrated that by the appropriate use of primers it is feasible to detect novel sequence arrangements created by gene targeting, even when a targeted cell represents only 1 cell in 10,000. The advantage of this 'brute force' screening approach is that it requires no selection and, therefore, does not require the use of a selectable marker for gene knockout. The disadvantage is that it is not feasible to screen for targeting events occurring at frequencies lower than about $10^{-6}$.

The use of PCR screening alone was employed successfully by Zimmer and Gruss [58] to knock-out the homeobox-containing Hox 1.1 gene in mouse ES cells. Approximately 1 in 150 microinjected cells were successfully targeted with a 20 bp insertion into the homeobox domain. Taking into account a microinjection efficiency of about 20%, the ratio of targeting to random insertion was 1:30, quite remarkable. Why did Zimmer and Gruss attain such an outstanding targeting efficiency? It is possible that the high efficiency was due in part to the investigators' purposeful decision to use microinjection as the means for transfection, since there was some evidence that this mode of transfection yields a higher frequency of targeting than other methods [16,58]. Another possibility that had been suggested by Zimmer and Gruss and others [1] was that the small amount of heterology (20 bp) transferred to the target might have accounted for the high efficiencies. For comparison, most targeted disruptions have involved the transfer of heterologous sequence in excess of 1 kb. However, recent work by Mansour et al. [49] (discussed above), showing that targeting frequencies do not appear to be affected by the size of transferred heterology between 8 bp and 12 kb, tends to discredit such a notion. It may well be that the Hox 1.1 locus is a hotspot for targeting due to some special structural feature of the chromatin of that region. The precise reason for the high targeting efficiency at Hox 1.1 is still open for debate.

Joyner and co-workers [59] used a scheme involving genetic selection combined with screening by PCR to knock-out the developmental En-2 gene in ES cells with a neo disruption. The targeting construct contained the neo gene driven by the human β-actin promoter and flanked by En-2 sequences so that gene targeting would result in the disruption of the homeobox-containing exon of En-2 with neo. Following electroporation of ES cells, selection for G418' was first applied to select for cells that had integrated the targeting construct at any
locus. These cells were then screened by PCR to identify the subset of cells that contained the desired targeted disruption. About 1 in 400 G418' cells were found to be targeted, with an overall targeting efficiency of 10^{-7}. This work illustrated how genetic selection can be combined with a PCR screen to isolate targeting events occurring at absolute frequencies too low to be recovered by a PCR screen alone.

Zijlstra et al. [43] as well as Koller and Smithies [60] introduced a neo disruption into exon two of the β2-microglobulin gene in a strategy involving selection for G418' combined with a PCR screen. Like the work of Joyner et al. [59], an initial selection for G418' allowed recovery of cells in which any integration of the targeting construct had occurred. Additionally, in the work of Zijlstra et al. [43], the neo gene was engineered to lack poly(A) sites so as to lessen the likelihood of neo expression upon random integration. In this case, upon targeting the neo gene would use the 3'-processing sequences of the β2-microglobulin gene. Screening of G418' clones by PCR allowed both research groups to identify those clones in which the construct had replaced part of the β2-microglobulin locus by targeting. Zijlstra et al. [43] achieved a targeting to random integration ratio of about 1:25 among the G418' clones (comparable to the ratio seen by Zimmer and Gruss who used no selection), with an overall targeting frequency of about 0.5 × 10^{-6}. The investigators commented that the exceptionally high percentage of targeting events might reflect the existence of a recombination hotspot in the β2-microglobulin gene [43]. Indeed, evidence exists for a high frequency of chromosomal rearrangements at the β2-microglobulin locus in lymphoma cell lines [61]. Koller and Smithies [60] also obtained a rather high efficiency of targeting, with a targeting to random integration ratio of 1:100 among G418' clones. This too is consistent with the idea that the locus contains a hotspot for homologous recombination. Interestingly, Koller and Smithies report that attempts at targeting the β2-microglobulin gene in mouse L cells failed. Curiously, the β2-microglobulin gene is highly expressed in mouse L cells, while it is very poorly expressed in mouse ES cells [60]. Therefore, it would appear that there is no correlation between the level of expression of a gene and the ease with which it can be targeted.

Le Mouellic et al. [56] recently reported the targeted disruption of the Hox 3.1 gene of ES cells by a construct containing the Escherichia coli lacZ gene configured in such a way to be placed under the control of the Hox 3.1 promoter upon targeting. The construct contained the neo gene as well. Again, selection for G418' was first applied followed by screening by PCR for gene targeting. Le Mouellic and co-workers used a construct that did not contain poly(A) signals for the neo gene, as did Zijlstra et al. [43]. The targeting construct also contained an A + T sequence contained in a Hox 3.1 intron sequence to direct the degradation of neo transcripts from randomly integrated constructs [58]. This RNA degradation signal was engineered to be lost upon targeted integration due to splicing of the Hox 3.1 transcript or by the recombination event itself. The scheme worked well and about 1/40 of the G418' clones obtained represented gene targeting, with an overall efficiency of about 10^{-6}. The authors commented that the ratio of 1 targeting event per 40 random insertions was in fact unexpectedly high since a similar construct containing 3' neo mRNA processing signals and lacking the mRNA degradation signal yielded only 2.4-times as many G418' clones and yet, with this construct, the ratio of targeting to random insertion was 1:900. The authors speculated that perhaps they inadvertently engineered a recombination hotspot into their vector, that perhaps the A + T mRNA degradation sequence promoted strand melting which in turn stimulated homologous recombination.

Another interesting aspect of the work by le Mouellic et al. [56] is that lacZ gene expression was placed under the control of the Hox 3.1 promoter in the targeted cells. LacZ expression is easily visualized at the level of just a few cells by staining with X-gal. Since the Hox3.1 gene is not expressed in ES cells [56] the targeted cells did not express LacZ. By making chimeric mice from targeted cells, le Mouellic and co-workers followed the fate of cells that express Hox 3.1 during embryogenesis. In this way, the investigators were able to confirm and expand data about the pattern of Hox3.1 transcription obtained previously from in situ hybridization determinations.

Soriano et al. [45] recently reported on the targeted disruption of the c-src proto-oncogene by a neo gene insert using a scheme combining selection for G418' and a PCR screen. In this report the absolute targeting efficiency was about 4 × 10^{-7} with a targeting to random integration ratio of 1:160 (quite high!).

Aside from being used as a screen to isolate targeted cells, PCR has also been used purely as an analytical tool to confirm gene targeting events that were recovered by direct genetic selection [62].

VII-B. Positive-negative selection

It is clear that the simplest way to recover a targeting event is by direct selection. Unfortunately, the vast majority of loci are not selectable and in fact the functions of many loci are not known. Recently, Mansour et al. [63] devised a selection scheme that greatly enriches for
gene targeting in a manner that is independent of gene function and indeed independent of gene expression. The enrichment procedure is known as 'positive-negative selection' (PNS). The procedure consists of a positive selection for cells that have integrated the targeting construct into any locus of a cell's genome and a negative selection against cells containing a random integration.

More specifically, the PNS targeting construct contains part of the gene to be targeted disrupted by a neo gene under the control of a strong promoter. Selection for expression of neo (G418') thus provides a positive selection for any integration of the construct into the genome. The construct also contains the Herpes simplex virus tk gene. In the positive-negative selection scheme, the targeting construct is linearized in such a way that the Herpes tk gene is at the terminus of the linear molecule and the sequence of interest is arranged to be colinear with the genomic target sequence (i.e., a gene 'replacement' construct, see above). The idea is that if the transfected DNA engages in targeted replacement of the genomic sequence, the tk gene is not inserted into the genome. On the other hand, if the construct integrates into a random locus, such integration will likely occur via the ends of the molecule [14] and so the entire construct, including the tk gene, will be inserted into the genome.

The Herpes tk protein has a relaxed substrate specificity relative to that of mammalian tk proteins. The compound gancyclovir is a substrate for the viral enzyme but not mammalian tk. Phosphorylation of gancyclovir by tk is toxic to the mammalian cell, perhaps by acting as a chain terminator in DNA synthesis. Cells expressing the Herpes tk gene are, therefore, sensitive to gancyclovir, so the presence of the tk gene in the genome provides a negative selection against random integrations. Cells in which targeting had occurred will remain resistant to gancyclovir. The simultaneous use of both the positive (G418') and negative (gancyclovir') selections therefore results in a great enrichment for targeted cells.

Mansour et al. [63] initially demonstrated the usefulness of this approach by successfully targeting the hprt and int-2 loci in ES cells. For the hprt locus, 19 out of 24 G418'-gancyclovir' colonies analyzed were successfully targeted. For the int-2 locus, 4 out of 81 G418'-gancyclovir' colonies contained a neo disruption in the int-2 gene. The authors estimate that the positive-negative selection scheme resulted in an approx. 2000-fold enrichment for targeting. In later work by the same authors [49], a lacZ reporter gene was inserted into the int-2 locus using PNS and in this work 9 out of 89 clones surviving the double selection were successfully targeted. Importantly, the lacZ gene was inserted as an in-frame fusion with the int-2 protein-coding region. Upon differentiation of the ES cells to embryoid bodies, the fusion marker gene accurately reproduced the expression pattern of int-2 mRNA. This technology might greatly aid in developmental studies in mice (similar to the work by le Mouellic et al. [56] described above).

In the initial work on PNS by Mansour et al. [63] the relative and absolute frequencies of targeting of the int-2 locus were over 10-fold lower than the frequencies obtained at the hprt locus. It was suggested [1,63] that this difference may have been due to the relatively low level of expression of the int-2 gene in ES cells. Indeed, it has been shown both in yeast [64,65] and mammalian [66] systems that transcription tends to enhance homologous recombination processes in mitotically dividing cells. However, early work by Smithies et al. [29] and later work by Koller and Smithies [60] (see above) indicated that there is no obvious correlation between the level of expression and the ability to target a locus. Additionally, in work published by Johnson et al. [67], PNS was applied equally successfully for disrupting the c-fos, adipsin, and adipocyte P2 (aP2) genes in ES cells. The c-fos gene is expressed at a low level, whereas there is no detectable level of expression of the latter two genes in ES cells [67]. Nevertheless, each of these genes was targeted using PNS and the frequency of targeting of these genes did not correlate with the level of expression of the genes (the absolute targeting efficiencies for c-fos and adipsin were both about 10^{-5}, while the frequency for aP2 was about 10^{-7}). In all cases, about 5 10% of the colonies surviving PNS exhibited a targeted disruption of the appropriate locus. An important point of the work by Johnson et al. [67] was that it demonstrated the feasibility of targeting unexpressed loci in ES cells using PNS.

In the original work describing the PNS procedure [63] the authors comment that the enrichment for targeting by PNS should be improved by placing two tk genes on either end of the transfected molecule. The authors argue that during transfection a tk gene might become damaged (or lost) and, therefore, the gancyclovir selection against random integration would be rendered ineffective. That this is a very real possibility was indicated in work by Zijlstra et al. [43] as well as work by Thomas and Capecchi [44]. In the former work involving disruption of the β2-microglobulin gene, a scheme to use PNS was abandoned when it was learned that 50% of randomly picked G418' colonies had lost the tk gene. In the work by Thomas and Capecchi involving targeting at the int-1 locus using PNS, only 1 out of 400 colonies surviving PNS was correctly targeted. This was very likely a reflection of loss or damage to the tk genes in the 399 non-targeted colonies that survived PNS.
In this latter work, two tk genes were actually included on the construct. The overall efficiency of targeting at the int-1 locus was also extraordinarily low, with an overall targeting to random integration ratio of 1 in $5 \times 10^6$ and an absolute frequency of $2 \times 10^{-9}$. It is not clear if there is any mechanistic link between the low targeting efficiency and the apparent high level of damage suffered to the tk sequences on the targeting construct. It should be noted that the transfection efficiency as measured by recovery of G418\(^{+}\) clones (i.e., the positive selection) was about 1 colony per 100 cells electroporated, which is not low and is consistent with efficiencies previously reported [63].

In theory, the chances of two tk genes contained on a targeting construct becoming impaired during transfection should be quite low, or certainly lower than the probability of one tk gene becoming damaged. Mansour et al. [63] report that placing two tk genes on a construct can in fact be effective at increasing the enrichment for targeting by PNS, although they do not show the data. (In the work involving targeting the int-1 gene described above, the use of two tk genes was apparently not very helpful.)

Mansour et al. [63] were somewhat surprised to learn that placing large heterologies (i.e., two tk genes) at the termini of a targeting construct does not interfere with homologous recombination. Thomas et al. [16] had previously demonstrated that linear DNA molecules are better targeting substrates than are circular DNA molecules. Since blocking the ends of a linearized construct with non-homology did not reduce the frequency of homologous recombination, Mansour et al. [63] inferred that linearization does not increase targeting efficiency by providing homologous ends for invasion of the target. Instead, they concluded that linearization probably provides a substrate that is better-suited topologically for recombination. That linear molecules are topologically better substrates for homologous recombination was previous demonstrated in studies of extrachromosomal homologous recombination in mammalian cells [68].

**VII-C. Other enrichment methods**

The PNS method is a clever method for enrichment of targeted disruption of genes for which there is no selection and for genes that may be totally unexpressed. Another methodology that is not dependent upon any target gene selection has been used successfully for enrichment of targeted disruptions into transcriptionally active loci. The general scheme involves the transfer of a selectable marker gene into a locus by targeting. The marker gene on the targeting construct is transcriptionally impaired, however, and will become transcriptionally active if it inserts into the target locus by homologous recombination. In this way, selection for the marker gene is effectively a positive selection for gene targeting, although the selection is not dependent upon the function of the target sequence.

One of the earliest uses of this approach was reported by Jasin and Berg [69]. In this work, the selectable E. coli gpt gene was transferred, via homologous recombination, into the SV40 early region that is contained within the genome of COS1 cells. In the targeting construct, gpt expression was placed under the control of the SV40 early promoter, but the promoter was crippled due to the lack of an SV40 enhancer. Targeting of the gpt gene into the SV40 early region in the COS1 genome placed the gpt gene immediately downstream from a functional SV40 promoter/enhancer, allowing expression of gpt. Random integration would likely not allow expression of gpt for lack of a functional promoter. The scheme worked well, and greater than 50% of the recovered gpt\(^{+}\) colonies were successfully targeted. The absolute efficiency of targeting was about $1 \times 10^{-5}$, using the calcium phosphate coprecipitation method of transfection.

A point of considerable interest in this work was that targeting events could only be recovered with high efficiency if the input DNA was cut in a region sharing homology with the target sequence. When the input DNA was cut in nonhomologous vector sequences, no targeting events were recovered. This observation is consistent with the double-strand break repair model that was originally described for recombination in yeast [70], but is in direct conflict with the finding of Mansour et al. [63] that nonhomologous sequences on the ends of input DNA did not impede gene targeting (see above). It is not clear how to reconcile these differing findings other than to say that they reflect differences in the sequences targeted, cell types used, or modes of transfection. Jasin and Berg also found that increasing the amount of homology between input DNA and target did not significantly effect the targeting frequency, a finding in possible conflict with earlier findings of Thomas and Capecchi [38].

More recently, Jasin et al. [71] have described a gene targeting protocol they have termed 'epitope addition'. In this work, a crippled murine Thy-1 gene was targeted to the endogenous CD4 locus of a human T-cell line. The targeting construct consisted of part of the mouse Thy-1 gene fused to part of the human CD4 gene. The CD4/Thy-1 fusion gene lacked a promoter and translation start. The targeting construct also contained the gpt selectable marker. By targeted recombination into the CD4 locus, the fusion gene would be expressed using the
endogenous CD4 promoter and translation start site. Furthermore, since CD4 is a cell surface protein, the fusion protein would be expressed at the cell surface because it would be linked to the leader sequence that directs the transport of the endogenous CD4 protein to the surface. The key point is that a gene targeting event would result in the expression of Thy 1 epitope at the cell surface, hence 'epitope addition'. The authors electroporated the cells and initially selected for gpt expression to select cells that had acquired the targeting construct. By the use of a fluorescently-tagged monoclonal antibody directed against Thy-1 in combination with fluorescence activated cell sorting (FACS), Jasin et al. were able to enrich the gpt+ cells for cells in which the construct had integrated into the CD4 locus by targeting. Seventy percent of the cells scored positive for Thy-1 expression by FACS were found to be correctly targeted. The authors estimated that the scheme resulted in a 100-fold enrichment for targeting events, with the ratio of homologous to nonhomologous integrations estimated at 1 in 900. They suggested that epitope addition may be useful for targeting other loci. In the case of a locus that does not encode a cell-surface marker, a crippled cell-surface marker can be introduced together with its own leader sequence. Although this technique is elegant, it is likely that more conventional approaches will be used for most targeting applications.

Sedivy and Sharp [53] used a crippled neo gene to disrupt the polyoma middle T (pmt) antigen gene in NIH 3T3 cells. A construct was used that contained the neo gene as an in-frame fusion to the coding region of the pmt gene. The fusion gene contained no promoter and would be expressed if it correctly targeted the pmt locus where it would acquire a promoter as well as a translation start site. This particular scheme worked with limited success, with the most successful attempt yielding a targeting event in 1 out of 100 G418' cells recovered after transfection and the least successful attempt yielding a targeting event in 1 out of 3000 G418' clones. The most successful attempt used gel-purified DNA fragments. The authors estimated that the scheme brought about a 100-fold enrichment in targeting events. They suggest that the ability to enrich for targeting events is somewhat limited by damage and fragmentation that occurs to DNA during transfection and suggest that negative selections against illegitimate integrations will be limited due to DNA damage as well. A similar conclusion was reached by Mansour et al. [63] in the development of the PNS strategy, as discussed above. Schwartzberg et al. [55] disrupted the c-abl locus of ES cells with a fusion gene consisting of a neo gene fused in-frame to the c-abl gene. Again, the promoterless marker gene would be expressed if it were correctly integrated into the target c-abl locus where it would be transcribed from the endogenous c-abl promoter. About 1 in 34 of G418' clones recovered exhibited a targeted gene replacement. The authors estimated that their procedure brought about a 100-fold enrichment, which means that the ratio of unselected homologous to nonhomologous integrations was about 1 in 3400.

VIII. Knocking-out both alleles of a gene by gene targeting

In order to determine the physiological role of a particular gene product, it may be necessary to knock-out both copies of a gene. te Riele et al. [72] successfully used selectable fusion genes to knock-out both alleles of the pim-1 proto-oncogene in vitro. Two targeting constructs were used. One construct contained a promoterless neo gene fused in-frame with the pim-1 coding sequence so that expression of neo could be driven by the pim-1 promoter and translation start site after targeting the pim-1 locus. The other construct was very similar except that it contained a defective hygromycin-resistance (hyg) gene fused to the pim-1 coding region. te Riele et al. [72] were successful at knocking-out a single pim-1 allele with either construct. Their success was in fact quite impressive: 85% of the neo' clones and 64% of the hyg' clones recovered were successfully targeted. Curiously, the enrichment factor in this targeting scheme appeared to be only 20-fold (that is, the use of targeting constructs in which the fusion marker genes contained their own pim-1 promoters increased colony recovery only 20-fold). This would imply that in the absence of enrichment perhaps 4% of all neo' or hyg' clones would represent targeting events at the pim-1 locus. If this were so, then the pim-1 locus might be quite a hotspot for targeted homologous recombination. Unfortunately there were too few colonies examined in the control experiments to judge what percent were targeted in the absence of enrichment.

te Riele and co-workers isolated a clone that had a neo disruption in one of the pim-1 alleles and electroporated into this clone the hyg targeting construct. By selecting for hyg' clones, they attempted to identify clones that had both alleles disrupted, one disrupted with the neo insert and one disrupted with the hyg insert. Almost 50% of the hyg' clones obtained in the double-knockout attempt were in fact found to contain the desired double-knockout. This is the first report of targeted knockout of both alleles of a gene in mammalian cells in vitro.

More typically, breeding of a targeted disruption to homozygosity would likely be used once chimeric mice are produced that transmit the targeted disruption into
the germ-line (see Thomas and Capecchi [44], and Zijlstra et al. [73] for example). It should of course be recognized that knockout of both alleles of a gene might be lethal, and so homozygous mutants might never be attainable. (For example, when a targeted disruption of the int-1 locus was bred to homozygosity, some mice died before birth while some survived with severe ataxia [44].)

IX. Additional approaches

IX-A. Direct injection into mouse eggs

With the use of ES cells so common in the current field of gene targeting, one might believe that there is no other approach to derive a mouse containing a targeted genomic alteration. Brinster et al. [54] have shown that this is not necessarily true. Motivated by the observation [42] that it is non-trivial to derive chimeric mice derived from ES cells that transmit the targeted genomic change to the germ line, Brinster et al. [54] examined the possibility that direct injection of DNA into mouse eggs might be a feasible approach in gene targeting. The logic was simple. In the construction of transgenic mice, in which DNA injected into fertilized eggs usually randomly integrates into the genome, the transgene is typically found to be represented in the germ line. If gene targeting would occur at a frequency approaching 1% of the eggs injected, Brinster and co-workers reasoned that injection of DNA into fertilized mouse eggs might provide a realistic methodology for producing mice carrying targeted alterations in the germ line.

Brinster et al. [54] used mice containing a deletion in the major histocompatibility (MIIC) class II Eα gene for their studies. Eggs from such mice were injected with various DNA molecules capable of correcting the MHC deletion via gene targeting. Out of 506 transgenic mice produced, a single mouse was found to have undergone correction of the Eα gene deletion by gene targeting. The analysis involved a brute-force approach using Southern blotting.

It was not possible to obtain an accurate frequency of gene targeting from the work of Brinster et al., but targeting was clearly fairly rare (1 out of 506 mice). The work did, however, demonstrate that producing an animal with targeted alterations by direct injection of DNA into fertilized eggs is in fact possible. The authors pointed out that the inefficiency of targeting in their particular system could have been due to the absolute lack of expression of the MHC gene target (the genomic deletion encompassed the gene's promoter). However, as discussed above, it does not appear to be generally true that gene expression affects targeting efficiency. Another possible reason for the low targeting efficiency obtained was the large number of heterologies between the target sequence and the input DNA used. To this day, there has been no systematic study of the effect of scattered sequence heterologies on the frequency of targeting, although it has been shown [50] that small amounts of heterology can have a profound effect on intrachromosomal homologous recombination.

Worth noting was the fact that the targeted transgenic mouse did not contain a long head-to-tail array of the injected DNA in the genome, which is so common in transgenic animals containing randomly integrated DNA. The correction of the deletion in the Eα gene was the result of a double crossover or gene conversion. The authors reported that there were many point mutations introduced into the corrected locus. The investigators acknowledge that a high mutation frequency would mitigate the usefulness of homologous recombination for gene therapy (as discussed above).

IX-B. Introducing subtle changes by gene targeting

IX-B.1. An 'in-out' targeting procedure

Almost all selection schemes for the enrichment of targeted gene disruptions have one limitation in common; the schemes result in the obligatory placement of a marker gene sequence into the target gene. If one desires to introduce a 'subtle' change into a genomic sequence, the schemes will not suffice. Of course, one could use a brute-force procedure of screening by PCR alone with no selection applied whatsoever, as was used by Zimmer and Gruss [58] to place a 20 base pair insertion into the Hox1.1 gene; however, it is worth noting that the work of Zimmer and Gruss remains the only report of successful targeting by PCR screening alone. As described above, investigators who have used PCR screening to isolate targeted clones have usually combined the screen with an initial selection for the marker neo gene to enrich for genomic integrations prior to applying PCR.

Seeing the need to devise a way of introducing subtle changes into genes by targeting, Valancius and Smithies [74] have recently tested an 'in-out' procedure for gene targeting. The gene that was studied was the hprt locus of an ES cell line that contained an hprt gene with a deletion of the promoter and the first two exons. The authors wanted to introduce a 4 bp insertion into the second intron of the hprt gene. To accomplish this, they used a construct containing hprt sequences carrying the desired 4 bp insertion. The construct contained a functional promoter as well as sequences through the first three exons of the hprt gene, but was lacking exons 4–9. Prior to electroporation into ES cells, this construct was
linearized within the exon 3 sequences so that the construct would insert into the third exon of the target locus upon targeting. The ES cells were electroporated with the construct and this first step of the procedure was dubbed the 'in' step. As a result of the in step, the targeting vector was integrated at the hprt locus, resulting in a duplication of the 5 kb of sequence shared by the target and the targeting construct. This duplication flanked the targeting construct and this first step of the procedure was directly selectable in HAT medium. Such targeted integrations were recovered at a frequency of about $2.8 \times 10^{-6}$ per electroporated cell.

The second step of the procedure, the 'out' step, involved excision from the genome of the vector sequences that had integrated in the in step. The excision event occurred by spontaneous homologous recombination between the duplicated hprt sequences. Excision of the integrated construct reverted the cell back to the hprt- phenotype since excision removed the vector-derived promoter and first two exons of the hprt gene. Since the hprt- phenotype is selectable in 6-TG, the excision was directly selectable. Importantly, depending on the precise location of the crossover between the duplicated hprt sequences, the 4 bp insertion would either remain in the genomic hprt sequence or would excise from the genome along with the remainder of the targeting construct. Hprt- clones were isolated at a frequency of $8 \times 10^{-7}$ per HAT+ cell plated into 6-TG. Out of 20 TG+ clones examined by Southern analysis, 19 retained the 4 bp insertion in the genomic hprt sequence. Thus, Valancius and Smithies were able to introduce a subtle change into the hprt locus.

What is needed to make the in-out procedure useful for the introduction of a subtle change into any locus? The key feature was the ability to select both for hprt expression (the in step) and against hprt expression (the out step). To use the in-out method therefore requires the use of an hprt- cell line. The targeting vector containing the desired subtle change to the target sequence would carry an hprt gene within vector sequences so that any integration (targeted or random) of the vector would be selectable in HAT. One might expect that about 1 in 1000 HAT+ cells would be correctly targeted, and so targeted cells should in theory be identifiable by PCR screening of HAT+ cells. Once a targeted clone is identified, the clone can be expanded and then selected in 6-TG to select for cells that excised the targeting construct (and, thus, have lost the hprt gene) due to homologous recombination between the duplicated target sequences. Clones retaining the subtle change in the genome would be identified by further PCR or Southern analysis.

**IX-B.2. Selectable subtleties**

Of course, one can fairly readily introduce a subtle change into a genome by targeting if the subtle change is selectable. Steeg et al. [75] were able to introduce two point mutations (subtle changes indeed) into the gene that encodes the largest subunit of murine RNA polymerase II in such a manner. A selection scheme was used in this study since one of the two point mutations conferred resistance to the mushroom toxin α-amanitin. Absolute and relative targeting frequencies were reportedly quite high (as high as 1 targeting event per $3.3 \times 10^4$ ES cells surviving electroporation, and one targeted cell per 30 that expressed stably integrated DNA). The authors suggest that the great degree of homology between the transfected DNA and the target (only two single mismatches) can explain the high efficiency in comparison to that obtained in other studies in which large heterologies are typically introduced into the target sequence. However, as discussed above, work by Mansour et al. [49] has shown that the frequency of targeting is not a function of the amount of heterology transferred during targeting. Therefore, the high efficiency of targeting at the RNA pol III locus likely has some other explanation.

Of interest in the study by Steeg et al. [75] was the observation that the majority of the cells that acquired α-amanitin resistance by targeting did not pick up the second point mutation that was only 20 bp away from the selectable point mutation on the targeting construct. The authors surmise that this indicates that the targeting event occurred via heteroduplex formation and subsequent mismatch repair and that the two mismatches present in the heteroduplex intermediate were corrected independently. Curiously, the authors 'contrast' this mechanism with gene conversion, which they say may explain the 30–40% of the events in which both point mutations were transferred to the genome during targeting. However, gene conversion is commonly conceptualized as arising from mismatch repair of heteroduplex (see Orr-Weaver and Szostak [76] for a review) and, thus, can sufficiently explain 100% of the events. Alternatively, in contrast to the authors’ argument, it is not necessarily unlikely that a crossover may occur within the 20 bp interval between the two mutations. There is virtually no data on the qualitative effects of small amounts of mismatch on homologous recombination in mammalian cells. It is possible that low levels of mismatch may promote crossing over.

**X. Summary and future perspectives**

Many clever tricks have been devised that allow investigators to recover rare gene targeting events. These
tricks have been applied quite successfully and it is not unreasonable to assume that any investigator who is patient enough can now knock-out any gene in cultured mammalian cells by gene targeting, as long as that gene (or part of the gene) has been cloned. However, there has been essentially no progress in finding ways of significantly improving global targeting efficiency. This is largely due to the lack of any systematic studies into the mechanism(s) of targeting. Such studies are needed. Absolute efficiencies are commonly about 1 in 10^6 cells transfected. Random integration of DNA is still the rule in most cases and the ratio of targeted to random integration is still often no better than 1 in 1000. If gene targeting is ever to be used for its most important potential application, human gene therapy to cure genetic disease, both its absolute and relative efficiencies must be improved.

How can gene targeting efficiency be improved? One possible approach to improve targeting efficiency might be the use of synchronized cells. It has been shown that extrachromosomal homologous recombination between transfected linear DNA molecules peaks at early to mid S-phase [77]. Perhaps gene targeting peaks at a certain point in the cell cycle as well. This hypothesis remains to be tested.

Another possibility is that using very large pieces of genomic DNA as input may significantly increase targeting efficiency. As discussed earlier in this review, there is some evidence that targeting may exhibit a strong dependence on homology length. With the use of yeast artificial chromosomes ("YACS"), it should be possible to work with exceptionally large segments of mammalian genomic DNA (say greater than 50 kb). Targeting attempts with such long sequences may prove to be fruitful.

A distinct approach to improve targeting efficiency would be to block random integration of DNA into the genome. This may channel more DNA into targeting. As discussed above, there is evidence in the literature [16,25,26] suggesting that the initial encounter between input DNA and the target is not rate-limiting. If this were indeed true, then random integration does not compete with targeting and therefore blocking random integration would not improve the absolute targeting efficiency. However, the issue as to precisely what limits the rate of targeting is certainly not resolved, so all reasonable approaches are worth a try. At the very least, blocking random integrations would eliminate or reduce the background of unwanted random integration events.

How may one block random integration? One approach that was considered by Chang and Wilson [78] was to add dideoxynucleotides to the termini of DNA molecules that were introduced into mammalian cells. They were able to show that such an addition effectively blocked extrachromosomal end-joining of transfected molecules while allowing extrachromosomal homologous recombination to occur. Since random integration of DNA is typically conceptualized as an end-joining process [14], the addition of dideoxynucleotides to the termini of transfected molecules may effectively block random genomic integration. Whether this is indeed so has not been reported.

Another approach to block random integration also makes use of the notion that random integration is an end-joining process but takes advantage of the fact that poly(ADP-ribosylation) is known to play a role in DNA end-joining. Farzaneh et al. [79] were able to show that treatment of cells with an inhibitor of poly(ADP-ribose)polymerase during transfection inhibited the ability of a variety of mammalian cells to randomly integrate DNA into their genomes. Waldman and Waldman [80] extended those studies by showing that 3-methoxybenzamide (3-MB), a potent competitive inhibitor of poly(ADP-ribose)polymerase, specifically inhibited random integration of DNA into the genome of mouse Ltk^- fibroblasts but did not inhibit extrachromosomal homologous recombination among the transfected DNA molecules. The observations held true for both linear as well as circular transfected molecules and for both inter- and intramolecular homologous recombination events. Waldman and Waldman have also determined that treatment of mouse fibroblasts with 3-MB stimulates intrachromosomal homologous recombination (unpublished observations). Taken together with the observations that 3-MB inhibits neither DNA uptake nor gene expression [79,80], it is possible that treatment of cells with 3-MB or similar compounds might be experimentally useful in gene targeting protocols.

On a different note, examination of apparent gene targeting hotspots might yield insight into what governs the rate of targeting. Certain loci, such as Hox 1.1 [58], β2-microglobulin [43,60], or pim-1 [72] appear to undergo targeting at high frequencies compared to many other loci. Perhaps by a careful study of the sequence, structure, or function of such loci, we may uncover certain features that make these loci different from others, why they are hotspots. Perhaps we may be able to apply whatever insight is gained to increase the targeting efficiency at any locus.

A problem other than inefficiency that will have to be overcome if gene targeting can be useful for gene therapy is the elimination of the mutagenic events and rearrangements at the target locus that are sometimes associated with targeting. It is not presently clear how this may be addressed, or if such mutations are intrinsic to the mechanism of gene targeting.
In summary, current technology allows gene targeting in mammalian cells to be used as a powerful genetic tool for the study of the structure and function of virtually any gene and, through the creation of chimeric mice from targeted ES cells, for the creation of animal models for genetic disease. The usefulness of gene targeting as a means of human gene therapy to cure disease awaits a greater understanding of the underlying mechanism(s) of the process.

References

3 Capecchi MR. Altering the genome by homologous recombination. Science 244: 1288, 1989.
39 Doetschman T, Gregg RG, Maeda N, et al. Targeted correction

Aragen/Transposagen Ex. 1012


