Cell, Vol. 44, 419–428, February 14, 1986, Copyright © 1986 by Cell Press

High Frequency Targeting of Genes to Specific Sites in the Mammalian Genome

Kirk R. Thomas, Kim R. Folger, and Mario R. Capecchi
Department of Biology
University of Utah
Salt Lake City, Utah 84112

Summary

We corrected a defective gene residing in the chromosome of a mammalian cell by injecting into the nucleus copies of the same gene carrying a different mutation. We determined how the number, the arrangement, and the chromosomal position of the integrated gene, as well as the number of injected molecules influence the gene-targeting frequency. Recombination between the newly introduced DNA and its chromosomal homolog occurred at a frequency of 1 in 10^3 cells receiving DNA. Correction events were mediated by either double reciprocal recombination or gene conversion. This resulted in sequences in the genome being replaced by sequences of the introduced DNA or, in separate experiments, sequences in the incoming DNA being replaced by chromosomal sequences. Both point mutations and deletion mutations were corrected; however, the nature of the mutation carried by the respective sequence influenced whether the integrated or injected sequence was corrected.

Introduction

Homologous recombination between DNA sequences residing in the chromosome and newly introduced DNA sequences, an event termed gene targeting, could provide the means for introducing specific mutations into the mammalian genome. In vitro mutagenesis could generate the desired mutation in a cloned DNA sequence, and gene targeting would introduce the mutation into the genome.

Following their introduction into mammalian cells, DNA molecules do recombine with each other through shared homologous regions (Folger et al., 1982; Miller and Temin, 1983; Robert de Saint Vincent and Wahl, 1983; Shapira et al., 1983; Small and Scangos, 1983; Subramani and Berg, 1983). A surprising feature of these reactions is that they occur at a frequency approaching 100% (Folger, Thomas, and Capecchi, 1985). Thus, somatic mammalian cells must possess the enzymatic machinery for efficiently mediating homologous recombination. The high efficiency suggests that newly introduced DNA is intrinsically highly recombinogenic; therefore, it is reasonable to suspect that recombination between newly introduced DNA and homologous chromosomal sequences also occurs.

Such homologous recombination between exogenous sequences and homologous sequences in the mammalian chromosome has been detected for viral DNA (Vogel, Glutzman, and Winocour, 1977; Vogel, 1980; Bandyopadhyay, Watanabe, and Temin, 1984; Shaul et al., 1985). Preliminary evidence for recombination between introduced cloned DNA sequences and the corresponding sequences in mammalian chromosomes has also been reported (Lin, Sperle, and Sternberg, 1985; Smith and Berg, 1984). In these studies the DNA was introduced by calcium phosphate coprecipitation and the frequency of recombination between the newly introduced and integrated plasmid sequences was approximately 1 in 10^8–10^9 transfected cells. When DNA is introduced into cells by calcium phosphate coprecipitation, most of the foreign DNA is incorporated into the host genome at random sites by mechanisms other than homologous recombination. Stable transfectants are obtained at a frequency of approximately 1 per 10^4 cells receiving DNA. If this frequency is used as a measure of the frequency of nonhomologous or illegitimate recombination, the estimated ratio of legitimate to illegitimate recombination is approximately 10^{-4}–10^{-5}. In this paper we report on homologous recombination between a newly introduced DNA sequence and the corresponding sequence residing in the mammalian genome occurring at a frequency of 1 per 10^3 cells receiving DNA. The ratio of legitimate to illegitimate recombination is 10^{-2}.

To monitor a gene-targeting event, we first established cell lines containing a mutant neomycin resistance gene (Neo^r) integrated into the host genome, and then we sought to specifically restore the gene via homologous recombination by injecting plasmid DNA carrying a different mutation in the Neo^r gene. Productive gene-targeting events were identified by selecting for cells resistant to the drug G418. We characterized the corrected Neo^r gene by Southern transfer analysis and by rescuing it from the mammalian genome.

Results

Recombinant Plasmids

In Figure 1 we illustrate the recombinant plasmids used for the gene-targeting experiments. The parental plasmid pRH140/TK contains sequences derived from the bacterial plasmid pBR322, the Herpes simplex virus thymidine kinase gene (HSV-tk), and the Neo^r gene from the bacterial Tn5 transposon. The pBR322 sequences supply an ampicillin resistance gene (Amp^r) and an origin of DNA replication that functions in bacteria. The Neo^r gene is transcribed from a promoter isolated from a long terminal repeat of an avian retrovirus (RSV-LTR) that functions both in bacteria and in mammalian cells (Hudziak et al., 1982). To allow expression of the Neo^r gene in mammalian cells, a poly(A) addition signal, derived from the HSV-tk gene, was added to the 3′ terminus of the gene. Transfection of bacteria with pRH140/TK DNA confers ampicillin and kanamycin resistance. DNA-mediated gene transfer of pRH140/TK DNA into a thymidine kinase-deficient mouse cell line, $LMtk^-$, renders the cells resis-

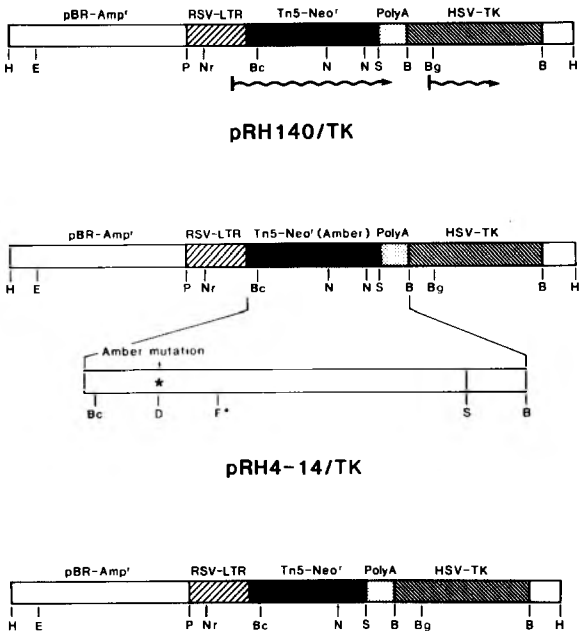

Figure 1. Maps of the Three Neo^r/Thymidine Kinase Plasmids, pRH140/TK, pRH4-14/TK, and pRH140ΔNae/TK

The three vectors were derived from pRH140 (Hudziak, 1982). The RSV-LTR provides an enhancer (Luciw et al., 1983) and a bifunctional promoter which allows expression of the Neo^r gene in both E. coli and mammalian cells. pRH4-14/TK contains an amber mutation in the Neo gene. pRH140ΔNae/TK contains a 284 bp deletion at the 3′ end of the Neo gene. The HSV-TK gene was inserted into the above vectors at the unique Bam HI site in the pBR322 sequence as a 3.6 kb Bam HI fragment. Each of the recombinant plasmids is represented in linear form from the unique Hind III site. The Bcl I site is 73 bp 5′ to the amber mutation. The restriction sites are as designated: H, Hind III; E, Eco RI; P, Pvu II; Bc, Bcl I; Nr, Nru I; N, Nae I; S, Sma I; B, Bam HI; Bg, Bgl II; D, Dde I and F, Fnu 4HI. The small asterisk next to the F indicates the loss of this site. PolyA, polyadenylation sequence from HSV-tk.

tant to the drug, G418, and allows growth in HAT medium.

Two derivative plasmids were used to study homologous recombination. One of these, pRH4-14/TK, contains an amber mutation in the Neo^r gene. This premature polypeptide chain termination signal renders the gene product defective in both bacteria and mammalian cells (Hudziak et al., 1982). The mutation that created the amber mutation concomitantly created a new Dde I site, which could be used as a diagnostic test for the presence of the amber mutation. The pRH14-14/TK also contains a silent point mutation in the Neo^r gene, resulting in the destruction of an Fnu 4HI restriction enzyme recognition site 59 bp downstream of the amber mutation. The Fnu 4HI site was used as a marker, independent of the amber mutation, for following the transfer of information to or from the pRH4-14/TK sequence.

The second derivative plasmid pRH140ΔNae/TK contains a 284 bp deletion at the 3′ end of the Neo^r gene which removes 52 amino acids from the carboxy-terminal end of the protein. pRH140ΔNae/TK was obtained by digesting pRH140 with the restriction endonuclease Nae I, religating, and cloning in E. coli. The HSV-tk gene was then cloned into the unique Bam HI site of pRH140ΔNae.

Recipient Cell Lines

We have used three recipient cell lines designated LM1, LM4, and LM5. They differ in the number of integrated plasmids that they contain, the arrangement of the plasmids within the host chromosome, and whether they contain pRH4-14/TK or pRH140ΔNae/TK. Each was derived by transforming LMtk$^-$ cells to tk$^+$ by injecting either pRH4-14/TK or pRH140ΔNae/TK plasmid DNA linearized with Hind III and selecting for HAT^R cells.

LM1 contains a single copy of the pRH4-14/TK plasmid integrated into chromosomal DNA by its Hind III ends. The cell line LM4 contains four copies of the pRH4-14/TK plasmid integrated by their Hind III ends at four different sites in the mammalian genome. Our best estimate is that the cell line LM5 contains five copies of the pRH140ΔNae/TK plasmid. They are integrated at a single site in the mouse chromosome as a head-to-tail concatemer. A schematic diagram of the integration pattern of the pRH4-14/TK and pRH140ΔNae/TK plasmids in LM1, LM4, and LM5 is illustrated in Figure 2. The organization of these plasmids in the host chromosome was deduced from Southern transfer analysis and from plasmid rescue.

For the rescue experiments, genomic DNA was digested with Bam HI, Hind III, or Bgl II and ligated under conditions that favored intramolecular ligation. This DNA was used to obtain Amp^r bacteria by transfection. From LM1 we rescued a single class of plasmid, which contained the 5′ junction with chromosomal DNA. From LM4 we rescued four classes of plasmids, designated J-1 through J-4. The chromosomal junction regions were shown to be from four independent integration sites by the absence of cross-hybridization. From LM5 we rescued two classes of plasmids, one containing the 5′ junction with chromosomal DNA, the other being of monomer plasmid length. Having rescued the plasmids from the genome, they were analyzed in detail to determine if the plasmids were altered in any way in the process of transfection into LMtk$^-$ cells. This is an important control since a number of investigators have shown that as much as 10%–15% of the plasmid DNA transferred into mammalian cells by calcium phosphate coprecipitations, DEAE-dextran, or protoplast fusion suffer mutations (Calos, Lebkowski, and Botchan, 1983; Razzaque, Mizusawa, and Seidman, 1983; Wake et al., 1984). Restriction enzyme analysis of the rescued plasmids using Hind III, Bam HI, Bgl II, Eco RI, Fnu 4HI, and Dde I agreed with the expected digestion patterns. We also sequenced the regions surrounding the 4-14 amber mutations and did not find a single base-pair change compared with the input, pRH4-14/TK, plasmid.

LM1, LM4, and LM5 are extremely stable. We have grown them in selective and nonselective medium for periods of up to 3 years without observing any changes in their Southern transfer pattern. The spontaneous reversion rate of LM1 and LM4 to G418^r is approximately $1/10^8$ cells/generation. We have never observed a spontaneous G418^r revertant of LM5 (i.e., $<1/10^9$ cells/generation). The

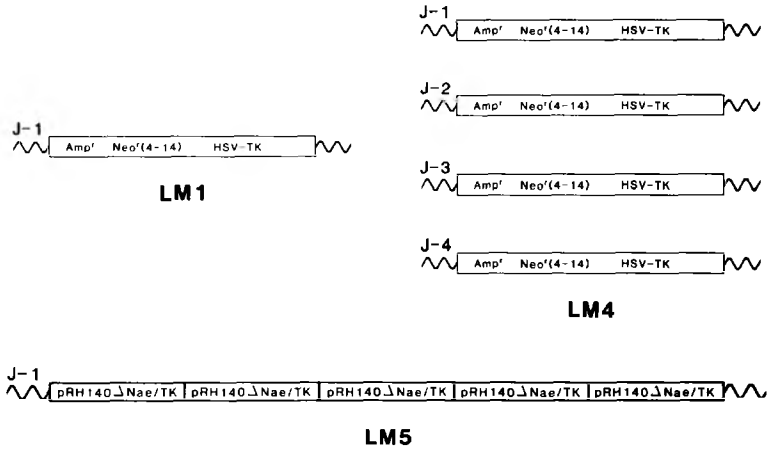

Figure 2. Integration Patterns of the pRH4-14/TK or pRH140ΔNae/TK Plasmids in the Recipient Cell Lines LM1, LM4, and LM5

LM1 contains a single copy of the pRH4-14/TK plasmid integrated into the host chromosome by its Hind III ends. The cell line LM4 contains four copies of the pRH4-14/TK plasmid integrated at four different sites in the mammalian genome, designated J-1 through J-4. Our best estimate is that the cell line LM5 contains five copies of the pRH140ΔNae/TK plasmid. They are integrated at a single site in the mouse chromosome as a head to tail concatemer.

latter is not unexpected since LM5 contains deletion mutants of the Neo^r gene. We have analyzed eight spontaneous revertants of LM4. All eight reverted by acquiring a suppressor rather than by a reversion of one of the 4-14 loci (Hudziak and Capecchi, unpublished).

Correction of a Deletion Mutation Residing in the Chromosome

To test for gene targeting, we injected 5–10 copies of pRH4-14/TK linearized with Bcl I into LM5 and selected for resistance to G418. G418^r cell lines arose at a frequency of 0.5–1 per 1000 cells receiving an injection. We did not obtain G418^r cell lines following control injections of linearized pRH140ΔNae/TK DNA (10,000 injections). Interestingly, we did not obtain G418^r cell lines following injection of LM5 with pRH4-14/TK supercoiled molecules (10,000 injections).

In Table 1 we show the number of G418^r cell lines obtained following injection with increasing numbers of linearized pRH4-14/TK molecules into LM5 (from 5 to 100 molecules per cell). Within the small sample size tested, 12,000 injections, we did not observe a correlation between the number of G418^r cell lines produced and the number of pRH4-14/TK molecules injected per cell.

Three of the G418^r cell lines, designated LM5-5, LM5-9, and LM5-11, were analyzed. In Figure 3 we illustrate Southern transfer patterns of these cell lines following digestion of genomic DNA with Bam HI. The intense band in the parental LM5 pattern corresponds to copies of the pRH140ΔNae unit length sequence, 5.5 kb long, generated by digestion of the pRH140ΔNae/TK head-to-tail concatemer with Bam HI. In the corresponding region of LM5-5, LM5-9, and LM5-11, we observe two bands 5.5 kb and 5.8 kb in length. The 5.8 kb band represents the wild-type size pRH140 unit length sequence. However, by this type of analysis we cannot distinguish between pRH140 and pRH4-14 since they contain the same number of base pairs. In the Southern transfer patterns of both LM5-5 and LM5-11 we also observe an additional new band, designated by an *n*, not present in the parental LM5 pattern. These bands correspond to a single copy of pRH4-14/TK integrated at a random site in the host chromosome. LM5-9 does not show an additional new band and presumably does not contain any copies of the injected pRH4-14/TK sequence. We have also analyzed these cell lines by Southern transfer following digestion of genomic DNA with Hind III (data not shown), and derived conclusions identical to those described above.

Table 1. Number of G418^r Cell Lines Obtained by Injecting LM5 Cells with pRH4-14/TK DNA

No. of DNA Molecules Injected Per Cell	No. of G418^r Cell Lines Obtained	No. of Cells Injected
5	2	3×10^3
10	3	3×10^3
25	1	3×10^3
100	2	3×10^3

LM5 cells were grown on glass cover slips (10 mm by 10 mm) in 35 mm petri dishes. Twenty-five cells per dish received nuclear injections with pRH4-14/TK linearized with Bcl I. After the injection the cells were incubated for 24 hr in nonselective medium at 37°C in a 5% CO_2 incubator and then switched to minimum essential medium supplemented with 400 μg of G418 per ml. After 2 weeks and then again after 3 weeks, the dishes were scored for large healthy colonies.

Digestion of genomic DNA with the restriction endonuclease Dde I yields a series of fragments that allow us to identify the presence of the wild-type Neo^r gene and each of the mutant alleles of Neo, including the 4-14 amber mutant, the ΔNae deletion mutant, and even a 4-14-ΔNae double mutant. As illustrated in Figure 4, the reason that we can distinguish the presence of these different alleles is that the 4-14 amber mutation creates a new Dde I site and the ΔNae deletion removes a Dde I site.

In Figure 5 we illustrate the Dde I Southern transfer patterns of LM5, LM5-5, LM5-9, and LM5-11. In the parental cell line, LM5, we observe only the Dde I fragment produced by the ΔNae deletion mutant. In addition to the ΔNae deletion fragment, each of the G418^r cell lines also contains the Dde I fragment produced by the presence of a wild-type Neo^r gene. Thus, in each of these cell lines one of the pRH140ΔNae/TK sequences residing in the

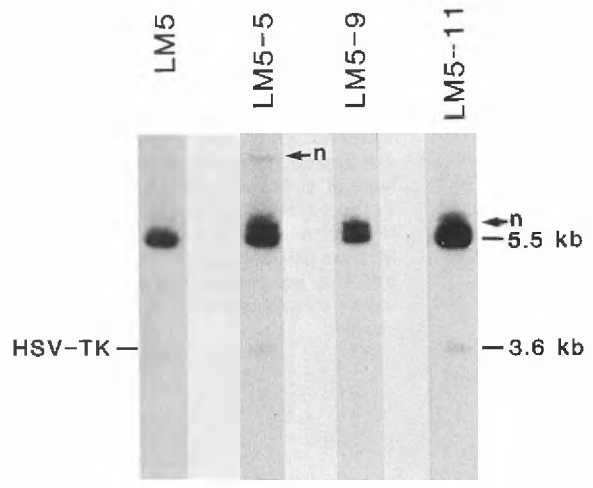

Figure 3. Bam HI Southern Transfer Patterns of LM5, LM5-5, LM5-9, and LM5-11

Five micrograms of genomic DNA from each cell line was digested with Bam HI, electrophoresed through 0.75% agarose, transferred to nitrocellulose paper, and probed with ^{32}P-labeled, nick-translated pRH140. The position of a monomer plasmid pRH140ΔNae, 5.5 kb, is shown as is the position of the 3.6 kb, Bam HI fragment of the HSV-tk gene.

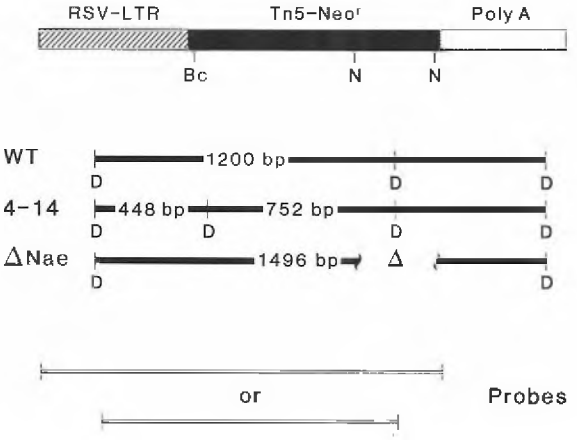

Figure 4. Lengths of the Neo^r Fragments of pRH140/TK, pRH4-14/TK, and pRH140ΔNae/TK Generated by Digestion with Dde I

In order to detect these fragments in a Southern transfer pattern we used as probe either a ^{32}P-labeled nick-translated 1465 bp Nru I–Sma I fragment of pRH140 or a ^{32}P-labeled nick-translated 1200 bp Dde I fragment of pRH140.

chromosome was corrected by recombination with the incoming pRH4-14/TK sequences. LM5-5 and LM5-11 also contain Dde I fragments produced by a 4-14 amber mutant. The 448 bp fragment is missing in LM5-11 because this fragment was disrupted when the pRH4-14/TK DNA integrated into the host genome by its Bcl I ends.

None of the G418^r cell lines contain a Dde I fragment 1048 bp long. Such a fragment would indicate the presence of a 4-14-ΔNae double mutant. The presence of a double mutant would be predicted if correction of the pRH140ΔNae/TK sequence resulted from a single reciprocal homologous recombination event between the 4-14 amber mutation and the ΔNae deletion mutation.

In order to extend our analysis of the G418^r lines, we rescued the integrated plasmid sequences from LM5, LM5-5, LM5-9, and LM5-11 (see Table 2). Analysis of the rescued plasmids provides an independent check of the results obtained by Southern transfer and permits a more detailed characterization of the recombination events. A wild-type, unit length pRH140 plasmid sequence was rescued from each of the G418^r cell lines. These sequences were indistinguishable from the parental pRH140 or pRH140/TK plasmid sequences when analyzed by a series of restriction enzymes (Bam HI, Hind III, Eco RI, Fnu 4HI, and Dde I). We also sequenced the region corresponding to the Nae I deletion and observed no differences in this region compared to the pRH4-14/TK and pRH140/TK sequences. Thus, in the process of correcting the pRH140ΔNae/TK sequence residing in the chromosome, the information donated by the incoming pRH4-14/TK plasmid was faithfully transferred.

As anticipated from the Southern transfer patterns, we did not rescue a DNA sequence containing the 4-14-ΔNae double mutation. The absence of the double mutant in the

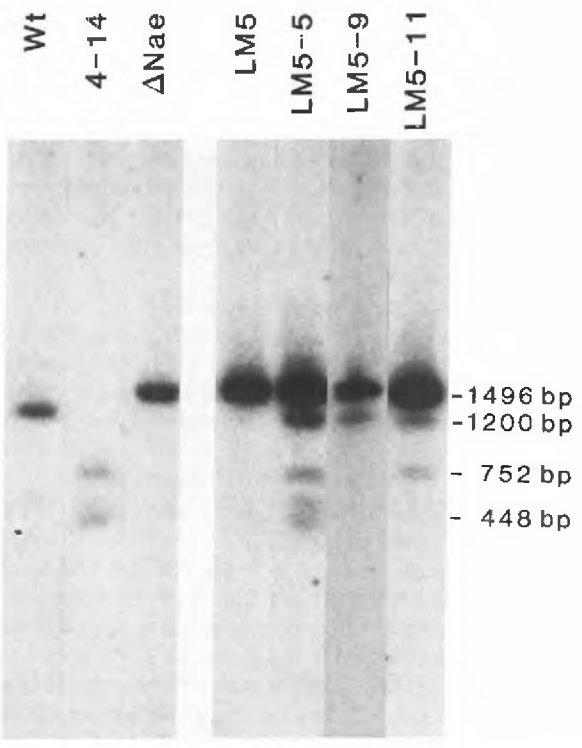

Figure 5. Dde I Southern Transfer Patterns of LM5, LM5-5, LM5-9, and LM5-11

DNA was purified from each cell line and digested with Dde I. Ten micrograms of each DNA sample was electrophoresed through 0.75% agarose, transferred to nitrocellulose paper, and probed with a ^{32}P-labeled, 1465 bp Nru I–Sma I, nick-translated fragment from the Neo^r gene (see Figure 4). As markers, purified plasmid DNAs (10 pg) were also digested with Dde I and processed with the cellular DNA. Wt, pRH140; 4–14, pRH4-14; ΔNae, pRH140ΔNae.

Table 2. Plasmids Rescued from LM5 and Its Derivatives

Cell Line	Locus	Kn^r	Dde I Polymorphism	Fnu 4HI Polymorphism	Number Rescued
LM5	unit length	–	ΔNae	wt	23
	J1	–	ΔNae	wt	1
	unit length	–	ΔNae	wt	30
LM5-5	unit length	–	4-14	4-14	11
	unit length	+	wt	wt	12
	J1				0
	J2	–	4-14	4-14	1
	unit length	–	ΔNae	wt	23
LM5-11	unit length	+	wt	wt	7
	J1	–	ΔNae	wt	1
	unit length	–	ΔNae	wt	16
LM5-9	unit length	+	wt	wt	3
	J1				0

DNA was purified from each cell line and digested with either Bam HI or Hind III. The digested DNA was ligated under conditions favoring intramolecular ligation and used to transform $CaCl_2$-shocked E. coli strain MH1. Transformants were selected for growth on LB agar containing ampicillin (100 μg/ml). Plasmid DNA was isolated from each transformant and analyzed for size and restriction polymorphisms, and tested for kanamycin sensitivity.

Locus: unit length refers to plasmids indistinguishable in length from parental plasmids, pRH140 and pRH140ΔNae, and thus representing internal members of a concatemer. J1 refers to plasmids containing chromosomal DNA corresponding to the junctions between plasmid DNA and chromosomal DNA in cell line, LM5. J2 refers to plasmids containing cellular DNA distinct from that found in J1, and thus representing another junction between plasmid and chromosomal DNA.

Kn^r: transformants were assayed for growth on minimal media agar containing kanamycin (100 μg/ml): + indicates growth; – indicates no growth.

Dde I polymorphism: plasmid DNA was digested with Dde I and electrophoresed through 1.5% agarose. wt indicates the presence of the 1200 bp fragment found in pRH140; 4-14 indicates the presence of the 752 bp and 448 bp fragments found in pRH4-14; ΔNae indicates the presence of the 1496 bp fragment found in pRH140ΔNae.

Fnu 4HI polymorphism: plasmid DNA was digested with Fnu 4HI and electrophoresed through 6% polyacrylamide. wt indicates the presence of a 90 bp fragment found in both pRH140 and pRH140ΔNae; 4-14 indicates the presence of a 101 bp fragment found in pRH4-14 (Folger, Thomas, and Capecchi, 1985).

The efficiency of rescuing a given DNA sequence from a mammalian cell line is not directly proportional to the representation of that sequence in the genome. Other factors, such as the length of the sequence influences this efficiency (Hanahan, 1983). In the above experiments, we observe that the J1 sequence, 12 kb in length, is rescued only 1/23 as frequently as the ΔNae unit length sequence, 5.5 kb in length. On the basis of the number of copies present in the genome, we would have expected the ratio of J1 to ΔNae unit length sequences rescued from LM5 to be 1:5.

concatemer argues against the generation of the wild-type Neo^r gene by a single reciprocal recombination event. It argues instead that the wild-type gene was generated either by a nonreciprocal recombination event, such as gene conversion, or by a double reciprocal recombination event followed by the subsequent loss of the free plasmid DNA carrying the double mutation.

From LM5-5 we rescued two classes of 4-14 plasmid sequences. One of these, containing chromosomal sequences, was probably derived from the random integration of an injected pRH4-14/TK plasmid into the genome. The second 4-14 plasmid was of unit length. We offer two hypotheses to account for the existence of this unit length 4-14 plasmid in LM5-5. After injection of LM5 with pRH4-14/TK, a head-to-tail dimer of pRH4-14/TK may have integrated at a random site in the host genome and the unit length plasmid represents the 3′ member of the dimer. Alternatively, one of the pRH140ΔNae/TK sequences in the LM5 concatemer may have been completely converted to pRH4-14/TK.

Correction of the Incoming Plasmid by the Sequence Residing in the Chromosome

Following injection of LM1 cells with 5–10 copies of Bcl I linearized pRH140ΔNae molecules, we obtained G418^r cells at a frequency of 0.5–1 per 1000 cells receiving an injection. The observed frequency of G418^r is five orders of magnitude greater than the spontaneous reversion frequency of LM1 to G418^r. No G418^r cell lines were obtained following injection of 10,000 LM1 cells with pRH4-14 DNA linearized at Bcl I.

We have analyzed a number of G418^r cell lines derived from LM1 and they fall into two classes. In the first class a wild-type gene was generated by homologous recombination between the introduced plasmid, pRH140ΔNae, and the chromosomal sequence, pRH4-14/TK. An example of such an event, as described below, is seen in the cell line LM1-3. In the second class a pseudo wild-type gene is generated by recombinational mutagenesis resulting from incorrect repair of a heteroduplex formed between the incoming pRH140ΔNae sequence and the chromosomal pRH4-14/TK sequence. In the latter class, the chromosomal pRH4-14/TK sequence still retains the amber mutation but has acquired a compensating frameshift mutation which allows the gene to function with near wild-type activity. The two classes arise at approximately the same frequency, 1 per 1000 cells receiving an injection. We will describe how the insertion of base pairs into the chromosomal pRH4-14/TK sequence occurred, how these events were mediated by the injected

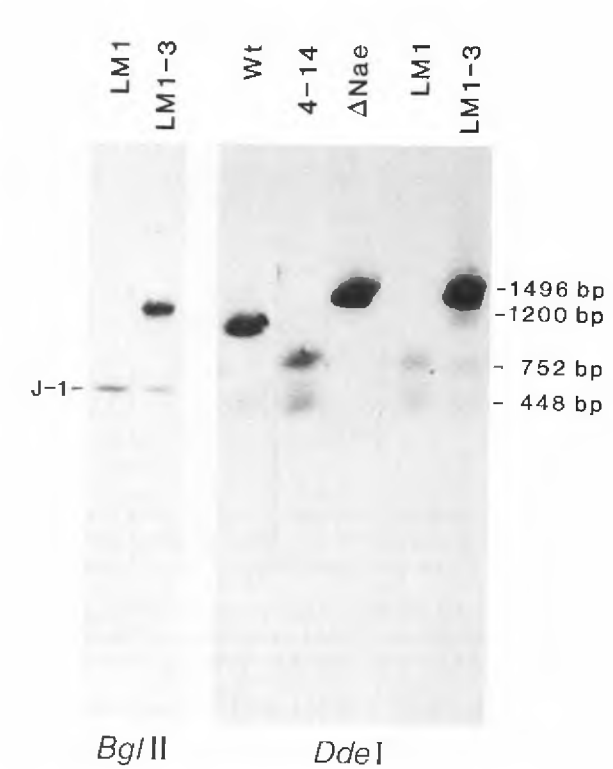

Figure 6. Southern Transfer Analysis of DNA from LM1 and LM1-3 Digested with Bgl II or Dde I

DNAs from LM1 and LM1-3 were digested with Bgl II or Dde I and electrophoresed through 0.75% agarose and transferred to nitrocellulose paper. The Bgl II Southern transfer was probed with ^{32}P-labeled, nick-translated pRH140. The Dde I Southern transfer was probed with a ^{32}P-labeled, 1465 bp Nru I–Sma I fragment with the Neo^r gene. The markers are as described in Figure 5.

pRH140ΔNae sequence, and how the Neo^r genes containing both the amber mutations and the frameshift mutations function with near wild-type activity in a separate communication (submitted).

In Figure 6 we illustrate the Southern transfer pattern of LM1 and LM1-3 genomic DNAs digested with Bgl II. In LM1 we see a single band (J1). In LM1-3 we observe the J1 band and an intensely hybridizing high molecular weight band not present in the parental LM1 cell line. This band represents a head-to-tail concatemer containing approximately four copies of the input pRH140ΔNae sequence and one copy of the wild-type pRH140 sequence. We will show that this wild-type sequence was generated by the correction of one of the incoming pRH140ΔNae plasmids by the pRH4-14/TK sequence residing in the chromosome of LM1.

In Figure 6 we also illustrate the Dde I Southern transfer pattern of LM1 and LM1-3. In LM1 we observe the Dde I fragments indicating the presence of the 4-14 amber mutant. In LM1-3 we observe the Dde I fragments indicating the presence of the 4-14 amber mutant, the ΔNae deletion mutant, and the corrected wild-type Neo^r gene. Thus, the single copy of the pRH4-14/TK sequence in LM1 was not corrected since it is still present in LM1-3, instead one of the incoming pRH140ΔNae sequences was corrected. This result was confirmed by plasmid rescue. In Table 3 we show that we rescued the pRH4-14/TK sequence residing in LM1-3. It contains the appropriate 5′ chromosomal junction DNA and is indistinguishable from the corresponding sequence isolated from LM1. From LM1-3 we rescued two additional classes of plasmid sequences, one corresponding to the unit length pRH140ΔNae and the other corresponding to the corrected wild-type, pRH140, sequence. We did not rescue a plasmid sequence containing a 4-14-ΔNae double mutation.

The results from the Southern transfer and plasmid rescue analysis of LM1-3 can only be interpreted in terms of a nonreciprocal or gene conversion event. A single reciprocal recombination event would have resulted in the integration of one of the incoming plasmids into the pRH4-14/TK chromosomal sequence forming a double mutant followed by a wild-type pRH140 sequence (i.e., – pRH4-14 – HSVTK – → – pRH140 (4-14-ΔNae) – pRH140 – HSV-TK-). Since the incoming plasmid does not contain a Bgl II site the J1 band in LM1 would have been shifted by 5.5 kb in LM1-3. A double reciprocal recombination event would also have converted the pRH4-14/TK sequence residing in the chromosome to a double mutant. Since none of these predictions are met we can conclude that the correction of the incoming pRH140ΔNae plasmid to wild-type occurred by gene conversion.

G418^r Cell Lines Derived from LM4

Following injection of LM4 cells with pRH140ΔNae we obtained G418^r cell lines at a frequency of 0.5–1 per 1000

Table 3. Plasmids Rescued from LM1 and LM1-3

Cell Line	Locus	Kn^r	Dde I Polymorphism	Fnu 4HI Polymorphism	Number Rescued
LM1	J1	–	4-14	4-14	9
	J1	–	4-14	4-14	3
LM1-3	unit length	–	ΔNae	wt	15
	unit length	+	wt	wt	6

Plasmid DNAs were rescued and analyzed as described in Table 2.
Locus: unit length refers to plasmids indistinguishable in size from pRH140 or pRH140ΔNae and therefore derived from a concatemer. J1 refers to a plasmid containing the junction between the integrated copy of pRH4-14 and chromosomal DNA.
Kn^r: + indicates resistance to kanamycin; – indicates sensitivity.
Dde I polymorphism: wt indicates the polymorphism found in pRH140; 4-14 indicates the polymorphism found in pRH4-14 and ΔNae indicates the polymorphism found in pRH140ΔNae.
Fnu 4HI polymorphism: wt indicates the polymorphism found in pRH140 and pRH140ΔNae; 4-14 indicates the polymorphism found in pRH4-14.

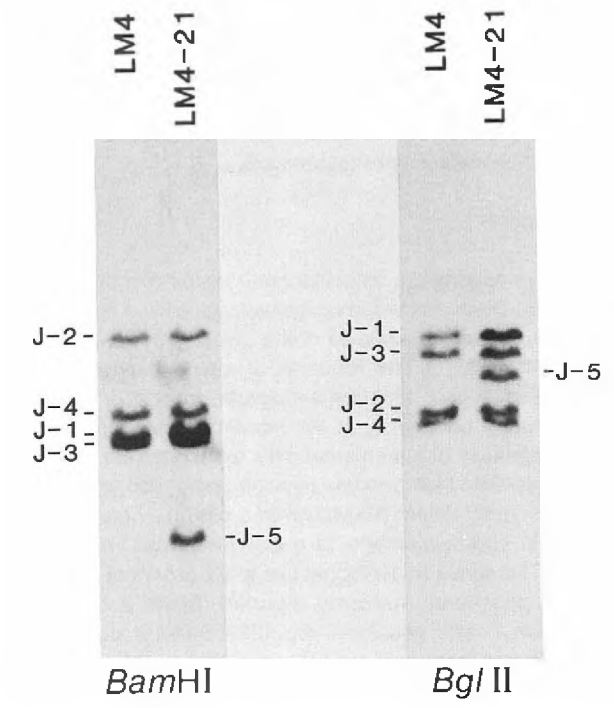

Figure 7. Bam HI and Bgl II Southern Transfer Patterns of LM4 and LM4-21

DNAs from LM4 and LM4-21 were digested with Bam HI or Bgl II and electrophoresed through 0.75% agarose. Following transfer to nitrocellulose paper, the DNAs were probed with a ^{32}P-labeled, nick-translated 1465 bp Nru I–Sma I fragment of the Neo^r gene.

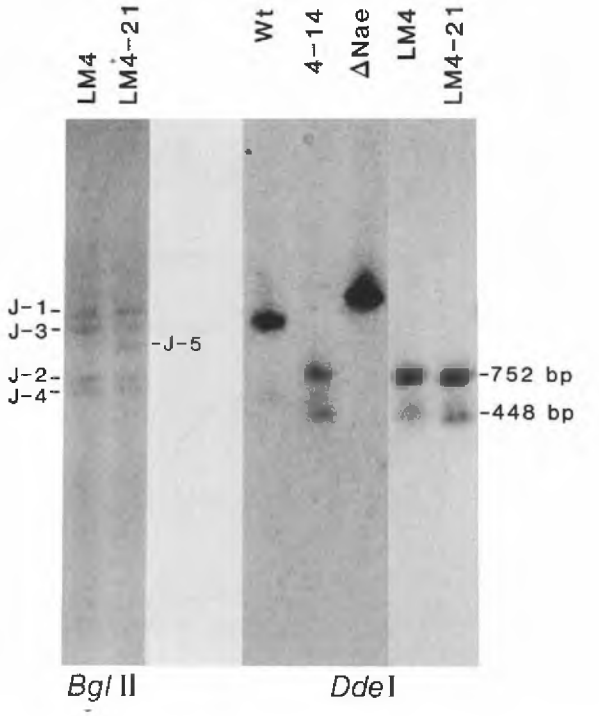

Figure 8. Southern Transfer Analysis of DNAs from LM4 and LM4-21 Digested with Bgl II or Dde I

DNAs from LM4 and LM4-21 were digested with Bgl II or Dde I, electrophoresed through 0.75% agarose, and transferred to nitrocellulose paper. The Bgl II Southern transfer was probed with a ^{32}P-labeled, 284 bp Nae I fragment from the Neo^r gene. This fragment corresponds to the fragment deleted from pRH140ΔNae. In control experiments it showed no cross-hybridization either to pRH140ΔNae or to DNA from the cell line LM5 (data not shown). The Dde I Southern transfer was probed with a ^{32}P-labeled, 1465 bp Nru I–Sma I fragment from the Neo^r gene. The markers are as described in Figure 5.

cells receiving an injection of plasmid DNA linearized with either Hind III or Nae I. No G418^r cell lines were obtained after injection of LM4 with pRH4-14 DNA linearized with Hind III or Bcl I or injection of LM4 with pRH4-14/TK linearized at Bgl II (20,000 injections).

In LM4 we again observe two classes of events, those in which the newly introduced DNA participates in homologous recombination with the chromosomal sequence and those in which initiation of homologous recombination between the incoming plasmid DNA and the chromosomal sequence leads to recombinational mutagenesis of the chromosomal sequence. Examples of the second class of events will be described elsewhere (submitted). In Figure 7 we illustrate the Bam HI and Bgl II Southern transfer patterns of LM4 and LM4-21. LM4-21 was obtained by injecting LM4 with pRH140ΔNae linearized with Nae I. These linear plasmid molecules therefore contain a gap in the Neo gene. As seen in Figure 7, LM4-21 contains a new band, designated by J-5, which is not present in the parental LM4 cell line. We will show that this band represents a single copy of pRH140ΔNae in which the gap was repaired by information provided by one of the pRH4-14/TK sequences residing in the chromosome. Following repair, the plasmid then integrated elsewhere into the host genome.

In Figure 8 we again show the Bgl II Southern transfer pattern of LM4 and LM4-21. However, this time the digested DNA was hybridized to a ^{32}P-labeled nick-translated probe prepared from the 284 bp Nae I fragment which is deleted in pRH140ΔNae. Note that the J-5 band hybridizes with this probe, showing that the incoming pRH140ΔNae plasmid has acquired this sequence.

In Figure 8 we also illustrate the Dde I Southern transfer pattern of LM4 and LM4-21. In both LM4 and LM4-21 we observe only the Dde I fragments produced by the 4-14 amber mutation. The Dde I fragment generated by a ΔNae mutation is not present. This is again consistent with the incoming pRH140ΔNae plasmid being converted by the pRH414/TK chromosomal sequence. However, rather than being converted to wild type it was completely converted to 4-14. In Figure 9 we show one pathway for the conversion of the incoming pRH140ΔNae to pRH4-14. This pathway is modeled after the double strand gap repair mechanism in yeast described by Orr-Weaver and Szostak (1983). After pRH140ΔNae was introduced into LM4, the gap in this plasmid was extended beyond the 5′-terminal end of the Neo gene. The gap was then repaired by the homologous sequence in the chromosome without a crossover, leading to the generation of a free, circular pRH4-14 plasmid. The plasmid then integrated into the host genome. From the Southern transfer patterns we know that this Neo gene is intact, but the J-5 Bam HI fragment, 4.5

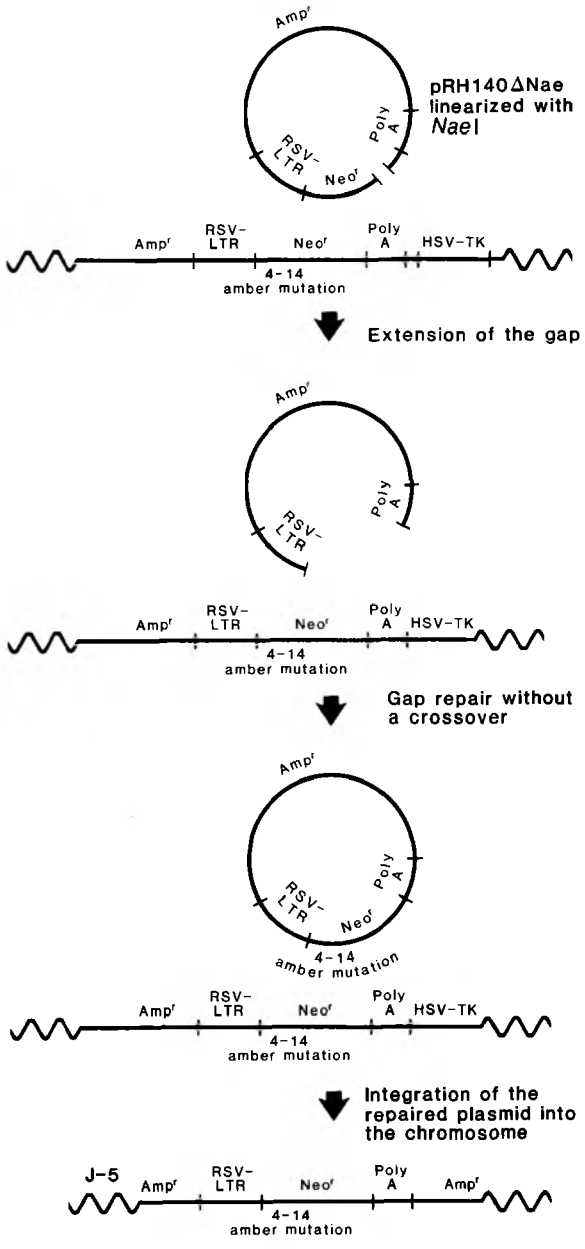

Figure 9. A Model for the Conversion of the Incoming pRH140ΔNae Plasmid to pRH4-14

The plasmid pRH140ΔNae, linearized with Nae I, was injected into nuclei of LM4. Following injection, we propose that the gap was extended beyond the region corresponding to the amber mutation (i.e., the TGG codon that encodes the 15th amino acid). The gap was then repaired by recombination with one of the pRH4-14/TK chromosomal sequences and the recombinant products resolved to yield a free pRH4-14 plasmid. The plasmid then integrated elsewhere in the genome through sequences in the Amp^r gene.

kb in length (see Figure 7), is not large enough to encompass an intact ampicillin gene. In the process of integration, it appears that the Amp^r gene was severed.

Two events occurred in LM4-21. During the first, which we have just described, the incoming pRH140ΔNae plasmid sequence was converted to pRH4-14 using information provided by one of the chromosomal pRH4-14/TK sequences. During the second event, which was responsible for the G418^r phenotype, one of the chromosomal pRH4-14/TK sequences was converted to a pseudo wild-type Neo^r gene by recombinational mutagenesis with the incoming pRH140ΔNae sequence. We will describe the latter events elsewhere (submitted).

Discussion

We have corrected a defective gene residing in the chromosome of a mammalian tissue culture cell line by injecting into the nucleus copies of the same gene carrying a different mutation. The frequency of such targeting events was much higher than we anticipated, particularly considering the complexity of the mammalian genome and the propensity of mammalian cells to incorporate exogenous DNA into their genome by nonhomologous or illegitimate recombination. We observed a productive targeting event in approximately 1 of every 1000 cells receiving DNA. The above frequency is five to six orders of magnitude higher than previously reported (Smith and Berg, 1984; Lin, Sperle, and Steinberg, 1985; Shaul et al., 1985). Considering that one in ten cells receiving a nuclear injection with a wild-type Neo^r gene becomes G418^r (Folger, Thomas, and Capecchi, 1985), the ratio of legitimate to illegitimate recombination in our experiments is approximately 1:100. This ratio of legitimate to illegitimate recombination is two to three orders of magnitude higher than previously reported (Smith and Berg, 1984; Lin, Sperle, and Sternberg, 1985).

The differences in the gene-targeting frequencies and in the ratios of legitimate to illegitimate recombination may reflect the different procedures, microinjection versus calcium phosphate coprecipitation, used to introduce the correcting plasmid DNA into the recipient cell line. Two factors, which differ between these methodologies, may play a role in determining the frequency of gene-targeting and the ratio of legitimate to illegitimate recombination: the amount of DNA delivered to the cell, and the quality of the DNA substrate delivered to the nucleus. By calcium phosphate coprecipitation large amounts of DNA are introduced into the cell. For example, Perucho, Hanahan, and Wigler (1980) have estimated that stable transformants derived by calcium phosphate coprecipitation contain, on average, 1000 kb of exogenous DNA. Prior to integration, this exogenous DNA is packaged into a large concatemer. Sequences within the large concatemer may actually be hindered from participating in homologous recombination with chromosomal sequences. In our experiments we delivered only a few copies, 5–10, of the secondary plasmid DNA to the recipient cell. Second, by calcium phosphate coprecipitation as well as by DEAE-dextran, protoplast fusion, or electroporation, the exogenous DNA must traverse the cytoplasm prior to reaching the nucleus. During this transit much of the DNA appears to travel through lysosomes where it acquires mutations, including point mutations, deletions, and rearrangements (Calos, Lebkowski, and Botchan, 1983; Razzaque, Mizusawa, and Seidman, 1983; Wake et al., 1984). The DNA, after traversing the cytoplasm, may be more heterogeneous than the

DNA added to the cells. Thus, the introduction into the nucleus of large amounts of DNA that is packaged into a large concatemer and/or the introduction of multiple DNA substrates could alter the relative rates of legitimate to illegitimate recombination and thereby influence the gene-targeting frequency.

The predominant correction events that we observed were mediated by either double reciprocal homologous recombination or by gene conversion, and in cases where we could distinguish between these alternative mechanisms, it was by gene conversion. By either of these mechanisms a sequence in the host genome is directly replaced by a sequence donated from the introduced exogenous DNA. The predominance of gene conversion or double reciprocal recombination events is a bonus if the intent is to use this protocol to do gene replacement experiments. It allows the direct transfer of defined mutations into endogenous genes without introducing either additional copies of the gene or exogenous vector sequences into the chromosome. Concern about introducing exogenous vector sequences into the mammalian genome has been heightened by the observation that mammalian genes are often inappropriately expressed when they are linked to plasmid sequences and introduced into transgenic mice (Chada et al., 1985; Magram et al., 1985; Townes et al., 1985). Removal of the plasmid sequences often leads to normal expression of the gene even though it is integrated at a random site within the host genome.

Our observation of gene conversion or double reciprocal homologous recombination events at the expense of single reciprocal homologous recombination events may reflect an inherent preference of the mammalian recombination machinery to mediate gene conversion events. It may have also been influenced by the nature of the DNA substrate injected into the recipient cells. In four of the five experiments that we described, we injected linear molecules where ends were near but outside the two nonoverlapping mutations in the Neo^r gene. Introduction of similar DNA substrates into yeast also leads predominantly to direct gene replacement (Rothstein, 1983).

It is interesting that in the cell line LM5, where the ΔNae deletion mutation resides in the chromosome, homologous recombination with the incoming pRH4-14/TK DNA led to the correction of the deletion mutation in the chromosomal gene, whereas in the cell lines LM1 and LM4, where the 4-14 amber mutation resides in the chromosome, homologous recombination with the incoming pRH140ΔNae DNA led to correction of the incoming plasmid DNA. This may have been fortuitous (i.e., five out of five cases) or it may reflect a preference for correction of the deletion mutation over the point mutation. The observation is of obvious practical importance if this technology is going to be extended to mutating endogenous genes within the mammalian genome. It will allow us to direct whether the chromosomal or incoming sequence acts as the recipient of information. As an example, if one wishes to mutate an endogenous gene, it would be wiser to introduce insertion mutations rather than deletion mutations.

Increasing the number of copies of the incoming plasmid injected into the recipient cell, increasing the number of plasmid sequences residing in the recipient cell, or providing these multiple copies in a tandem head-to-tail array did not increase the gene-targeting frequency. This was an unexpected result and indicates that the initial step in the targeting reaction, that of finding the homologous sequence in the genome, is not the rate-limiting step. Were the initial step rate limiting, then targeting would be dependent on both the concentration of the input secondary plasmid and the number of plasmid molecules residing in the host genome. The absence of such a concentration dependence argues that a subsequent step in the homologous recombination reaction, such as formation of a heteroduplex, formation of a Holiday junction, or resolution of a Holiday junction, is rate limiting.

We have compared the frequency of gene-targeting following injection of the recipient cell lines with linear and supercoiled DNA. To date, we have not observed productive targeting events following injection of supercoiled DNA. This is consistent with the observation that linear DNA molecules are much more recombinogenic in mammalian cells than their corresponding supercoiled molecules (Folger et al., 1982; Kucherlapati et al., 1984; Lin, Sperle, and Sternberg, 1984; Folger, Thomas, and Capecchi, 1985; Wake, Vernaleone, and Wilson, 1985).

One of the surprising observations of Lin, Sperle, and Sternberg (1985) was that only one in ten TK$^-$ cell lines containing defective tk genes could be transformed to Tk$^+$ by homologous recombination with an incoming complementary defective tk gene. They interpreted this observation to indicate that possibly only certain chromosomal regions are available for homologous recombination with an incoming plasmid sequence. In contrast we found that sequences in all three cell lines that we have tested were capable of undergoing recombination with the incoming plasmid. In a separate communication, we show that formation of a heteroduplex between the incoming plasmid and the homologous chromosomal sequence can initiate mutagenesis of the chromosomal gene (submitted). In LM4, which contains four copies of the defective gene integrated at four different sites, the probability of correcting any given copy appears to be equal. We have analyzed four transformants and observed correction at three of the four loci. This result again argues that the position of integration of the gene does not strongly influence the availability of that gene for interactions with the incoming plasmid. We do not have an explanation for the different results reported by Lin, Sperle, and Sternberg and ourselves. Perhaps the local chromosomal environment of the introduced gene does affect its availability for recombination. In our experiments the genes are introduced into nuclei in the absence of "carrier" DNA.

There is no reason to suspect that we have achieved the optimum frequency for targeting genes to specific sites in the mammalian genome. Many parameters need to be examined which could influence the gene-targeting frequency, such as using different DNA substrates, injecting DNA complexed with proteins involved in homologous recombination, or experimentally modulating the recipient

cell line. Extension of gene targeting to mammalian organisms such as the mouse will probably require increasing the gene-targeting frequency by an order of magnitude. The present frequency is certainly sufficient to permit manipulation of the mammalian genome within tissue culture cell lines.

Experimental Procedures

The methods used for culturing the mammalian cells, selecting the G418^r cell lines, microinjection, Southern transfer analysis, plasmid rescue, constructing the recombinant plasmids, plasmid analysis, analyzing the Dde I and Fnu 4HI restriction fragment length polymorphisms, and DNA sequencing have been described elsewhere (Capecchi, 1980; Folger, Thomas, and Capecchi, 1984; Folger, Thomas, and Capecchi, 1985).

Plasmid DNA Preparations

For the gene-targeting experiments it was imperative that the plasmid DNAs harboring the mutant Neo^r genes were neither cross-contaminated with each other nor contaminated with plasmid DNA containing a wild-type Neo^r gene. Great care was taken to obviate any source of possible contamination. Furthermore, each plasmid DNA preparation was tested for possible contamination by injecting the DNA into LMtk$^-$ cells and selecting for G418^r cell lines. We also transfected the DNA into E. coli and tested for Kn^r. No G418^r cell lines were ever obtained following injection of LMtk$^-$ cells with pRH4-14/TK, pRH140Δ Nae/TK, or related mutant vectors. Nor have we obtained Kn^r bacteria following transfection of E. coli with the mutant plasmids. In order to obviate any possibility for cross-contamination during the microinjection process itself, glass micropipettes used to deliver the DNA into nuclei were used for only a single experiment involving a single species of DNA.

Acknowledgments

We thank Laurie Fraser for her excellent technical assistance with experiments involving mammalian tissue culture and John Weiland for his expertise and assistance with DNA sequencing.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 20, 1985; revised November 20, 1985

References

Bandyopadhyay, P. K., Watanabe, S., and Temin, H. M. (1984). Recombination of transfected DNAs in vertebrate cells in culture. Proc. Natl. Acad. Sci. USA *81*, 3476–3480.

Calos, M. P., Lebkowski, J. S., and Botchan, M. R. (1983). High mutation frequency in DNA transfected into mammalian cells. Proc. Natl. Acad. Sci. USA *80*, 3015–3019.

Capecchi, M. R. (1980). High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell 22, 479–488.

Chada, K., Magram, J., Raphael, K., Radice, G., Lacy, E., and Constantini, F. (1985). Specific expression of a foreign β-globin gene in erythroid cells of transgenic mice. Nature *314*, 377–380.

Folger, K. R., Wong, E. A., Wahl, G., and Capecchi, M. R. (1982). Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. *2*, 1372–1387.

Folger, K. R., Thomas, K. R., and Capecchi, M. R. (1984). Analysis of homologous recombination in cultured mammalian cells. Cold Spring Harbor Symp. Quant. Biol. *49*, 123–138.

Folger, K. R., Thomas, K. R., and Capecchi, M. R. (1985). Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. Mol. Cell. Biol. *5*, 59–69.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. *166*, 557–576.

Hudziak, R. M., Laski, F. A., RajBhandary, U. L., Sharp, P. A., and Capecchi, M. R. (1982). Establishment of mammalian cell lines containing multiple nonsense mutations and functional suppressor tRNA genes. Cell *31*, 137–146.

Kucherlapati, R. S., Eves, E. M., Song, K. Y., Morse, B. S., and Smithies, O. (1984). Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. Proc. Natl. Acad. Sci. USA *81*, 3153–3157.

Lin, F. L., Sperle, K., and Sternberg, N. (1984). A model for homologous recombination during the transfer of DNA into mouse cells: a role for DNA ends in the recombination process. Mol. Cell. Biol. *4*, 1020–1034.

Lin, F. L., Sperle, K., and Sternberg, N. (1985). Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. Proc. Natl. Acad. Sci. USA *82*, 1391–1395.

Luciw, P. A., Bishop, J. M., Varmus, H. E., and Capecchi, M. R. (1983). Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell *33*, 705–716.

Magram, J., Chada, K., and Constantini, F. (1985). Developmental regulation of a cloned adult β-globin gene in transgenic mice. Nature *315*, 338–340.

Miller, C. K., and Temin, H. M. (1983). High efficiency ligation and recombination of DNA fragments by vertebrate cells. Science *220*, 606–609.

Orr-Weaver, T. L., and Szostak, J. W. (1983). Yeast recombination: the association between double-strand gap-repair and crossing-over. Proc. Natl. Acad. Sci. USA *80*, 4417–4421.

Perucho, M., Hanahan, D., and Wigler, M. (1980). Genetic and physical linkage of exogenous sequences in transformed cells. Cell *22*, 309–317.

Razzaque, A., Mizusawa, H., and Seidman, M. M. (1983). Rearrangement and mutagenesis of a shuttle vector plasmid after passage in mammalian cells. Proc. Natl. Acad. Sci. USA *80*, 3010–3014.

Robert de Saint Vincent, B., and Wahl, G. (1983). Homologous recombination in mammalian cells mediates formation of a functional gene from two overlapping gene fragments. Proc. Natl. Acad. Sci. USA *80*, 2002–2006.

Rothstein, R. (1983). One-step gene disruption in yeast. Meth. Enzymol. *101*, 202–211.

Shapira, G. J., Stachelek, J. L., Letsou, A., Soodak, L., and Liskay, R. M. (1983). Novel use of synthetic oligonucleotide insertion mutants for the study of homologous recombination in mammalian cells. Proc. Natl. Acad. Sci. USA *80*, 4827–4831.

Shaul, Y., Laub, O., Walker, M. D., and Rutter, W. J. (1985). Homologous recombination between a defective virus and a chromosomal sequence in mammalian cells. Proc. Natl. Acad. Sci. USA *82*, 3781–3784.

Small, J., and Scangos, G. (1983). Recombination during gene transfer into mouse cells can restore the function of deleted genes. Science *219*, 174–176.

Smith, A. J. H., and Berg, P. (1984). Homologous recombination between defective neo genes in mouse 3T6 cells. Cold Spring Harbor Symp. Quant. Biol. *49*, 171–181.

Subramani, S., and Berg, P. (1983). Homologous and non-homologous recombination in monkey cells. Mol. Cell. Biol. *3*, 1040–1052.

Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L., and Palmiter, R. D. (1985). Erythroid-specific expression of human β-globin genes in transgenic mice. EMBO J. *4*, 1715–1723.

Vogel, T. (1980). Recombination between endogenous and exogenous Simian virus 40 genes. Virology *104*, 73–83.

Vogel, T., Gluzman, Y., and Winocour, E. (1977). Recombination between endogenous and exogenous Simian virus 40 genes. J. Virol. *24*, 541–550.

Wake, C. T., Gudewicz, T., Porter, T., White, A., and Wilson, J. A. (1984). How damaged is the biologically active subpopulation of transfected DNA? Mol. Cell. Biol. *4*, 387–398.

Wake, C. T., Vernaleone, F., and Wilson, J. H. (1985). Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. Mol. Cell. Biol. *5*, 2080–2089.