Nucleic Acids Research, 1993, Vol. 21, No. 7 1541–1548

The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells

Satish Kumar and J.Paul Simons*
Department of Molecular Genetics, AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian EH25 9PS, UK

Received January 29, 1992; Revised and Accepted March 3, 1993

ABSTRACT

We have examined the effects of placing non-homologous DNA on the ends of an insertion-type gene targeting vector. The presence of terminal heterologies was found to be compatible with insertion targeting, and the terminal heterologies were efficiently removed. Terminal heterologies reduced the frequency of gene targeting to variable extents. The degree of inhibition of targeting was dependent on the length and the position of the heterology: 2.1kb heterologous sequences were more inhibitory than shorter regions of heterology, and heterology placed on the end of the long (4.8kb) arm of homology was more inhibitory than heterology positioned on the end of the short (0.8kb) arm. When heterology was placed on both arms of the targeting vector the targeting efficiencies were similar to or higher than when heterology was present on the long arm only. These results suggest that terminal sequences are removed simultaneously from both ends of targeting vectors. The removal of terminal sequences probably occurs by exonucleolytic degradation of both strands at each end, and removal of at least one of the strands is intimately coupled with the process of homologous recombination. These findings have implications for the design of gene targeting vectors.

INTRODUCTION

DNA introduced into cells can undergo recombination with homologous chromosomal sequences. The modification of chromosomal sequences by homologous recombination is termed gene targeting and is being widely used in embryonic stem (ES) cells for the generation of mice carrying defined mutations (1–3). There are broadly two types of gene targeting: replacement targeting and insertion targeting (1–3). Replacement targeting is used much more widely than insertion targeting, and results in the substitution of chromosomal sequences by sequences from the targeting vector. Insertion targeting on the other hand results in the integration of the targeting vector with no loss of chromosomal sequences, and a consequence of this is the presence of a sequence duplication in the targeted locus. Mutations generated by replacement targeting cannot revert to wild-type; mutations generated by insertion targeting can excise the targeting vector by homologous recombination between the duplicated sequences. Targeting with insertion vectors which carry a mutation within the homologous sequences, followed by selection for excision of the vector results in either complete reversion to wild-type or to reversion with the mutation retained in the chromosome. The usual products of gene targeting experiments are gross disruptions; the two-step insertion-excision method (known as hit-and-run targeting [4], and as in-out targeting [5]) has been shown to be useful for the targeting of subtle mutations, and is likely to be of increasing importance.

There have been many studies of the mechanism of homologous recombination in mammalian cells (see ref. 6 for review). In the majority of experiments the two substrates for the recombination were introduced together; in individual studies, the experiments are usually designed to assess either extrachromosomal recombination (prior to integration) or intrachromosomal recombination (between co-integrated homologous sequences). A comparison of extrachromosomal and intrachromosomal recombination has clearly demonstrated that fundamental differences exist in the mechanisms of recombination in these two experimental systems (7). Extrachromosomal recombination proceeds predominantly by a non-conservative pathway (8–10). Intrachromosomal recombination and gene targeting occur by a conservative pathway (11,12). The currently most favoured model for the mechanism of conservative recombination is double-strand-break (gap) repair (DSBR) (13–16).

Valancius and Smithies (16) found that predictions of the double-strand-break repair model hold for gene targeting in ES cells. Adjacent homologous ends were found not to be required for insertion vector gene targeting in ES cells, and that terminal gaps were always repaired during integration. The presence of small gaps in the targeting vector did not reduce the efficiency of gene targeting, arguing that the same recombination pathway is operating whether gapped or non-gapped targeting vectors are used. Further, it was shown that internal sequence heterologies within either of the arms of the targeting vector were repaired during gene targeting. The efficiency of such repair was very high for heterologies a short distance (0.2kb) from the double strand break, approaching 100%, while heterologies 0.8kb from

* To whom correspondence should be addressed

the double strand break were repaired, but at a significantly lower frequency. The repair of heterologies was unidirectional: the vector-borne heterologies were always corrected by the chromosomal target. Together, the data suggested that the normal pathway of insertion vector gene targeting includes an exonucleolytic mechanism for the removal of terminal sequences from a targeting vector, and for the use of chromosomal target sequences as template to repair gaps introduced by this process.

Chromosomal integration of exogenous DNA in mammalian cells occurs predominantly by non-homologous recombination. In order to minimise the amount of screening required in gene targeting experiments, Mansour et al (17) devised the positive-negative selection (PNS) method to select against cells in which the gene targeting vector had been integrated at random. This method is based on the finding that random integration occurs predominantly via the ends of the exogenous DNA molecule, whereas targeting of a replacement vector results from recombination within the two arms of homology. To adopt the PNS approach, a negatively selectable gene is placed on the end of one or both arms of homology; selection is applied for the positively selectable marker included within the targeting vector and against the negatively selectable gene. Very significant enrichment for gene targeting can result from the use of this method.

In principle, it would be possible to use a PNS approach to the enrichment of gene targeting with insertion vectors. A prerequisite for the application of PNS to insertion vectors is that the presence of a long heterologous DNA tail on one or both ends of the vector should not interfere with gene targeting. In the experiments described here, we have used the properties of the HPRT gene (negatively selectable and hemizygous in male cells) to allow the facile determination of targeting frequencies and isolation of targeted clones. We find that the presence of terminal heterologies does have effects on the frequency of insertion vector targeting of the HPRT gene, and that the effects depend on the position and length of the heterologous sequences. Gene targeting with vectors carrying terminal heterologies is similarly accurate as targeting with the control vector, the terminal heterologies are removed in the majority of targeted clones. We discuss these results in particular with respect to the mechanism of removal of the terminal heterologous sequences.

MATERIALS AND METHODS

Targeting vectors

The mutation in the neomycin resistance gene (18) in the pMC1neo and pMC1neopola (19, Stratagene) cassettes was corrected by exchange of an *Xma*III−*Nco*I fragment with the equivalent fragment form pSV2neo; the corrected plamids, pMC1neo(C) and pMC1neopola(C), gave significantly higher transformation frequencies than the original plasmids. The *Bam*HI site in pMC1neopola(C) was destroyed by filling in, to yield pMC1neopola(C)ΔB.

pIV was constructed by sequential insertion into pGEM4 of a 254bp *Eco*RI−*Sau*3A PCR-derived fragment from intron 2 and exon 3 of the HPRT gene (inserted into *Eco*RI and *Bam*HI-digested vector), the 5.4kb *Eco*RI fragment (intron 1, exon 2, intron 2) of the mouse HPRT gene (20), inserted into the *Eco*RI site, and the *Sal*I−*Xho*I cassette from pMC1neopola(C)DB, inserted into the *Sal*I site. Digestion of pIV with *Bam*HI yielded vector IV.

pIV+ was constructed by linearisation of IV with *Bam*HI, filling in and insertion of an *Xho*I linker (cctcgagg). Insertion of the 2.1kb *Sal*I−*Xho*I HSVtk cassette from pSPTK into the XhoI site of pIV+ yielded pIV+2.1L and pIV+2.1S, which differed in the orientation of the insert. pIV+0.7L and pIV+0.7S were derived from pIV+2.1L and pIV+2.1S respectively by digestion with *Xho*I and *Sma*I, filling in and religation. pIV+0.33L and pIV+0.33S were derived by insertion into the XhoI site of pIV+ in either orientation of a 334bp *Sal*I−*Xho*I fragment from pMC1neoD, itself derived from pMC1neopola(C) by deletion of the *Xma*III−*Bam*HI fragment. pIV+0.19L and pIV+0.19S were obtained by insertion into the XhoI site of pIV+ in either orientation of a 194bp *Sal*I−*Xho*I fragment which contains the enhancer from pMC1neo. Digestion of these plasmids with *Xho*I gave the corresponding targeting vector (e.g. pIV+2.1S×*Xho*I = vector IV2.1S); pIV+2.1L when digested with *Sma*I, yielded targeting vector IV+ 0.7L1.4S. For experiments H and I (table 1) the targeting vectors were all derived from pIV+2.1L: Vector IV was pIV+2.1L×*Bam*HI, IV+2.1L was pIV+2.1L×*Xho*I; IV+ 0.7L was pIV+2.1L×*Xho*I + *Sma*I and IV+0.7L1.4S was pIV+2.1L×*Sma*I.

The control vector (WAPneoTK) is a replacement vector designed for targeting of the whey acidic protein (WAP) gene which contains an unique *Xho*I site, and comprises the pSP72 vector, 6kb of WAP sequences, the neo cassette from pMC1neopola(C), the HSVtk cassette from pSPTK.

Cell culture and electroporation

E14 embryonic stem cells (21) were cultured in gelatin-treated flasks or petri dishes in the absence of feeder cells in DMEM (high glucose formulation, Life technologies), supplemented with NEAA, 0.1mM β-mercaptoethanol, 5% FCS and 5% NBCS, and recombinant DIA/LIF (22). For electroporation, plasmids were digested with the appropriate restriction enzymes and purified by phenol extraction, ethanol precipitation and washing with ethanol. Growing cells (passage 24 to 40) were harvested by trypsinisation and resuspended at 10^8 ml^{-1} in PBS with DNA at a concentration of 5nM. Each electroporation consisted of 0.8 ml cells which were electroporated using a Bio-Rad Gene Pulser at 500μF, 230V (575 Vcm^{-1}) (experiments A−F) or 3μF, 800V (2000 Vcm^{-1}) (experiments G−J). After 10 minutes at room temperature, the cells were diluted to 10^6 ml^{-1} in complete medium and 5×10^6 cells were plated per 6cm diameter petri dish (in experiment C, 3.2×10^6 cells were plated in each dish). G418 selection was either applied after 2 days (0.5mg Geneticin ml^{-1} [~0.25mg/ml active G418], experiments A−F) or after 24 hours (0.3mg Geneticin ml^{-1} [~0.15mg/ml active G418], experiments G−J). Six days after electroporation, 2.0mM 6-TG selection was applied to 80% of the plates, plates were fixed and stained for colony counts after a further four days (G418-selected) or after a further eight days (G418- and 6-TG-selected). When required, colonies were picked immediately prior to fixation from independent plates, and expanded for preparation of DNA.

Southern blot analysis

DNA was isolated, digested and electrophoresed according to standard methods and blotted onto HybondN (Amersham). Probes were labelled by random priming, and hybridised at 65°C by the method of Church and Gilbert (23); post-hybridisation washes were performed with 40mM sodium phosphate pH7.2, 1% sds,

1mM EDTA at 65°C. The probes used were a 339bp *Hinc*II−*Hpa*II fragment of mouse HPRT cDNA clone pHPT5 (24) which is essentially specific for exons 2 and 3 (20), linearised pGEM4-neo which contains the neo gene from pMC1neo(C) in pGEM4, and a 1.5kb *Eco*RV HSV-tk fragment from pSPTK.

RESULTS

Insertion vector targeting of the mouse HPRT gene

An insertion type targeting vector, designated IV, was constructed for targeting of the mouse HPRT gene. This vector contains approximately 5.6kb of (non-isogenic) HPRT sequences, the neo^r gene from pMC1neoPolA(C) and the plasmid pGEM4 (see Figure 1). There is an unique *Bam*HI site 0.8kb from the 5′ end of the HPRT sequences; linearisation at this site yields an insertion vector with asymmetric homologous arms: a long arm of 4.8kb and a short arm of 0.8kb. The structure predicted following targeted integration of IV is shown in figure 1. The region of homology in the unmodified endogenous HPRT gene lies within an 8.0kb *Pst*I fragment. Following targeted integration of IV, the absence of the 8.0kb wild-type *Pst*I fragment and new 10.4kb and 6.4kb *Pst*I fragments diagnostic of targeting are predicted. The 6.4kb fragment contains the 5′ recombination junction and the 10.4kb fragment contains the 3′ recombination junction; these fragments are derived from recombination with the target locus of the long and the short arms (respectively) of the targeting vector.

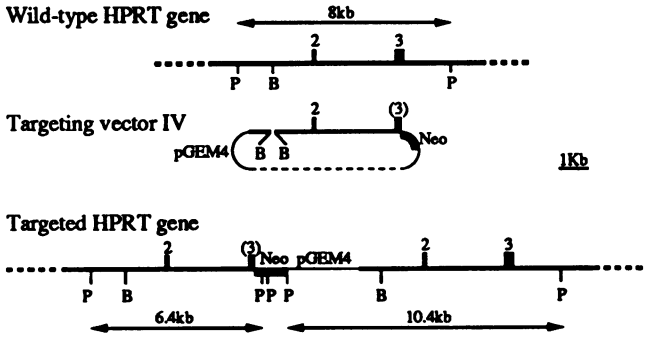

Figure 1. Targeting of the HPRT gene with insertion vector IV. Integration of IV into the HPRT gene by homologous recombination results in the insertion of the entire vector, with the loss of the wild-type 8.0Kb *Pst*I fragment, and new fragments of 6.4kb and 10.4kb. Thick lines: HPRT sequences; boxes numbered 2, 3 and (3): HPRT exons 2, 3 and incomplete exon 3. Boxes labelled Neo: Neo^r gene; thin line, pGEM4 plasmid vector sequences. P: *Pst*I sites; B: *Bam*HI sites.

Cells of the male ES cell line E14 were electroporated with IV, and with the control vector WAPneoTK, which contains no HPRT sequences and which yields G418-resistant colonies at a similar frequency to IV. G418 and 6-thioguanine (6-TG) selection were applied to select transformed cells which did not express HPRT. In each experiment, IV yielded numerous 6-TG-resistant colonies whereas WAPneoTK yielded at most one colony per experiment (Table 1), strongly suggestive of HPRT targeting by IV. To confirm that the 6-TG resistant colonies obtained following electroporation with IV were HPRT-deficient due to targeting, representative colonies were picked and expanded for Southern blot analysis. None of the thirteen clones analysed gave the 8.0kb *Pst*I fragment as found in the wild-type HPRT gene after probing with HPRT cDNA; twelve gave the 10.4 and 6.4kb *Pst*I fragments (see Figure 2), as predicted if targeted integration had occurred. One of the clones yielded, in addition to the predicted fragments, an extra ~8.7kb fragment (Figure 2, lane 5), consistent with targeted integration of multiple copies of the targeting vector in a head-to-tail tandem array (Figure 3A). The

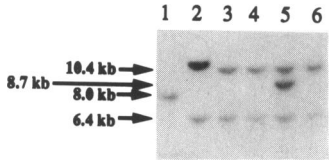

Figure 2. Southern blot analysis of clones targeted with IV. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lane 1: wild-type E14 DNA; lanes 2−6: DNA from G418- and 6-TG-resistant clones. Eight further clones gave patterns indistinguishable from those in lanes 3, 4 and 6. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the 8.7kb fragment indicates the presence of a head-to-tail repeat of the targeting vector (see Figure 3A).

Table 1. Targeting frequencies

Experiment	Relative targeting efficiency (No. of colonies) A	B	C	D	E	F	G	H	I	J	Comments
Vector											
IV	1.0 (65)	1.0 (50)	1.0 (14¶)	1.0 (43¶)	1.0 (21)	1.0 (154)	1.0 (112)	1.0 (272)	1.0 (179)	1.0 (108)	No heterology: positive control
IV+	0.98 (64)									0.88(95)	*Xho*I linker
IV+2.1L	0.28 (18)	0.36 (18)	0.29 (4¶)	0.28 (12¶)				0.33 (89)	0.36 (65)		2.1kb on long arm
IV+2.1S		0.64 (32)	0.71 (10¶)	0.72 (31¶)							2.1kb on short arm
IV+0.7L		0.40 (20)						0.59 (160)	0.57 (102)	0.55 (59)	0.7kb on long arm
IV+0.7S		0.70 (35)								0.83 (90)	0.7kb on short arm
IV+0.33L				0.57 (21)						0.30 (32)	0.33kb on long arm
IV+0.33S				1.03 (38)						0.53(57)	0.33kb on short arm
IV+0.19L					0.57 (12)		1.02 (114)				0.19kb on long arm
IV+0.19S					0.76 (16)		1.16 (130)				0.19kb on short arm
IV+0.7L1.4S				1.03 (38)	1.10 (23)	0.85 (131)		0.59 (161)	0.54 (97)		0.7kb on long arm, 1.4kb on short arm
WAPneoTK	0.02 (1)	<0.02 (0)	<0.072 (0)	<0.03 (0)	<0.05 (0)	<0.007 (0)	0.009 (1)	0.004 (1)	<0.006 (0)	0.009 (1)	Negative control

Summary of the relative targeting efficiencies with each vector. For each experiment (A−J), the targeting frequencies were normalised to the positive control vector IV (relative targeting efficiency = 1.0), and were calculated from the total number of colonies resistant to G418 and 6-TG on 12 petri dishes (except ¶: 14 petri dishes). Each petri dish received 5×10^6 electroporated cells, except in experiment C in which 3.2×10^6 cells were plated per dish. The figures for the negative control vector represent frequencies of spontaneous HPRT$^-$ mutations, and are not targeting frequencies.

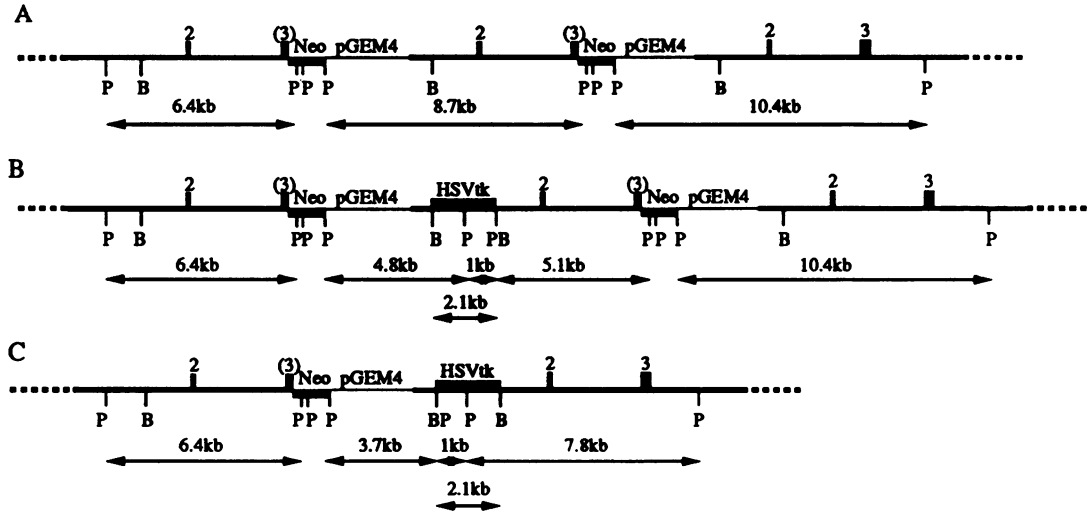

Figure 3. Structures of targeted loci in abnormally targeted clones. **A)** Structure of the targeted HPRT gene in one clone, abnormally targeted with vector IV (see Figure 2, lane 5). Tandemly repeated (Head-to-tail) copies of the vector were integrated (two copies are shown here). **B)** Structure of the targeted HPRT gene in one clone, abnormally targeted with vector IV+2.1L (see Figure 5A, lane3 and 5B, lanes 1 and 3). Tandemly repeated (Head-to-tail) copies of the vector were integrated; the terminal heterology (HSVtk) was retained internally, and removed from the end copy of the targeting vector. **C)** Structure of the targeted HPRT gene in two clones abnormally targeted with vector IV+2.1S (see Figure 5C, lanes 1 and 5; 5D, lanes 1−4). Targeting ocurred by homologous recombination within the long arm, and the HSVtk terminal heterology was retained. A single copy of the targeting vector was integrated in one clone, tandem head-to-tail repeats (not shown) of the targeting vector were integrated in the other.

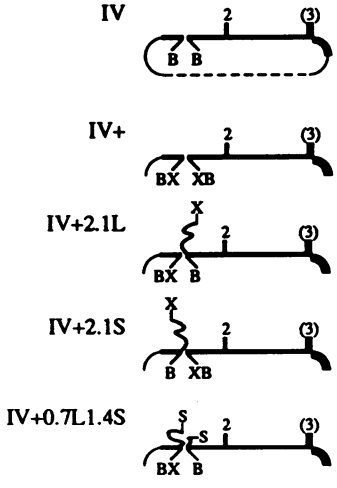

Figure 4. Targeting vectors with terminal heterology. The vectors illustrated carry: no heterology (IV); *Xho*I linker heterology on both ends (IV+); 2.1kb of heterology on the long arm and *Xho*I linker heterology on the short arm (IV+2.1L); 2.1kb of heterology on the short arm and *Xho*I linker heterology on the long arm (IV+2.1S); and 0.7kb of heterology on the long arm and 1.4kb of heterology on the long arm (IV+0.7L1.4S). The vectors which are not shown here are similar to IV+2.1L and IV+2.1S, with shorter regions of terminal heterology. Thick straight lines: HPRT sequences; boxes labelled 2 and (3): HPRT exon 2 and incomplete exon 3; thin lines: pGEM4 sequences; unlabelled boxes neo^r gene; wavy lines: terminal heterologies; B: *Bam*HI; X: *Xho*I; S: *Sma*I.

intensity of the 8.7kb fragment is higher than either the 6.4 or 10.4kb fragments, suggesting that at least three copies of the targeting vector were integrated into the HPRT gene in this clone. The remaining clone gave a fragment of 6.4kb, indicative of homologous recombination of the long arm of the vector; the 10.4kb band was absent, and a fragment of approximately 11kb was observed (Figure 2, lane 2); the intensity of this fragment indicates the integration of multiple copies of the targeting vector. Although this clone is the product of a complex integration event and the structure cannot be precisely determined, the absence of the 8.0kb fragment and the presence of the 6.4kb fragment make it likely that the HPRT gene in this clone underwent one-sided homologous recombination, being targeted by the long arm of the vector.

To confirm this interpretation of the results, the blots were reprobed with pGEM4neo, a plasmid containing all of the sequences in IV other than those derived from the HPRT gene (data not shown). As expected, in the twelve clones which gave the 10.4kb fragment when probed with HPRT cDNA, this fragment also hybridised strongly using pGEM4neo as probe. In the two clones which gave fragments which were not predicted, these fragments also hybridised to pGEM4neo, as expected if the interpretation is correct. Although only eleven out of the thirteen clones were targeted by simple insertion, the two other clones were both targeted.

These results demonstrate that the targeting vector IV integrates into the HPRT gene by homologous recombination, and show that the frequency of 6-TG-resistance is a good approximation to the targeting frequency.

The effects of terminal heterologies on insertion vector targeting efficiencies

To examine the effects of terminal heterologies on the efficiency of insertion vector targeting, a series of derivatives of IV were made. In addition to the sequences present in IV, these vectors contain inserts at the *Bam*HI site; varying from 12nt (*Xho*I linker) to 2.1kb (HSV1-tk). By digestion with *Xho*I, these vectors can be opened to leave very short heterology on both arms (vector IV+), or short heterology on one arm and long heterology on the other (see Figure 4). The designation of these vectors indicates

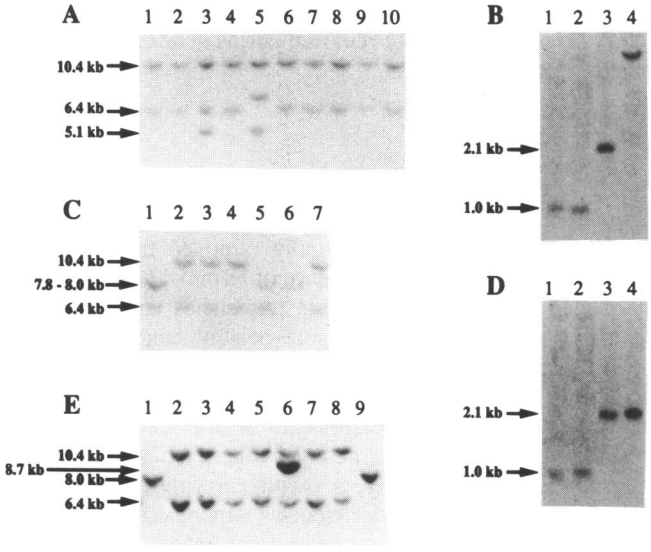

Figure 5. Southern blot analysis of clones targeted with vectors carrying terminal heterology. **A)** Clones derived from electroporation with IV+2.1L. Lanes 1–10: *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the 5.1kb fragment indicates retention of the terminal heterology (see Figure 3B). **B)** Confirmation of the presence of the terminal heterology in two clones derived from electroporation with IV+2.1L. Lanes 1 and 3: DNA from the clone shown in Figure 5A, lane 3; lanes 2 and 4: DNA from the clone shown in Figure 5A, lane 5. Lanes 1 and 2: *Pst*I digests; lanes 2 and 4: *Bam*HI digests. Probed for HSVtk sequences. **C)** Clones derived from electroporation with IV+2.1S. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lane 6: wild-type E14 cell DNA; lanes 1–5 and 7: G418- and 6-TG-resistant clones. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting; the wild-type gene yields a fragment of 8.0kb; head-to-tail integration of the targeting vector without removal of the terminal heterology gives a 7.8kb fragment (see Figure 3C). **D)** Confirmation of the presence of the terminal heterology in two clones derived from electroporation with IV+2.1S. Lanes 1 and 3: DNA from the clone shown in lane 1 of Figure 5C; lanes 2 and 4: DNA from the clone shown in Figure 5C, lane 5. Lanes 1 and 2: *Pst*I digests; lanes 2 and 4: *Bam*HI digests. Probed for HSVtk sequences. **E)** Clones derived from electroporation with IV+0.7L1.4S. *Pst*I-digested DNA from representative clones hybridised with an HPRT cDNA probe. Lanes 1 and 9: wild-type E14 DNA; lanes 2–8: DNA from G418- and 6-TG-resistant clones. Sixteen further clones gave patterns indistinguishable from those in lanes 2–5, 7 and 8. The 10.4kb and 6.4kb fragments are diagnostic of gene targeting with removal of terminal heterology from both arms; the 8.7kb fragment in lane 6 indicates the presence of a head-to-tail repeat of the targeting vector (see Figure 3A).

the length and position of the heterology (e.g. IV+2.1L carries 2.1kb of heterology on the long arm).

The frequency of targeting into HPRT with these vectors was measured and the data are summarised in table 1. In common with others (25–27) we observed significant variability between experiments of absolute targeting efficiency with any given vector. Targeting efficiencies are consistent within experiments and therefore we have normalised the results relative to a positive control (IV) included in each experiment. Targeting efficiencies are expressed as a ratio of targeting frequency obtained with each vector to that obtained with the IV control.

Two experiments were performed with vector IV+; the short (one base pair and four nucleotides) linker heterologies which were present on both arms of the vector had little or no effect on the targeting frequency (Table 1, experiments A & J). In contrast, long terminal heterology consistently reduced the efficiency of gene targeting (Table 1). This effect was most pronounced with the vector carrying 2.1kb of heterologous sequences on the long arm (IV+2.1L): in six experiments, the relative targeting efficiencies ranged between 0.28 and 0.36. Suppression of gene targeting by 2.1kb of heterology on the short arm (vector IV+2.1S) was also observed: in each of three experiments the relative targeting efficiencies ranged from 0.64 to 0.72. Similarly, in each of six experiments with shorter regions of terminal heterology (0.7kb, 0.33kb or 0.19kb), the relative targeting efficiency was higher when the heterology was on the short arm than when it was on the long arm. The data (Table 1) also suggest a possible effect of the length of the heterologous sequences on the degree of suppression of gene targeting: the shorter heterologies suppressing targeting to a lesser degree.

The effect on targeting efficiency of long heterologies on both arms

Given that long terminal heterology on either arm of the targeting vector reduces the targeting efficiency, it may be expected that this effect would be compounded by the presence of heterology on both arms. To examine this, experiments were performed using a vector carrying 0.7kb of heterologous sequences on the long arm and 1.4kb on the short arm (IV+0.7L1.4S). The relative targeting efficiencies in five experiments ranged from 0.54 to 1.10 (Table 1). This contrasts with the relative targeting efficiencies with 0.7kb of heterology on the long arm, which ranged between 0.40 and 0.59 (four experiments). In four of the experiments, other vectors were assessed in parallel. Without exception, vectors which carried 0.7kb or less heterology on the long arm gave targeting efficiencies which were similar to or lower than IV+0.7L1.4S (Table 1). Surprisingly, these data show that when there is 0.7kb of terminal heterology on the long arm, targeting is insensitive to, or is enhanced by the presence of 1.4kb of heterology on the short arm.

Terminal heterologies and the fidelity of gene targeting

Southern blot analysis was performed to determine the structures of the HPRT gene in representative clones targeted with the vectors IV+2.1L, IV+2.1S and IV+0.7L1.4S.

Of the ten clones analysed from targeting with IV+2.1L none gave the wild-type 8.0kb *Pst*I fragment, and nine gave 10.4 and 6.4kb *Pst*I fragments when probed with HPRT cDNA (Figure 5A), as predicted if targeting was accompanied by removal of the terminal heterologous DNA. One of the nine clones gave an additional band of approximately 5.1kb (Figure 5A, lane 3); the presence of three hybridising fragments indicates the integration of at least two copies of the targeting vector. A *Pst*I fragment of 5.1kb is expected if the heterologous HSV-tk sequences were retained. *Pst*I- and *Bam*HI-digested DNAs were probed for 3′ HSV-tk sequences; respectively, 1kb and 2.1kb fragments were detected (Figure 5B, lanes 1 and 3), confirming the incorporation of the heterology. A 2.1kb *Bam*HI fragment is predicted on recircularisation or head-to-tail ligation of the targeting vector. A further prediction of recircularisation or head-to-tail ligation is a 4.8kb *Pst*I fragment containing pGEM4 sequences; a band of this size was obtained following probing with pGEM4-neo (data not shown). Together, these results suggest that recombination has occurred either by insertion of a head-to-tail vector multimer with removal of the heterology from the end copy, or by replacement recombination of a head-to-tail multimer. The probable structure of the HPRT locus in this clone is shown in Figure 3B.

The tenth clone, when probed with HPRT cDNA (Figure 5A, lane 5) gave the predicted 10.4kb *Pst*I fragment, indicative of homologous recombination of the short arm of the targeting vector, and two other fragments of 7.4 and 5.1kb. These data again indicate the involvement of more than one copy of the targeting vector and incorporation of HSV-tk sequences. Probing of *Pst*I- and *Bam*HI-digested DNAs confirmed the presence of HSV-tk sequences. The internal 1kb *Pst*I fragment was detected (Figure 5B, lane 2); with *Bam*HI, a large fragment was detected (Figure 5B, lane 4), the origin of which cannot be accounted for by a simple vector ligation event. The short-arm origin of the 10.4kb *Pst*I fragment was confirmed by hybridisation with pGEM4-neo (data not shown). This clone is thus the product of an imprecise targeting event, and the precise structure of the modified HPRT locus is not clear.

Six clones from targeting with IV+2.1S were analysed (Figure 5C). Four of the clones gave both the predicted 6.4kb and 10.4kb *Pst*I fragments when probed with the HPRT cDNA probe, demonstrating that these clones were targeted and that the terminal heterology was removed. The remaining two clones gave the 6.4kb fragment indicative of targeting via the long arm and a fragment of approximately 8.0kb (Figure 5C, lanes 1 and 5). The presence of the wild-type 8.0kb *Pst*I fragment is incompatible with gene targeting which is indicated by the presence of the 6.4kb fragment. This inconsistency can be reconciled if recircularisation or head-to-tail ligation of this targeting vector occurred. This would be predicted to give a fragment of 7.8kb; targeting via the long arm of such an intermediate would yield *Pst*I fragments of 6.4kb and ~8kb, with integration of the HSVtk terminal heterology (Figure 3C). Probing for HSV-tk sequences of *Pst*I- and *Bam*HI-digested DNAs gave fragments of 1kb and 2.1kb (Figure 5D), as predicted by this explanation. This explanation is further supported by the presence of a 3.7kb *Pst*I fragment which hybridises to pGEM4-neo (data not shown); a fragment of this size is predicted by the above explanation, but because this size of fragment would also be obtained following random integration of IV+2.1S, this support is weak. The intensity of the 7.8kb *Pst*I band in one clone is higher (Figure 5C, lane 1), indicating that this band is derived from multiple copies of the vector. Of the six IV+2.1S clones analysed, four clones were targeted with removal of the terminal heterology. The other two clones were also targeted, but with retention of the terminal heterology. In these two clones recombination occurred within the long arm of the vector, and in one it appears that multiple copies of the targeting vector were integrated in a head-to-tail tandem array.

Twenty three G418- and 6-TG-resistant clones obtained with vector IV+0.7L1.4S were analysed. None gave the 8.0kb wild-type *Pst*I fragment when probed with the HPRT cDNA probe. Remarkably, all of these clones gave the 6.4 and 10.4kb *Pst*I fragments (see Figure 5E); one clone gave a single intense additional band of approximately 8.7kb (Figure 5E, lane 6). When DNA from this clone was analysed further, no hybridisation was detected in *Pst*I- and *Bam*HI-digested DNA probed for HSV-tk (data not shown), showing that the terminal heterology had been removed. Taken together, these data suggest that the HPRT gene in this clone carries multiple copies of the targeting vector from which the terminal heterology has been removed (see Figure 3A). This interpretation of the data was confirmed by hybridisation with pGEM4neo (data not shown). All of the IV+0.7L1.4S clones were targeted with removal of the terminal heterology: twenty two carry a single copy of the targeting vector and one carries multiple copies.

With each of three vectors carrying terminal heterology, the terminal heterologies were removed during gene targeting in most or all clones.

DISCUSSION

We have shown that the presence of terminal heterologies is compatible with insertion vector gene targeting and that the heterologous sequences are usually removed from the vector during gene targeting. The presence of terminal heterologies was associated with reduced targeting frequencies, the extent of the reduction being dependent on the position and length of the heterologies. Heterologies positioned on the long arm of homology consistently reduced the targeting frequency to a greater extent than did heterologies on the short arm; longer heterologies reduced the targeting frequency more than shorter heterologies. Unexpectedly, a vector carrying heterologies on both arms gave targeting frequencies which were similar to or greater than vectors bearing heterology on only one arm.

Hasty et al (25) targeted the HPRT gene in mouse ES cells with insertion vectors possessing a 5.6kb long arm of homology and a 1.2kb short arm; one of the vectors carried 2.3kb of heterologous sequences on the long arm of homology. The heterology reduced targeting efficiency in each of four experiments, giving relative targeting efficiencies from 0.13 to 0.46. The terminal heterology was removed in most clones. The data of Hasty et al (25) are consistent with the results we present here.

Gene targeting by a single targeting vector can occur by different pathways, and this can result in different structures of the targeted gene. For example, one targeting vector designed for replacement type gene targeting was found to undergo the predicted simple replacement targeting in only a minority of clones (28). The structures of most of the targeted genes suggested either replacement targeting by a head-to-tail multimer of the original vector, or insertion type recombination of a recircularised vector. With an insertion vector such as IV, it is not possible to distinguish between simple insertion targeting, recircularisation and insertion targeting or replacement recombination between the target gene and adjacent copies of IV in a multimer, because all three pathways would yield the same structure. Targeting by vectors carrying terminal heterology by either of the latter two possible pathways would however be detectable by the absence of one of the junction fragments and the retention of the terminal heterologous sequences. In our experiments, the majority of clones had undergone a simple insertion whether the vector carried heterology on the long arm, the short arm or both arms, showing that simple insertion targeting with removal of terminal heterologies is more efficient than targeting by either of the above alternative pathways.

Although we have shown that long terminal heterologies reduced the frequencies of gene targeting, the degree of suppression was not very great: at most 3.6-fold. This, together with the efficient removal of the terminal heterologies argues that targeting proceeds by essentially the same pathway of recombination whether or not the vector ends carry terminal heterologies. A prediction of this argument is that the pathway of insertion vector targeting must include the means for removal of terminal heterologies.

How are the terminal heterologies removed? A trivial explanation would be that random physical or endonuclease-induced double strand DNA breaks separate the terminal heterologous sequences from the homology. This is unlikely for

the following reasons. Firstly, frequent random double strand breaks would lead to a low frequency of incorporation of full length molecules; randomly integrated DNA is usually integrated without suffering deletions (29–33). Secondly, targeting efficiencies are significantly increased following the introduction of double strand breaks into the targeting vector (16,25); if random double strand breaks were common, pretreatment of targeting vectors would have little or no effect on targeting frequencies. Because both strands of the terminal heterologies must be removed, there are three alternative explanations for the removal of heterologies: exonuclease digestion of both strands, non-random double-strand endonuclease cleavage, and single-strand exonuclease digestion with non-random single-strand endonuclease cleavage of the other strand.

Are these alternative mechanisms compatible with the effects we observed on targeting frequency? Non-random endonuclease digestion could act before the homology search at preferred sites within the targeting vector, or alternatively endonuclease could act after the homology search, recognising the presence of terminal heterology. For the first explanation to be valid, there would have to be at least one preferred recognition site for the endonuclease within each arm of the targeting vector, and cleavage of both sites would have to be efficient; this is unlikely. Further, a prediction of this mechanism is that with the vectors used, the heterology would be removed efficiently in non-targeted clones. We have found that the majority of randomly integrated copies retain the terminal heterologies (data not shown), arguing strongly against this mechanism. This argument applies for both double-strand endonuclease, and single-strand exonuclease with single-strand endonuclease, and effectively rules out the possibility of non-random endonuclease cleavage prior to the homology search. There is no simple way to reconcile double strand endonuclease recognition and removal of terminal heterologies after the homology search with the observed differential effects of heterologies on targeting frequencies.

Removal of both strands of terminal heterologies by exonuclease(s) is compatible with the effects we observed of heterologies on targeting efficiency if exonuclease digestion occurs concurrently from both ends of targeting vectors. Removal of terminal heterology from one arm of the vector would be accompanied by removal of sequences from the other arm. If for example we take the case of vector IV+2.1L, which carries 2.1kb of terminal heterology on the long arm, then if exonuclease digestion from the two ends is precisely synchronous, the 0.8kb short arm of homology would be completely removed before complete removal of the 2.1kb heterologous sequences, and this would have profound effects on targeting. Complete removal of one arm of homology is not compatible with the structures of targeted loci that were predominantly obtained. Partial removal of homologous sequences would lead to a gap within the region of homology: it is known that such gaps are repaired efficiently during gene targeting (16). Concurrent but not precisely synchronous exonuclease digestion from both ends could explain the reduced targeting frequency observed with this vector. If we consider the case of vectors carrying terminal heterology on the short arm, removal of the heterology would be accompanied in general by incomplete removal of homologous sequences from the long arm, and therefore a smaller reduction in targeting frequency. Again, this accords with the results we obtained. The effect of terminal heterology on both arms of a targeting vector would be to reduce the extent of degradation of sequences from the arms of homology when compared with the same amount of total heterology positioned on a single arm. Another way of looking at this is that, given terminal heterology on one arm of a targeting vector, heterology on the other arm would be protective against removal of the homologous sequences. Again, this is entirely consistent with the effects we observed on targeting frequencies. All of the observed effects of length and position of terminal heterologies on targeting frequencies can be accounted for by removal of both strands of the heterologous sequences by exonuclease(s).

The exonuclease digestion would have to be coupled with recombination for this mechanism to be compatible with the efficient removal of terminal heterologies during gene targeting and with the usual retention of terminal sequences after random integration. This coupling could be by exonuclease digestion occurring after the homology search. Alternatively, exonuclease digestion could proceed until a homologous target is found; failure to find homology would result in complete degradation of the vector. A requirement of the homology search preceding exonuclease digestion is that free homologous ends are not required for the search. Recombinase activity in human cell extracts has been found to require free homologous ends (34) arguing against this possibility although it remains possible that other recombinase activities exist which do not require free ends. A second activity which promotes DNA strand exchange has been identified and purified (homologous pairing protein-1 or HPP-1 [35]), which possesses 3′→5′ exonuclease activity in addition to its pairing activity (36). An activity like HPP-1 could couple the removal of heterology with the homology search.

Let us now consider the third possible mechanism for removal of terminal heterologies: single-strand exonuclease digestion followed by non-random single-strand endonuclease cleavage. Random removal of one strand by single strand exonuclease activity is compatible with randomly integrated DNA retaining terminal sequences, because ligation of single-stranded vector ends to chromosomal DNA could occur followed by filling in of the single strand gap(s). The argument used above, of simultaneous exonuclease digestion from both ends of targeting vectors could apply, but this would not explain the effects on targeting efficiencies with the different vectors, because homologous sequences would remain on the second strand. If however, for efficient targeting there is a requirement for a region of double stranded DNA within both arms of homology, this mechanism would fit the data. Complete removal of one strand from the one arm of homology would preclude the formation of a Holliday junction by that arm, as predicted by the DSBR model of recombination (13); the absence of a Holliday junction however need not be incompatible with recombination. Removal of a single strand by 3′→5′ exonuclease would lead to a single stranded 5′ end. Although according to the Meselson-Radding model of recombination, the target sequence would be invaded by a 5′ end, this is incompatible with gap repair, which we invoke and which has been shown to occur (16). Thus, if one strand is removed by exonuclease and the other by endonuclease, the polarity of the exonuclease would have to be 5′→3′.

The findings of Valancius and Smithies (16) that internal heterologies are frequently lost during gene targeting and that the heterologies are always corrected to the chromosomal sequence suggested that the terminal sequences are subjected to exonucleolytic degradation during gene targeting. The loss of terminal homologous (and internal heterologous) sequences is likely to occur by the same mechanism as the removal of terminal heterologies. The unidirectional correction of internal heterologies (16) argues against the presence of the heterologies within regions of heteroduplex DNA, because repair or replication of heterology-

containing heteroduplexes would give bidirectional correction. This argument would be invalidated if repair of heteroduplexes is highly efficient and intrinsically directional. The removal of terminal (homologous or heterologous) sequences from insertion-type targeting vectors is thus probably due to exonucleolytic digestion of both DNA strands.

The results of our experiments suggest that PNS could be of use for insertion vector gene targeting. Even if gene targeting is suppressed by terminal heterologies, a level of enrichment significantly greater than the degree of suppression would suffice to increase the proportion of targeted clones amongst those analysed. Although we have used only one target locus, the results suggest guidelines which could be of use in the design of insertion vectors intended for PNS: the heterology should be as short as possible, and if the arms are of different sizes, the heterology should be positioned on either the shorter arm or on both arms. In preliminary experiments however, we obtained only poor enrichment with vectors IV+2.1L and IV+2.1S when ganciclovir selection was applied. We do not understand the reason for the low level of enrichment, but this result is consistent with the results of others (37–39) and with our experience with replacement vectors.

Finally, the hit-and-run/in-out method (4,5) for the generation of subtle mutations may be complicated by the unwanted correction of the subtle mutation during the insertion step (16). If this proves to be a problem, positioning terminal heterology on the arm of the vector which carries the mutation could be useful to reduce the frequency of the correction.

ACKNOWLEDGEMENTS

We would like to thank John Clark for support, Martin Hooper for E14 cells and for comments on the manuscript, David Melton for HPRT clones, Austin Smith for DIA/LIF and a DIA/LIF expression vector; Ray Ansell for technical assistance, Anthea Springbett for advice and Norrie Russell for photography. This work was supported in part by AFRC grant TAP2B. SK was supported by an Indian National Scholarship and an ORS award from the CVCP.

REFERENCES

1. Bradley,A., Hasty,P., Davis,A. and Ramirez-Solis,R. (1992) *Bio/Technology*, **10**, 534–539.
2. Capecchi,M.R. (1989) *Science*, **244**, 1288–1292.
3. Koller,B.H. and Smithies,O. (1992) *Ann. Rev. Immunol.*, 10, 705–730.
4. Hasty,P., Ramírez-Solis,R., Krumlauf,R. and Bradley,A. (1991) *Nature*, **350**, 243–246.
5. Valancius,V. and Smithies,O. (1991) *Mol. Cell. Biol.*, 11, 1402–1408.
6. Bollag,R.J., Waldman,A.S. and Liskay,R.M. (1989) *Ann. Rev. Genet.*, **23**, 199–225.
7. Waldman,A.S. and Liskay,R.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5340–5342.
8. Lin, F.-L.M., Sperle,K. and Sternberg,N. (1990) *Mol. Cell. Biol.*, **10**, 103–112.
9. Seidman,M.M. (1987) *Mol. Cell. Biol.*, **7**, 3561–3565.
10. Lin, F.-L., Sperle,K. and Sternberg,N. (1984) *Mol. Cell. Biol.*, **4**, 1020–1034.
11. Bollag,R.J. and Liskay,R.M. (1988) *Genetics*, **119**, 161–169.
12. Pennington,S.L. and Wilson,J.H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9498–9502.
13. Szostak,J.W., Orr-Weaver,T.L., Rothstein,R.J. and Stahl,F.W. (1983) *Cell*, **33**, 25–35.
14. Sun,H., Treco,D., and Szostak,J.W. (1991) *Cell*, **64**, 1155–1161.
15. Zenvirth,D., Arbel,T., Sherman,A., Goldway,M., Klein,S. and Simchen,G. (1992) *EMBO J.*, **11**, 3441–3447.
16. Valancius,V. and Smithies,O. (1991) *Mol. Cell. Biol.*, 11, 4389–4397.
17. Mansour,S.L., Thomas,K.R. and Capecchi, M.R. (1988) *Nature*, 336, 348–352.
18. Yenofsky,R.L., Fine,M. and Pellow,J.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3435–3439.
19. Thomas,K.R. and Capecchi,M.R. (1987) *Cell*, **51**, 503–512.
20. Melton,D.W., Konecki,D.S., Brennand,J. and Caskey, C.T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2147–2151.
21. Handyside,A.H., O'Neill,G.T., Jones,M. and Hooper, M.L. (1989) *Roux's Arch. Dev. Biol.*, **198**, 48–55.
22. Smith,A.G., Heath, J.K., Donaldson, D.D., Wong, G.G. Moreau, J., Stahl,M. and Rogers, D. (1988) *Nature*, **336**, 688–690.
23. Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
24. Konecki,D.S., Brennand,J., Fuscoe,J.C., Caskey,C.T. and Chinault,A.C. (1982) *Nucl. Acids Res.*, **10**, 6763–6775.
25. Hasty,P., Ramírez-Solis,R. and Bradley,A. (1992) *Mol. Cell. Biol.*, **12**, 2464–2474.
26. Piedrahita,J.A., Zhang, S.H., Hagaman,J.R., Oliver,P.M. and Maeda,N. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 4471–4475.
27. Koller,B.H., Kim,H.-S., Latour, A.M., Brigman,K., Boucher,R.C.,Jr., Scambler,P., Wainwright,B. and Smithies,O. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10730–10734.
28. Hasty,P., Ramírez-Solis,R., Chang,C. and Bradley,A. (1991) *Mol. Cell. Biol.*, **11**, 4509–4517.
29. Macleod,D., Lovell-Badge,R., Jones,S. and Jackson,I. (1991) *Nucl. Acids Res.*, **19**, 17–23.
30. Folger,K.R., Wong, E.A., Wahl,G. and Capecchi,M.R. (1982) *Mol. Cell. Biol.*, 2, 1372–1387.
31. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) *Mol. Cell. Biol.*, 5, 2599–2607.
32. Thomas,K.R., Folger,K.R. and Capecchi,M.R. (1986) *Cell*, 44, 419–428.
33. Toneguzzo,F., Keating,A., Glynn,S. and McDonald,K. (1988) *Nucleic Acids Res.*, **16**, 5515–5532.
34. Hsieh,P., Meyn,M.S. and Camerini-Otero,R.D. (1986) *Cell*, **44**, 885–894.
35. Moore,S.P. and Fishel,R. (1991) *J. Biol. Chem.*, **265**, 11108–11117.
36. Fishel,R., Derbyshire,M.K., Moore,S.P. and Young, C.S.H. (1991) *Biochimie*, 73, 257–267.
37. Le Mouellic,H., Lallemand,Y. and Brûlet,P. (1992) *Cell*, **69**, 251–264.
38. Mombaerts,P., Clarke,A.R., Hooper,M.L. and Tonegawa,S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3084–3087.
39. Tybulewicz,V.L.J, Crawford,C.E., Jackson, P.K., Bronson,R.T. and Mulligan, R.C. (1991) *Cell*, **65**, 1153–1163.