© 1997 Oxford University Press
Nucleic Acids Research, 1997, Vol. 25, No. 19 **3937–3943**

Stability of *HPRT* marker gene expression at different gene-targeted loci: observing and overcoming a position effect

David W. Melton*, Ann-Marie Ketchen and Jim Selfridge

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK

Received May 27, 1997; Revised and Accepted August 10, 1997

ABSTRACT

For sophisticated gene targeting procedures requiring two sequential selective steps to operate efficiently it is essential that the marker genes used are not prone to position effects. The double replacement gene targeting procedure, to produce mice with subtle gene alterations, is based on the use of hypoxanthine phosphoribosyltransferase (*HPRT*) minigenes in HPRT-deficient embryonic stem cells. Our standard *HPRT* minigene, under the control of the mouse phosphoglycerate kinase-1 gene promoter, was stably expressed at five of six target loci examined. At the remaining locus, DNA ligase I (*Lig1*), expression of this minigene was highly unstable. A different minigene, under the control of the mouse *HPRT* promoter and embedded in its natural CpG-rich island, overcame this position effect and was stably expressed when targeted to the identical site in the *Lig1* locus. The promoter region of the stably expressed minigene remained unmethylated, while the promoter of the unstably expressed minigene rapidly became fully methylated. The difference in the stability of *HPRT* minigene expression at the same target locus can be explained in the context of the different lengths of their CpG-rich promoter regions with associated transcription factors and a resulting difference in their susceptibility to DNA methylation, rather than by differences in promoter strength.

INTRODUCTION

The ability of chromatin to affect the expression of randomly integrated genes has been extensively documented in transgenic mouse studies (for recent reviews see 1–3). Conventional knock-out gene targeting in cultured mammalian cells involves the insertion into or replacement of part of the target locus with a selectable marker gene. Stringent demands are not placed on the marker genes used, which need only be expressed for long enough at a sufficiently high level to permit screening of colonies surviving selection to identify targeted clones. Consequently, although gene targeting could provide a powerful method to analyse the mechanism by which endogenous sequences exert position effects on integrated transgenes, there has been no stimulus to do so. More sophisticated gene targeting procedures, which result in subtle gene alterations without the retention of a marker gene, require two sequential selection steps and consequently place greater demands on the stability of marker gene expression. The advantage of these more complex targeting procedures has been emphasized by the increasing number of reports that integrated marker genes can have unanticipated effects on the expression of genes adjacent to the target locus (for a review see 4).

We have developed the double replacement gene targeting procedure in the hypoxanthine phosphoribosyltransferase (HPRT)-deficient embryonic stem (ES) cell line HM-1 (5) for making subtle gene alterations (6,7). The procedure, which was originally suggested by Reid and co-workers (8), is based on *HPRT* selectable markers (9) and the ability to select both for and against *HPRT* expression. In the first step, to inactivate the target, a region of the target locus is replaced with an *HPRT* minigene, with HAT (hypoxanthine/aminopterin/thymidine; 10) selection for *HPRT* marker expression. In the second targeting step the *HPRT* minigene is itself replaced with an altered region of the target gene to reconstitute the locus, with selection for loss of the *HPRT* marker using the purine analogue 6-thioguanine (6-TG). Starting from the knock-out cells generated in the first targeting step, a series of second steps carried out in parallel can be used to generate a panel of mouse strains with different alterations in the same target gene.

We have used this method successfully to generate mice with subtle alterations at three loci: α-lactalbumin (7), prion protein (*PrP*; 6) and DNA ligase I (*Lig1*). The method is reliant for success on the 6-TG selection identifying clones where the *HPRT* marker has been physically lost as a result of the second replacement step. The efficiency of the procedure will be compromised if 6-TG^R clones also arise due to instability of *HPRT* minigene expression. Although the frequency for the second step by double replacement targeting compared favourably with that reported for other methods, such as 'hit-and-run' (12), particularly in the case of our initial work on *Lig1*, this frequency has generally been lower than anticipated. A similar two step selection strategy has also been devised around a combination of the *neo* (positive selection for first step) and Herpes simplex virus thymidine kinase (HSV *tk*; negative selection for second step) genes (13,14). Difficulties caused by the frequent loss of HSV *tk* expression, as opposed to the physical loss of the marker by the second step of the targeting process, have been discussed (13).

Here we compare the stability of expression of an *HPRT* minigene, under the control of the constitutively expressed mouse

*To whom correspondence should be addressed. Tel: +44 131 650 5393; Fax: +44 131 668 3870; Email: david.melton@ed.ac.uk

phosphoglycerate kinase (*PGK*-1) promoter, at six different targeted loci. This represents the first survey for position effects at gene-targeted loci. *HPRT* marker expression was stable at five loci, but at the sixth, *Lig1*, expression was rapidly lost and this was associated with methylation of the *PGK*-1 promoter. When the same part of the *Lig1* gene was retargeted with a different *HPRT* minigene, under the control of the mouse *HPRT* promoter itself and containing the CpG island in which the *HPRT* promoter resides, a high level of *HPRT* expression was stably maintained and the CpG island remained unmethylated.

MATERIALS AND METHODS

HPRT minigenes

Construction and expression of the mouse *HPRT* minigenes *PGK–HPRT* and pDWM110 have been described (9). In the referenced publication *PGK–HPRT* was described as *PGK*/pDWM1. The 2.7 kb minigene (see Fig. 4) is under the control of a 510 bp *Eco*RI–*Taq*I fragment from the mouse *PGK*-1 promoter (15). The promoter is fused to the truncated *HPRT* 5′-untranslated region, 63 bp downstream of the main *HPRT* transcriptional start site. The minigene contains the remainder of the 5′-untranslated region, entire coding region and 3′-untranslated region, with the coding region being interrupted by introns 7 and 8. The 2.7 kb minigene was excised from *PGK*/pDWM1 as an *Eco*RI fragment and cloned into the *Eco*RI site of pBluescript II SK+ (Stratagene), to give pBT/PGK-HPRT (RI), which was used for targeting vector construction.

The 5.2 kb minigene pDWM110 differs from *PGK-HPRT* at its 5′-end (see Fig. 4). It is under the control of 637 bp of the natural mouse *HPRT* promoter, rather than the *PGK*-1 promoter, and contains the entire 5′-untranslated region and parts of introns 1 and 2. Intron 1 contains a key control element required by the *HPRT* promoter. The presence of the truncated intron 2 fulfils a non-specific intron requirement for efficient *HPRT* minigene expression. The 5.2 kb minigene was cloned as a blunt-ended fragment into *Sal*I/*Eco*RV-cut pBluescript II SK+, to give pBT/DWM110, which was used for targeting vector construction.

Gene-targeted clones

Gene targeting was carried out in the HPRT-deficient ES cell line HM-1 (5). For the *PrP* gene the entire coding region, which is contained in a single exon (exon 3), was replaced by the *PGK–HPRT* minigene in the opposite transcriptional orientation (6). For the β-casein gene *PGK–HPRT* was inserted into the 3′-flanking region in the same transcriptional orientation (A.Kind, personal communication). For the α-lactalbumin locus a 2.7 kb fragment containing the entire gene was deleted and replaced with *PGK–HPRT* in the same transcriptional orientation (7). For the *Lig1* gene a 4.9 kb fragment containing the 3′-end (exons 23–27) of the gene was replaced by an *HPRT* minigene (either *PGK–HPRT* or DWM110) in the same transcriptional orientation (11). For the cytokeratin 10 gene an internal fragment containing exons 3–7 was deleted and replaced with *PGK–HPRT* in the opposite transcriptional orientation (16). For the cytokeratin 18 gene a 5′-fragment containing part of exon 1 and the whole of exon 2 was deleted and replaced with *PGK–HPRT* in the opposite transcriptional orientation.

Cell culture

Culture conditions for HM-1 cells have been described (17). Cells were grown in Glasgow modified Eagle's medium supplemented with 5% foetal calf serum, 5% newborn calf serum and human LIF. Targeted clones were selected in HAT medium. Cultures were removed from HAT medium and grown for 2–3 days in medium supplemented with hypoxanthine and thymidine, before being transferred to ordinary medium. After 7–10 days of non-selective growth cells were plated at $1–1.5 \times 10^6$/90 mm dish in medium containing 5 μg/ml 6-TG to determine the spontaneous frequency of 6-TG^R in individual targeted clones.

Nucleic acid hybridization

Genomic DNA was prepared from targeted clones which had been maintained under HAT selection and from cultures grown non-selectively for 7–10 days and Southern hybridization was carried out as described (18). Probes were derived from the full-length mouse *HPRT* cDNA clone pHPT5 (19). The 5′-probe was a 254 bp *Pst*I–*Xho*I fragment, extending from the 5′-end of the cDNA to a *Xho*I site at the beginning of exon 3 (see Fig. 4). The 3′-probe was a 449 bp *Hin*dIII–*Eco*RI fragment extending from a *Hin*dIII site in exon 9 to the 3′-end of the cDNA (see Fig. 4). Total RNA was prepared from targeted clones grown non-selectively for 7–10 days and northern hybridization was carried out as described (18). Filters were probed first for *HPRT* mRNA, using pHPT5, and then reprobed with a mouse α-actin cDNA (20).

*Hpa*II PCR assay

Genomic DNA (1 μg and 100 ng), prepared from targeted clones grown non-selectively for 7–10 days, was restricted overnight with 20 U *Hpa*II or *Msp*I. The digests were treated at 95°C for 15 min to inactivate the enzyme and PCR reactions were set up with 100 and 10 ng amounts of uncut and restricted (100 ng from the 1 μg digest, 10 ng from the 100 ng digest) DNA. The cycle conditions (34 cycles) were: 1 min at 94°C, 1 min at 65°C, 30 s at 72°C. The primers used to amplify a 162 bp fragment from the CpG island of minigene DWM110 were 5′-GCCGGCAGCGTTT-CTGAGCCATTGCTGAGG (located just upstream of the main *HPRT* transcription initiation site, positions 826–855 from the mouse *HPRT* promoter sequence; GenBank accession no. J00423) and 5′-CGACGCTGGGACTGCGGGTCGGCATGA-CGG (located in *HPRT* exon 1, positions 988–959). The primers used to amplify a 150 bp fragment from the CpG island of *PGK–HPRT* were 5′-CGCACGCTTCAAAAGCGCACGTCT-GCCGCG (located just downstream of the *PGK*-1 transcription initiation site, positions 864–893 from the mouse *PGK*-1 promoter sequence; GenBank accession no. M18735) and the same *HPRT* exon 1 primer as for the DWM110 reaction above. Primer locations are indicated in Figure 5.

RESULTS

To date, *PGK–HPRT* has been our minigene of choice for gene targeting in HM-1 cells. It combines the convenience for targeting vector construction of a compact size (2.7 kb) with a high level of transient expression in ES cells and a good transformation frequency in both HPRT-deficient ES cells and fibroblasts. We have used it successfully to target a number of different loci (6,7,11,16). Unlike *PGK–HPRT*, which is controlled by a 510 bp fragment from the mouse *PGK*-1 promoter, the larger (5.2 kb) DWM110 minigene is controlled by the natural mouse *HPRT*

promoter and contains part of intron 1. At the *HPRT* locus itself the promoter resides in a 1.3 kb CpG island (21), with all but the 3′ 200 bp of this island being present in the DWM110 minigene. The mouse *PGK*-1 promoter is also located within a CpG island, but only 510 bp of this is present in the *PGK–HPRT* minigene. Because this minigene lacks any intron 1 sequences, the *PGK* promoter is not located in the same CpG island context in *PGK–HPRT* as the *HPRT* promoter is in the DWM110 minigene. The larger minigene performs slightly better than *PGK–HPRT* in transient assays (2-fold) and significantly better in transformation assays (7-fold) in ES cells (9).

Stability of *HPRT* minigene expression in targeted clones

The spontaneous frequency of 6-TG^R in six different gene-targeted clones, all containing the *PGK–HPRT* minigene, is shown in Table 1. For five of the six clones tested (*PrP*, β-casein, α-lactalbumin, cytokeratin 10 and cytokeratin 18) the frequency was very low (ranging from $<2\times10^{-8}$ to 4×10^{-6}), indicating that *HPRT* minigene expression was stable at each of these targeted loci in the absence of selection. However, for the sixth target locus, *Lig1*, two independent clones targeted with the *PGK–HPRT* minigene gave a dramatically higher level of 6-TG^R ($1–2\times10^{-2}$). To address the possibility that this very high level of 6-TG^R might be due to the carry over, in the targeted *Lig1* clones, of parental HPRT-deficient HM-1 cells that were surviving in HAT medium by metabolic cooperation (7), we subcloned the two *Lig1* clones under HAT selection and then repeated the determination of 6-TG^R on three subclones of each. All subclones assayed showed the same high level of 6-TG^R as the targeted *Lig1* clones themselves (data not shown). The 6-TG^R derivatives of these *Lig1* clones do not arise as a result of the physical loss of *PGK–HPRT* minigene DNA: all 6-TG^R colonies analysed by PCR were positive for minigene DNA (data not shown). We conclude that *PGK–HPRT* minigene expression is extremely unstable when targeted into the 3′-end of the *Lig1* gene.

The identical region of *Lig1* was then targeted with the DWM110 minigene. Two independent clones analysed showed a very low level of 6-TG^R ($<0.5–4\times10^{-6}$), indicating that this minigene can overcome the position effect and is stably expressed in the same situation that abolishes expression of *PGK–HPRT*.

Level of *HPRT* minigene mRNA in targeted clones

The levels of *HPRT* mRNA in targeted clones were determined by northern analysis after the cells had been grown non-selectively for 7–10 days (see Fig. 1). The targeted clones are all derivatives of the HPRT-deficient HM-1 line, which produces no *HPRT* mRNA as the result of a deletion which has removed the 5′-end of the gene (18). The level of *HPRT* mRNA in the wild-type ES cell line E14 (22) was the standard against which the levels of minigene-encoded *HPRT* mRNA in the targeted clones were compared. As expected, the five clones targeted with the *PGK–HPRT* minigene, which had shown a low frequency of 6-TG^R, all contained high levels of *HPRT* mRNA, although there was variation in *HPRT* mRNA levels between members of this group, particularly when the actin reprobe, to correct for variations in loading, was taken into account. The two *Lig1* clones targeted with the *PGK–HPRT* minigene contained barely detectable levels of *HPRT* mRNA, while the equivalent clones targeted with the DWM110 minigene both had high *HPRT* mRNA levels. Thus, there was a good correlation between the high frequency of 6-TG^R and the low level of *HPRT* mRNA in the *Lig1*/*PGK–HPRT* clones, confirming the conclusion that expression of the *PGK–HPRT* minigene, but not the DWM110 minigene, is highly unstable at this locus.

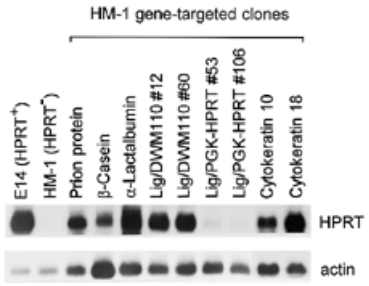

Figure 1. *HPRT* mRNA levels in gene-targeted clones. RNA (30 µg) extracted from wild-type ES cells (E14), HPRT-deficient cells (HM-1) and a range of clones targeted with *HPRT* minigenes (*PGK–HPRT* or DWM110) was subjected to northern analysis, probed for *HPRT* mRNA and then the same filter was reprobed for α-actin mRNA as a control for variations in RNA loading. Unless otherwise indicated, the *PGK–HPRT* minigene was used for gene targeting. Lig, DNA ligase I.

Table 1. Spontaneous frequency of 6-TG resistance in HM-1 clones targeted with an *HPRT* minigene

Gene targeted clone[a]	Frequency of 6-TG resistance[b]
Prion protein	0.75×10^{-6}
β-Casein	$<0.2\times10^{-6}$
α-Lactalbumin	$<2\times10^{-8}$
Lig1/DWM110 #12	$<0.5\times10^{-6}$
Lig1/DWM110 #6	4×10^{-6}
Lig1/*PGK–HPRT* #53	2×10^{-2}
Lig1/*PGK–HPRT* #106	1.5×10^{-2}
Cytokeratin 10	2×10^{-6}
Cytokeratin 18	0.4×10^{-6}

[a]Unless otherwise indicated, the *PGK–HPRT* minigene was used for gene targeting. [b]Individual gene-targeted clones were released from HAT selection and grown non-selectively for 7–10 days, before being plated on 5 µg/ml 6-TG at a density of $1–1.5\times10^6$ cells/90 mm dish, to determine the frequency of 6-TG^R colonies. Frequencies are expressed as less than a certain value where no resistant colonies were obtained from the cells plated.

Methylation status of *HPRT* minigenes in targeted clones

DNA methylation, particularly in the context of CpG-rich promoter regions (23), was clearly a prime candidate for involvement in unstable expression of the *PGK–HPRT* minigene at the *Lig1* locus. Consequently, we used Southern analysis with *Hpa*II and *Msp*I digestion to study the methylation status of *HPRT* minigenes and flanking target sequences in three selected clones. The *PrP* gene, targeted with *PGK–HPRT*, was chosen as an example where this minigene was stably expressed. The *Lig1* gene, targeted with *PGK–HPRT*, was selected as a situation where expression was highly unstable, while the same gene targeted with the DWM110 minigene provided an example where a different minigene was stably expressed at the same integration site. Southern blots for one clone of each type are shown in Figures 2 (5′-probe) and 3 (3′-probe), with the interpretation of

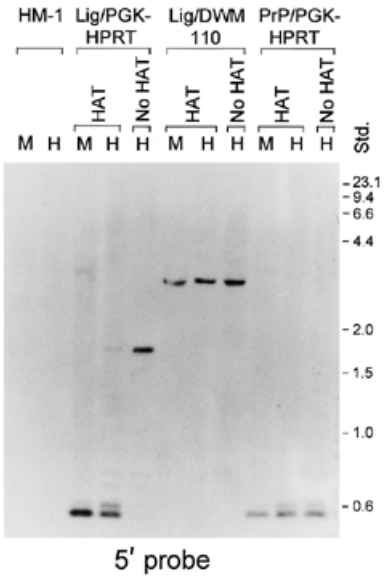

Figure 2. Methylation state around the 5′-end of *HPRT* minigenes in targeted clones. DNA (10 µg), extracted from HM-1 cells and three clones targeted with *HPRT* minigenes (*Lig1*/*PGK–HPRT*, *Lig1*/DWM110 and *PrP*/*PGK–HPRT*), was digested (M, *Hin*dIII/*Msp*I; H, *Hin*dIII/*Hpa*II) and subjected to Southern analysis. DNA was extracted from cultures maintained under HAT selection (HAT) or grown non-selectively for 7–10 days (no HAT). Std. indicates mobility of markers (kb) on the same gel. The 5′-probe hybridizes to *HPRT* exons 1 and 2.

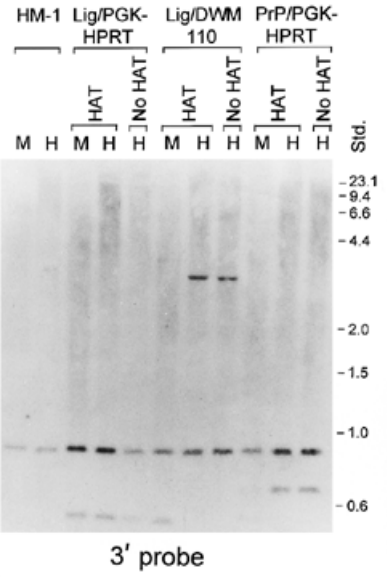

Figure 3. Methylation state around the 3′-end of *HPRT* minigenes in targeted clones. The filter shown in Figure 2 was stripped and reprobed with a 3′-probe which hybridizes to the *HPRT* 3′-untranslated region.

the results given in Figure 4. For each of the three situations identical results were obtained from two independent clones (data not shown).

The *PGK* promoter and *HPRT* 5′-untranslated region present in *PGK–HPRT* contain 10 sites, spread over 590 bp, for *Hpa*II (methylation sensitive) and *Msp*I (methylation insensitive). The

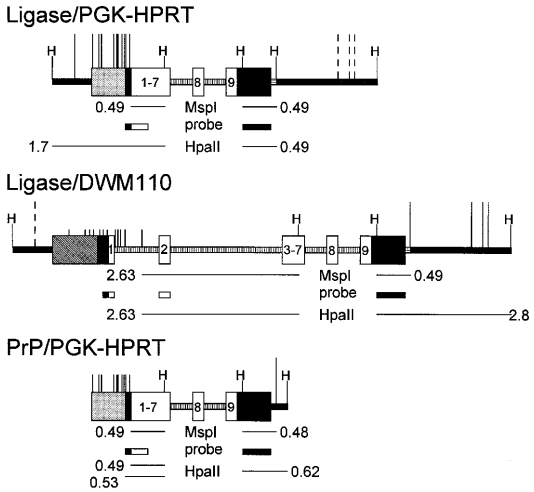

Figure 4. Methylation state of *HPRT* minigenes in targeted clones. The Southern blot data from Figures 2 and 3 are summarized here, for *Lig1*/*PGK–HPRT*, *Lig1*/DWM110 and *PrP*/*PGK–HPRT*. Open numbered boxes, *HPRT* exons; closed boxes, *HPRT* untranslated regions; light stippling, *PGK* promoter; diagonal shading, *HPRT* promoter; vertical shading, *HPRT* introns; dark stippling, flanking target sequences; horizontal shading, polylinker. H, *Hin*dIII. The locations and deduced methylation status of all *Hpa*II sites within the minigenes and flanking target sequences are indicated by the unlabelled vertical lines on the maps: short lines, non-methylated; intermediate lines, partially methylated; long lines, fully methylated. The locations of *Hpa*II sites within the *PGK*-1 and *HPRT* promoter regions were obtained from the DNA sequence [*HPRT* (25), GenBank accession no. J00423; *PGK*-1 (15), GenBank accession no. M18735]. For the flanking target sequences *Hpa*II sites were mapped using cloned material. Note that the methylation status of *Hpa*II sites in the 3′-flanking sequence of *LigI*/*PGK–HPRT* could not be probed by Southern blotting, due to the presence of a *Hin*dIII site in the polylinker at the 3′-end of the *PGK–HPRT* minigene. For this reason *Hpa*II sites in this region are shown as broken lines, assuming the same fully methylated status as in the equivalent region of *LigI*/DWM110, which could be probed. Similarly, the methylation status of the *Hpa*II site in the 5′-flanking sequence of *Lig1*/DWM110 could not be probed, due to the non-methylation of *Hpa*II sites within the DWM110 promoter region. For this reason, the *Hpa*II site in this region is shown as a broken line, assuming the same fully methylated status as in the equivalent region of *Lig1*/*PGK–HPRT*, which could be probed. For each targeted locus the regions recognized by each probe and the locations and sizes of the fragments observed following *Hin*dIII/*Msp*I and *Hin*dIII/*Hpa*II digestion are indicated. Note that the *Hin*dIII/*Hpa*II data refer only to cells grown non-selectively.

HPRT promoter/exon 1/intron 1 region of the DWM110 minigene, which comprises most of the CpG island of the *HPRT* gene, contains 12 sites for *Hpa*II, spread over 1.1 kb. There are no *Hpa*II sites in the rest of either minigene.

Lig1/PGK–HPRT. DNA was double digested with *Hin*dIII and *Msp*I or *Hpa*II and probed with a *HPRT* 5′-probe, which recognizes exons 1 and 2 and so did not give any signal with parental HM-1 DNA (see Fig. 2). DNA prepared from *Lig1*/*PGK–HPRT* cells maintained under HAT selection and restricted with *Hin*dIII/*Msp*I showed a 0.49 kb fragment extending from a *Hin*dIII site within the coding region to the first *Hpa*II site encountered in the 5′-untranslated region. When the same DNA was restricted with *Hin*dIII/*Hpa*II a ladder of higher molecular weight bands was observed in addition to the 0.49 kb fragment, indicating partial methylation of a number of *Hpa*II sites in the *PGK* promoter region. When the *Hin*dIII/*Hpa*II digest was repeated on DNA from *Lig1*/*PGK–HPRT* cells grown non-selectively for 7–10 days only a 1.7 kb fragment was detected. This fragment

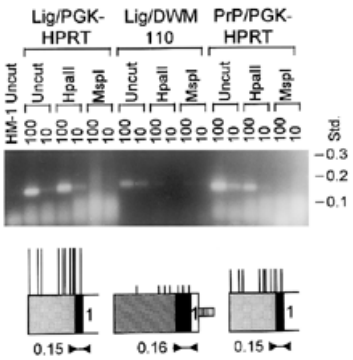

Figure 5. *Hpa*II PCR assays on the promoter regions of *HPRT* minigenes in targeted clones. DNA prepared from targeted clones (*Lig1*/*PGK–HPRT*, *Lig1*/DWM110 and *PrP*/*PGK–HPRT*) grown non-selectively for 7–10 days was subjected to PCR analysis. The DNA used (100 and 10 ng) was either uncut or restricted with *Hpa*II or *Msp*I prior to the PCR reaction. The location of the primers used (arrowheads) and sizes of the PCR products obtained for both the *PGK–HPRT* (0.15 kb) and DWM110 (0.16 kb) minigenes are shown below the picture of the gel. The same conventions as in Figure 3 are used to describe the minigenes. Note that *Hpa*II sites are indicated by the unlabelled vertical lines: short lines, non-methylated; intermediate lines, partially methylated; long lines, fully methylated. The 0.15 kb PCR product from *PGK–HPRT* spans four *Hpa*II sites, the 0.16 kb product from DWM110 spans three *Hpa*II sites. A control PCR reaction with the *PGK–HPRT* primers on 100 ng uncut HM-1 DNA is also shown. Std. indicates mobility of markers (kb) on the same gel.

extends from the *Hin*dIII site within the minigene coding region to a flanking *Hin*dIII site in the *Lig1* locus itself. This indicated that all 10 *Hpa*II sites in the *PGK–HPRT* promoter region and the single *Hpa*II site in the *Lig1* flanking region were heavily methylated and refractory to *Hpa*II digestion.

The filter was then stripped and reprobed with a 3′-probe, which recognizes both the 3′-untranslated region of the endogenous *HPRT* gene and minigene. The 0.8 kb *Hin*dIII fragment, present in all lanes in Figure 3, contains the 3′-end of the endogenous gene. Digestion of *Lig1*/*PGK–HPRT* DNA did not generate any useful information concerning the methylation state of *Hpa*II sites in the flanking target gene sequences, since a *Hin*dIII site was present in the polylinker sequence immediately 3′ of the minigene.

Thus, for the *PGK–HPRT* minigene at the *Lig1* locus we conclude that *Hpa*II sites within the *PGK* promoter region are partially methylated even when the cells are maintained under HAT selection and all sites become heavily methylated as soon as selection is removed. The methylation correlates well with the extreme instability of minigene expression at this site.

Lig1/DWM110. The same 2.63 kb fragment was detected by the 5′-probe with DNA prepared from *Lig1*/DWM110 cells maintained under HAT selection or grown non-selectively for 7–10 days and with *Hin*dIII/*Msp*I and *Hin*dIII/*Hpa*II digestion. The fragment extends from a *Hin*dIII site within the central coding region to the first *Hpa*II site encountered in intron 1. Thus, this most 3′ site within the CpG island of DWM110 remains unmethylated in the absence of selection for minigene expression. The 2.63 kb fragment is detected by the 5′-probe because it contains exon 2. The probe will also detect exon 1-containing fragments. Although we are able to detect restriction fragments containing an intact exon 1 by Southern hybridization (18), the exon 1-containing fragments generated by complete *Hpa*II digestion of DWM110 were below the limit of detection. Methylation of the *Hpa*II sites in this region would result in larger exon 1-containing fragments which would be detectable by blotting. We have been unable to detect any such exon 1-containing fragments following *Hpa*II digestion of *Lig1*/DWM110 DNA and, using this method of analysis, can find no evidence of any methylation within the CpG island of the DWM110 minigene.

Reprobing *Lig1*/DWM110 clone DNA with the 3′-probe was informative for the methylation status of *Lig1* flanking sequences because this locus, unlike *Lig1*/*PGK–HPRT*, does not contain a *Hin*dIII site in the polylinker immediately 3′ of the minigene. Digestion with *Hin*dIII/*Msp*I of DNA prepared from cells maintained under HAT selection gave a 0.49 kb fragment extending from the *Hin*dIII site in the 3′-untranslated region of the minigene to a *Hpa*II site at the junction with *Lig1* flanking sequences. *Hin*dIII/*Hpa*II digestion of DNA from cells maintained under HAT selection or grown non-selectively gave the same 2.8 kb fragment extending from the *Hin*dIII site in the minigene to a *Hin*dIII site in the flanking *Lig1* locus. This fragment would be produced if all four *Hpa*II sites in the flanking sequence were heavily methylated and refractory to digestion.

Thus, for the DWM110 minigene at the *Lig1* locus we see no evidence for any methylation of *Hpa*II sites within the CpG island region, despite the heavy methylation in flanking target sequences. The lack of detectable methylation correlates well with the stability of DWM110 minigene expression at this locus.

PrP/PGK-HPRT. Hybridization with the 5′-probe to *Hin*dIII/*Msp*I-digested DNA prepared from *PrP*/*PGK–HPRT* cells maintained under HAT selection detected the same 0.49 kb *Hin*dIII/*Hpa*II fragment seen in *Lig1*/*PGK–HPRT* cells. The same fragment was also prominent following *Hin*dIII/*Hpa*II digestion of DNA prepared from cultures maintained under HAT selection or grown non-selectively for 7–10 days. In addition, both digests showed a faint 0.53 kb fragment. Such a pattern would be compatible with there being a very low level of methylation over the *PGK* promoter. The prominent 0.49 kb fragment would arise if the first *Hpa*II site within the 5′-untranslated region was unmethylated in most cells of the culture. In a minority of cells this site would be methylated and the *Hpa*II would read through to the next site, to give the 530 bp fragment.

Reprobing *Hin*dIII/*Msp*I-cut DNA with the 3′-probe gave a 0.48 kb fragment extending from the *Hin*dIII site in the 3′-untranslated region of the minigene to a *Hpa*II site in the *PrP* flanking sequence. (Note that this fragment is only just detectable in Figure 3.) *Hin*dIII/*Hpa*II digestion revealed that this *Hpa*II site was methylated in DNA prepared both from cultures maintained under HAT selection and cultures grown non-selectively for 7–10 days, to give a 0.62 kb fragment reading through to a *Hin*dIII site in the flanking sequence.

Thus, the *PGK–HPRT* minigene at the *PrP* locus is largely unmethylated and this state is maintained in the absence of selection for *HPRT* expression. This correlates well with the stability of minigene expression at this locus.

HpaII PCR assays. Although Southern blotting gave no indication of any methylation of *Hpa*II sites in the CpG island of the minigene in *Lig1*/DWM110 cells, our inability to detect very small exon 1-containing fragments meant that we could not exclude the possibility that some methylation was present. Consequently, we carried out a series of *Hpa*II PCR assays (24) to supplement the blotting data. We were unable to establish PCR reactions to amplify the entire promoter region from either minigene, presumably due to the very high CG content of the

template DNA. The PCR reaction used for the DWM110 minigene amplified a 162 bp fragment from exon 1, spanning three *Hpa*II sites (see Fig. 5). The reaction for *PGK–HPRT* amplified an equivalent 150 bp fragment, spanning four *Hpa*II sites. Control, non-targeted HM-1 cells did not give a product with either primer pair. The quantitative nature of both PCR reactions was demonstrated using 100 and 10 ng uncut genomic DNA as template. In each case considerably more product was obtained from the reaction with 100 ng template. All reactions were carried out on the same DNA, isolated from targeted clones that had been grown non-selectively for 7–10 days, that was used for the Southern blot analysis. Data from one clone of each type (*Lig1*/*PGK–HPRT*, *Lig1*/DWM110 and *PrP*/*PGK–HPRT*) are shown in Figure 5. Equivalent results were obtained from a second independent clone of each type (data not shown).

*Msp*I digestion of *Lig1*/*PGK–HPRT* DNA prior to PCR prevented appearance of the 150 bp product. As expected, prior digestion with *Hpa*II had no effect on appearance of the product, confirming the conclusion from the Southern blots that the *Hpa*II sites in this region are heavily methylated and so refractory to restriction. Digestion of *Lig1*/DWM110 DNA with *Msp*I or *Hpa*II prior to PCR generated the same result: the almost complete disappearance of product, providing direct evidence in support of the blotting data, that the region of the promoter amplified by the PCR reaction is unmethylated.

*Msp*I digestion of *PrP*/*PGK–HPRT* DNA prior to PCR prevented appearance of the product. The product was present following *Hpa*II digestion, but at a reduced level compared to uncut DNA, confirming the conclusion from the Southern blot that there is some methylation of the CpG island in this targeted allele.

DISCUSSION

We have developed a flexible system for gene targeting in ES cells, based on the HPRT-deficient cell line HM-1 (5) and the use of *HPRT* minigenes as selectable markers (9). The system can be used for conventional gene knock-out (6,7,11,16), but is also ideally suited to more sophisticated two step gene targeting procedures, designed to introduce subtle gene alterations (6,7), and also for making conditional gene knock-outs by combining gene targeting with a site-specific recombination system, such as Cre/*loxP*.

Using the *PGK–HPRT* minigene, under the control of a 510 bp fragment from the mouse *PGK*-1 gene promoter, we have never failed to identify knock-out clones at a range of different target loci. At five of the six targeted loci examined expression of the *PGK–HPRT* minigene was stable. At the sixth locus, *Lig1*, expression was highly unstable and, although this did not prevent successful isolation of knock-out clones (11), the very high spontaneous frequency of 6-TG^R observed did preclude the use of double replacement targeting to introduce subtle alterations at this locus. The second selection step in this procedure, introduction of the subtle alteration to the target locus, involves selection in 6-TG for physical loss of the *HPRT* marker. The frequency of gene targeting is of the order of 10^{-5}–10^{-6}, so clearly a spontaneous 6-TG^R frequency of 10^{-2}, due to unstable minigene expression, would prevent easy identification of second step targeted clones at this locus.

The position effect leading to instability of *PGK–HPRT* minigene expression at the *Lig1* locus was overcome by the use of another minigene, DWM110. Two subtle gene alterations were subsequently carried out by double replacement gene targeting in *Lig1*/DWM110 cells at an equivalent frequency (1/5 to 1/10 of the 6TG^R clones screened contained the desired subtle alteration) to that which we have already reported for the α-lactalbumin (7) and *PrP* loci (6). While it is impossible to extrapolate from one position effect experienced/six loci targeted to the implications for targeting many other loci, we do make the following practical suggestions. If there is definitely only a requirement for simple gene knock-out at a particular locus, then we see no compelling reason to switch from *PGK–HPRT* to the larger minigene, DWM110, whose size will complicate the construction of targeting vectors. However, if two step gene targeting procedures, such as double replacement, are required, then potential difficulties associated with unstable minigene expression should be minimized by using DWM110 rather than *PGK–HPRT*.

DWM110 is expressed more strongly than *PGK–HPRT* in ES cells, giving 2-fold higher transient expression and a 7-fold higher transformation frequency (9). This difference in transient expression between the two minigenes is not sufficient to explain the dramatic difference in expression of the two minigenes at the *Lig1* locus. We believe that the key is the stability, rather than the absolute level, of minigene expression. DWM110 expression is stable in the absence of selection, while *PGK–HPRT* expression is highly unstable and is rapidly lost from cultures grown non-selectively, leading to a drop in HPRT levels below the threshold needed for acquisition of 6-TG^R. This is clearly an extreme example of a difference in stability of minigene expression at a particular targeted locus. The 7-fold higher transformation frequency for DWM110, compared with *PGK–HPRT*, could reflect a less extreme indication of the same difference in stability of minigene expression, but this time following random DNA integration.

In one of the locations, the *PrP* locus, where the *PGK–HPRT* minigene is stably expressed the *PGK* promoter remained largely unmethylated. The extreme instability of *PGK–HPRT* expression at the *Lig1* locus was correlated with complete methylation of all the *Hpa*II sites in the promoter region. In extreme contrast, expression of the DWM110 minigene targeted to precisely the same region of the *Lig1* locus was stable and there was no evidence for any methylation of the promoter region. For both loci all *Hpa*II sites examined in flanking target sequences were fully methylated.

The promoters of the mouse *HPRT* (25) and *PGK*-1 (15) genes, in common with many other housekeeping and some tissue-specific genes, are located in CpG-rich regions of the genome (for a review see 23). While the rest of the genome is usually fully methylated, there is a strong correlation between lack of methylation at CpG islands and gene expression. In situations where genes with a CpG island are not expressed, such as genes on the inactive X chromosome in female cells, then the island is heavily methylated. The binding of transcription factors to sites within the CpG island is believed to deny access to methylases and so keep the islands non-methylated (26,27).

Lig1 maps to mouse chromosome 7 (28). The wild-type gene is strongly expressed in ES cells, but no *Lig1* mRNA could be detected from the allele targeted with the *PGK–HPRT* minigene (11). Targeting has resulted in deletion of the 3′-end of the gene, including the polyadenylation signal. We have presumed that the failure to detect any *Lig1* mRNA from the target locus was due to a failure to polyadenylate primary transcripts, but it could also arise from a failure to transcribe the targeted locus from the *Lig1*

promoter. The position effect leading to instability of *PGK–HPRT* minigene expression at this locus could result from a particular property of this region of the genome or could be due to an alteration in the expression properties of the locus associated with the targeted event itself. While we cannot discriminate between these two possibilities, either could apply to many other target loci or targeting gene structures. Of equal interest is the mechanism by which the DWM110 minigene is able to overcome this position effect and continue to be stably expressed. The endogenous mouse *HPRT* promoter is definitely not a stronger natural promoter than its *PGK*-1 equivalent. While both genes are constitutively expressed, *HPRT* mRNA is present at low levels in all mouse cell lines and tissues that we have examined, with the exception of the brain, where expression is elevated (18). In contrast, the strength of the 510 bp *PGK*-1 promoter fragment has been equated to that of viral promoters, such as SV40 (29). Both promoters lie in CpG islands and bind a similar range of generalized transcription factors (30–32). The density of CpG dinucleotides in the promoter regions of the two minigenes is also similar: 7.1 CpGs/100 bp for DWM110, 8.5 for *PGK–HPRT*. However, in man the endogenous *HPRT* promoter is more resistant to methylation on the inactive X chromosome than its *PGK*-1 equivalent (for a review see 23). We are not dealing with the entire CpG islands for the mouse *HPRT* and *PGK*-1 genes, so it is not clear that the apparent greater susceptibility of the *PGK* island to methylation can account for our results. We suggest that the size of the CpG-rich promoter regions could be the key difference: 590 bp in *PGK–HPRT* (*PGK* promoter and *HPRT* exon 1), 1.1 kb in DWM110 (*HPRT* promoter, exon 1 and intron 1), comprising all but the 3′ 200 bp of the natural CpG island. For the *HPRT* gene we have identified key control elements (9) and Sp1 binding sites within intron 1. Thus, for the *HPRT* promoter in DWM110 the assembled transcription factors are likely to extend from the promoter region itself and into intron 1. This will not be the case for the *PGK*-1 promoter in *PGK–HPRT*. Perhaps this more extensive array of transcription factor binding sites permits the assembly of a more stable transcription complex, which can better resist the attention of methylases to the underlying DNA. Another, less likely, possibility to explain the differential methylation sensitivity of the two minigenes at the *Lig1* locus is that the additional, non-CpG island components (parts of intron 1 and 2) present in DWM110 but lacking in *PGK–HPRT* could act to protect any flanking CpG island from methylation.

Conventional transgenesis, with its random integration of transgenes, has led to the discovery of locus control regions, which render transgenes non-susceptible to position effects (for a review see 1). However, this method of analysis cannot probe the mechanism by which specific chromatin regions can affect expression of integrated genes. The observation of major differences in the stability of expression from different promoters targeted into the same chromosomal location could be useful in understanding the key interactions between transcription factors, chromatin structure and DNA methylation which determine position effects at the molecular level. In practical terms, if marker genes with CpG islands are generally more stably expressed than their non-island-containing counterparts, then this could result in valuable improvements in the efficiency of a range of gene targeting procedures.

ACKNOWLEDGEMENTS

D.W.M. is grateful to Alex Kind (PPL Therapeutics, Edinburgh) for sharing his unpublished data on β-casein gene targeting. The DNA ligase I gene targeting was supported by The Cancer Research Campaign (grant no. SP2095/0103). The prion protein gene targeting was supported by The Biotechnology and Biological Sciences Research Council (grant no. BS204512).

REFERENCES

1 Dillon,N. and Grosveld,F. (1994) *Curr. Opin. Genet. Dev.*, **4**, 260–264.
2 Clark,A.J., Bissinger,P., Bullock,D.W., Damak,S., Wallace, R., Whitelaw,C.B. and Yull,F. (1994) *Reprod. Fertil. Dev.*, **6**, 589–598.
3 Bonifer,C., Huber,M.C., Jagle, U., Faust,N. and Sippel,A.E. (1996) *J. Mol. Med.*, **74**, 663–671.
4 Olson,E.N., Arnold,H.-H., Rigby,P.W.J. and Wold,B.J. (1996) *Cell*, **85**, 1–4.
5 Magin,T.M., McWhir,J. and Melton, D.W. (1992) *Nucleic Acids Res.*, **20**, 3795–3796.
6 Moore,R.C., Redhead, N.J., Selfridge,J., Hope,J., Manson,J.C. and Melton,D.W. (1995) *BioTechnology*, **13**, 999–1004.
7 Stacey,A., Schnieke,A., McWhir,J., Cooper,J., Colman,A. and Melton,D.W. (1994) *Mol. Cell. Biol.*, **14**, 1009–1016.
8 Reid,L.H., Gregg,R.G., Smithies,O. and Koller,B.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4299–4303.
9 Magin,T.M., McEwan,C., Milne,M., Pow,A.M., Selfridge,J. and Melton,D.W. (1992) *Gene*, **122**, 289–296.
10 Littlefield,J.W. (1964) *Science*, **145**, 709.
11 Bentley,D.J., Selfridge,J., Millar,J.K., Samuel,K., Hole,N., Ansell, J.D. and Melton,D.W. (1996) *Nature Genet.*, **13**, 489–491.
12 Hasty,P., Ramirez-Solis,R., Krumlauf,R. and Bradley,A. (1991) *Nature*, **350**, 243–246.
13 Askew,G.R., Doetschman, T. and Lingrel, J.B. (1993) *Mol. Cell. Biol.*, **13**, 4115–4124.
14 Wu,H., Liu,X. and Jaenisch,R. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 2819–2823.
15 Adra, C.N., Boer,P.H. and McBurney, M.W. (1987) *Gene*, **60**, 65–74.
16 Porter,R.M., Leitgeb,S., Melton,D.W., Swensson,O., Eady,R.A.J. and Magin,T.M. (1996) *J. Cell Biol.*, **132**, 925–936.
17 Redhead,N.J., Selfridge,J., Wu,C.-L. and Melton,D.W. (1996) *Hum. Gene Ther.*, **7**, 1491–1502.
18 Thompson,S., Clarke,A.R., Pow,A.M., Hooper,M.L. and Melton,D.W. (1989) *Cell*, **56**, 313–321.
19 Konecki,D.S., Brennand,J., Fuscoe,J., Caskey,C.T. and Chinault,A.C. (1982) *Nucleic Acids Res.*, **10**, 6763–6775.
20 Minty,A. J., Caravatti, M., Robert,B., Cohen,A., Daubas,P., Weydert,A., Gros, F. and Buckingham,M. (1981) *J. Biol. Chem.*, **256**, 1008–1014.
21 Melton,D.W. (1987) In Maclean,N. (ed.), *Oxford Surveys on Eukaryotic Genes*. Oxford University Press, Oxford, UK, Vol. 4, pp. 34–76.
22 Hooper,M., Hardy,K., Handyside,A., Hunter,S., and Monk,M. (1987) *Nature*, **326**, 292–295.
23 Cross,S.H. and Bird,A.P. (1995) *Curr. Opin. Genet. Dev.*, **5**, 309–314.
24 Singer-Sam,J., LeBon,J.M., Tanguay,R.L. and Riggs,A.D. (1990) *Nucleic Acids Res.*, **18**, 687.
25 Melton,D.W., Konecki,D.S., Brennand,J. and Caskey,C.T. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2147–2151.
26 Brandeis,M., Frank,D., Keshet,I., Siegfried,Z., Mendelsohn,M., Nemes,A., Temper,V., Razin,A. and Cedar, H. (1994) *Nature*, **371**, 435–438.
27 Macleod,D., Charlton,J., Mullins,J. and Bird,A.P. (1994) *Genes Dev.*, **8**, 2282–2292.
28 Gariboldi,M., Montecucco,A., Columbano,A., Ledda-Columbano,G.M., Savini,E., Manenti,G., Pierotti,M.A. and Dragani,T.A. (1995) *Mol. Carcinogen.*, **14**, 71–74.
29 McBurney,M.W., Sutherland,L.C., Adra,C.N., Leclair, B., Rudnicki,M.A. and Jardine,K. (1991) *Nucleic Acids Res.*, **19**, 5755–5761.
30 Melton,D.W., McEwan,C., McKie,A. and Reid,A.M. (1986) *Cell*, **44**, 319–328.
31 Jiralerspong,S., and Patel,P.I. (1996) *Proc. Soc. Exp. Biol. Med.*, **212**, 116–127.
32 Pfeifer,G.P. and Riggs,A.D. (1991) *Genes Dev.*, **5**, 1102–1113.