Activation of the *lac* Repressor in the Transgenic Mouse

Heidi Scrable* and Peter J. Stambrook†

**Department of Neuroscience, University of Virginia, Charlottesville, Virginia 22908 and †Department of Cell Biology, Neurobiology, and Anatomy, University of Cincinnati, Cincinnati, Ohio 45267*

Manuscript received December 10, 1996
Accepted for publication June 13, 1997

ABSTRACT

We have introduced sequences encoding the *lac* repressor of *Escherichia coli* into the genome of the mouse. One sequence was derived from the bacterial *lac* operon and the other was created by re-encoding the amino acid sequence of *lacI* with mammalian codons. Both versions are driven by an identical promoter fragment derived from the human β-actin locus and were microinjected into genetically identical pronuclear stage embryos. All transgenes utilizing the bacterial coding sequence were transcriptionally silent in all somatic tissues tested. The sequence re-encoded with mammalian codons was transcriptionally active at all transgene loci and expressed ubiquitously. Using methylation-sensitive enzymes, we have determined the methylation status of *lac* repressor transgenes encoded by either the bacterial or mammalian sequence. The highly divergent bacterial sequence was hypermethylated at all transgene loci, while the mammalian sequence was only hypermethylated at a high copy number locus. This may reflect a normal process that protects the genome from acquiring new material that has an abnormally divergent sequence or structure.

THE introduction of foreign genes and DNA fragments into the genome of the mouse has led to the elucidation of the function of many normal genes and to an understanding of how mutations in particular genes disrupt phenotype. The ability to introduce exogenous DNA sequences that code for either normal or mutant gene products, however, has been limited to those whose expression results in benign or, at the least, sublethal phenotypic changes. In an effort to circumvent problems that arise from the unregulated expression of introduced sequences, we have been constructing a regulatable transgenic system for the mouse that is based on the lactose (*lac*) operon of *Escherichia coli* (JACOB and MONOD 1961). Like analogous systems that use temperature-sensitive mutations to study lethal mutations in bacteria and lower eukaryotes, this system would allow the introduction and analysis of embryonic lethal genes at the organismal level without compromising the viability of their host. The capacity to regulate genes in the mouse *in vivo* would greatly expand the repertory of genes that can be altered and analyzed within the context of an organism more closely related to the human.

Use of regulatory elements derived from the bacterial *lac* operon in our transgenic system required that we introduce sequences coding for the *lac* repressor (*lacI*) that would be expressed in the mouse. Data from many other experiments in which bacterial genes have been introduced into the mammalian genome have implicated methylation in the inability of these genes to be expressed. Sporadic expression of reporter genes like *lacZ* in cultured mouse embryos (NILSSON and LENDAHL 1993) and in the transgenic mouse (BEDDINGTON *et al.* 1989; reviewed in CUI *et al.* 1994) has been attributed to the gradual acquisition of methylated sequences at transgenic as well as endogenous loci as the methylation patterns characteristic of the mature genome are established and maintained.

To avoid suppression of gene activity by methylation directed at the bacterial *lacI* coding sequence, we re-encoded the amino acid sequence of the *lac* repressor with mammalian codons (ZHANG *et al.* 1991). To test the effect on expression of altering only the coding sequence, we prepared two transgene constructs that were identical except for the sequence encoding *lacI* and microinjected them into genetically identical mice. A comparison of gene activity and methylation status of the two types of repressor transgene suggests that during the process of transgenesis the host genome can respond differentially to introduced sequences. This response appears to depend on how closely the introduced sequences resemble the host genome and results in transcriptional activation of the closely related mammalian *lacI* sequence and transcriptional suppression of the widely divergent bacterial sequence.

Corresponding author: Heidi Scrable, Department of Neuroscience, University of Virginia, Charlottesville, Virginia 22908.
E-mail: hs2n@virginia.edu

MATERIALS AND METHODS

Synthesis of mammalian *lacI*: Segments of 80 nucleotides in length corresponding to either strand of the gene were synthesized on an Applied BioSystems DNA synthesizer, annealed, and ligated in groups of three overlapping fragments. These larger fragments [~240 base pairs (bp)] were ligated to each other in turn to produce the final product. The mam-

```
atg aaa cca gta aca ttg tat gat gtt gca gag tat
gcc ggt gtc tct tat cag act gtt tcc aga gtg gtg
aac cag gcc agc cat gtt tct gcc aaa acc agg gaa
aaa gtg gaa gca gcc atg gca gag ctg aat tac att
ccc aac aga gtg gca caa caa ctg gca ggc aaa cag
agc ttg ctg att gga gtt gcc acc tcc agt ctg gcc
ctg cat gca cca tct caa att gtg gca gcc att aaa
tct aga gct gat caa ctg gga gcc tct gtg gtg gtg
tca atg gta gaa aga agt gga gtt gaa gcc tgt aaa
gct gca gtg cac aat ctt ctg gca caa aga gtc agt
ggg ctg atc att aac tat cca ctg gat gac cag gat
gcc att gct gtg gaa gct gcc tgc act aat gtt cca
gca ctc ttt ctt gat gtc tct gac cag aca ccc atc
aac agt att att ttc tcc cat gaa gat ggt aca aga
ctg ggt gtg gag cat ctg gtt gca ttg gga cac cag
caa att gca ctg ctt gcg ggc cca ctc agt tct gtc
tca gca agg ctg aga ctg gcc ggc tgg cat aaa tat
ctc act agg aat caa att cag cca ata gct gaa aga
gaa ggg gac tgg agt gcc atg tct ggg ttt caa caa
acc atg caa atg ctg aat gag ggc att gtt ccc act
gca atg ctg gtt gcc aat gat cag atg gca ctg ggt
gca atg aga gcc att act gag tct ggg ctg aga gtt
ggt gca gat atc tca gta gtg ggt tat gat gat act
gaa gac agc tca tgt tat atc cca ccc tca acc acc
atc aaa caa gat ttt aga ctg ctg ggg caa acc agt
gtg gac aga ttg ctg caa ctc tct caa ggc caa gca
gtg aag ggc aat cag ctg ttg cca gtc tca ctg gtg
aag aga aaa acc acc ctg gca ccc aat aca caa act
gcc tct ccc cgg gca ttg gct gat tca ctc atg cag
cta gca aga cag gtt tcc aga ctg gaa agt ggg cag
gca gct ctg ccc aag aag aag cga aag gtg tga
 A   A   L   P   K   K   K   R   K   V
```

FIGURE 1.—DNA sequence of mammalian *lacI*. Each triplet encodes the same amino acid found in the bacterial *lac* repressor. The translation product of the nuclear localization signal from SV40 is indicated beneath the 30-bp extension appended to the 3′ end of the *lacI* coding region immediately 5′ to STOP (tga).

malian *lacI* sequence was also modified to include the nuclear localization signal (NLS) from SV40, which results in the sequestration of most repressor protein in the nucleus, as described in LIU *et al.* (1992). The NLS was appended to the carboxy terminus of the protein, with two additional alanines serving as a linker. The sequence was also designed to avoid inclusion of potential splice donor or splice acceptor sites. The final product was sequenced and several cloning artifacts corrected by PCR-based, site-directed mutagenesis (BOWMAN *et al.* 1990). The sequence of the mammalian *lac* repressor is given in Figure 1 above, with the translation product of the NLS and alanine linker indicated at the 3′ end.

Production of transgenic mice: Mice transgenic for the bacterial gene coding for the *lac* repressor (*lacI*) were made by microinjection into (C57BL/6 × SJL)F_1 hybrid pronulear stage embryos. Thirteen founders and their offspring were analyzed for the data used in this article (01-14, Table 1). In addition, to circumvent anticipated problems with expression of *lacI* as a result of methylation, we attempted to derive *lacI* transgenics in inbred DBA/2J embryos. DBA/2J mice have been reported to hypomethylate transgene DNA (ENGLER *et al.* 1991). In a typical attempt, 325 eggs were harvested from 30 superovulated DBA/2J females; 100 of the 325 eggs had been fertilized, and 0 of 100 survived implantation and/or embryogenesis. We switched to hybrids of DBA/2J with SJL, another hypomethylating strain (ENGLER *et al.* 1991). Forty-seven eggs from DBA/2J females fertilized *in vitro* with sperm from SJL males were microinjected and cultured overnight. Thirty-four progressed to the two-cell stage and were implanted into isogenic foster mothers. Three animals were born and two of them were transgenic. One founder and the line derived from it were analyzed for the data used in this article (D2, Table 1).

DNA extraction: DNA was extracted from tissue by lysis in 10 mM Tris, pH 8; 10 mM NaCl; and 1 mM EDTA in the presence of 1% sarkosyl and proteinase K (100–200 μg/ml). Proteins and nucleic acids were separated by extracting three times with phenol and once with phenol/$CHCl_3$/isoamyl alcohol (25:24:1). Nucleic acids were precipitated from the aqueous phase by adding 0.5 volumes of 7.5 M ammonium acetate, and then 1 volume isopropanol. High molecular weight DNA was spooled onto a glass rod and transferred to TE (10 mM Tris/1 mM EDTA) and resuspended overnight. The DNA was reprecipitated the next day by the addition of 0.1 volume 3 M sodium acetate and 2 volumes of ethanol precooled to −20°, washed in 70% ethanol, and resuspended in TE. DNA was extracted from mouse tail by digesting 1 cm of tissue at 58° overnight in 700 μl 50 mM Tris-HCl, pH 8/100 mM EDTA, pH 8/100 mM NaCl/1% SDS to which was added 35 μl 10 mg/ml Proteinase K (BRL, Bethesda, MD) just prior to incubation. Following digestion with 20 μl of RNAseA (Sigma, St. Louis, MO; 10 mg/ml made up in water, then boiled 15 min to destroy DNAse) for 2 hr at 37°, the samples were extracted with equal volumes of phenol (3×), phenol/$CHCl_3$/isoamyl alcohol (25:24:1) (2×), or $CHCl_3$/isoamyl alcohol (24:1) (1×). The DNA was precipitated out of the aqueous phase by adding 0.5 M NH_4OAc to the top of the tube, spooled onto a glass rod, and suspended in 500 μl TE^{-4} (10 mM Tris/0.1 mM EDTA).

Southern blotting: DNA (5 μg) was electrophoresed through 0.8% agarose after digestion with appropriate restriction endonucleases (GIBCO/BRL). Transfer was to Hybond-N^+ (Amersham) in 0.25 M NaOH/1.5 M NaCl following treatment of the gel in 0.25 M HCl (2 × 15 min) and 0.5 M NaOH/1.5 M NaCl (30 min). The resultant Southern blot was dried at 80° for 10 min and then was prehybridized overnight at 42° in 5× SSC, 10× Denhardt's, 0.05 M phosphate (pH 6.7), 1% SDS, 5% dextran sulfate, and 50% formamide, with 500 μg ml^{-1} boiled salmon sperm DNA added. The blots were hybridized either at 42° or at 47° in 5× SSC, 2× Denhardt's, 0.02 M phosphate (pH 6.7), 1% SDS, 10% dextran sulfate, and 50% formamide, with 200 μg/ml salmon sperm DNA and 2×10^6 cpm/ml labeled probe added. Probes were labeled by random priming and were usually isolated inserts. After 36 hr hybridization, the blots were washed in 0.1× SSC/0.1% SDS, either at 55° or at 65°. The blots were subjected to autoradiography with Dupont/NEN Reflection film and intensifying screens.

RNA extraction and Northern blotting: Tissue was pulverized in a liquid nitrogen-cooled mortar and total RNA was extracted by the acid phenol method using TRI Reagent (Molecular Research Products, Inc., Cincinnati, OH). RNA was electrophoresed through a 1% agarose/0.67 M formaldehyde gel after formaldehyde denaturation for 15 min at 55°. RNA was transferred to a nylon membrane (Hybond-N; Amersham) by capillary blotting in 20× SSC and fixed to the membrane by UV crosslinking (Stratalinker; Stratagene). Blots were prehybridized overnight at 42° in 50% formamide, 5× Denhardt's, 5× SSPE, and 0.1% SDS, with 200 μg/ml boiled salmon sperm DNA and 1 μg/ml poly(A) added. Blots were hybridized at 42° in 50% formamide, 1× Denhardt's, 0.04 M phosphate (pH 6.7), and 0.05% SDS, with DNA and poly(A) added, as above, and 2×10^6 cpm/ml $\alpha[^{32}P]$dCTP labeled probes.

Probes: Probes were labeled by the method of random priming (FEINBERG and VOGELSTEIN 1983). *lacI* probes were derived from either the bacterial sequence contained in p*CMVlacI* or the mammalian sequence contained in the plasmid p*I5*, both in the plasmid collection of the Stambrook lab. The β-actin probe was derived from a murine cDNA (BUCKINGHAM 1985) given to us by MARGARET BUCKINGHAM. The MyoD probe was a murine cDNA in plasmid p*VZC13β* (DAVIS *et al.* 1987) given to us by ANDREW LASSAR, and p53 was a gift from ALAN BERNSTEIN (ROVINSKI *et al.* 1987).

TABLE 1

Comparison of copy number, methylation status, and expression levels of wild-type and modified *lacI* transgenes

Transgene	Coding sequence	Background	Copy number[a]	Methylation[b]	Expression[c]
01	Bacterial	B6SJL	High	High	None
02	Bacterial	B6SJL	High	High	None
03	Bacterial	B6SJL	Medium	High	None
04	Bacterial	B6SJL	High	High	None
05	Bacterial	B6SJL	Medium	High	None
06	Bacterial	B6SJL	Low	High	None
07	Bacterial	B6SJL	High	High	None
09	Bacterial	B6SJL	Low	High	None
10	Bacterial	B6SJL	Medium	High	None
11	Bacterial	B6SJL	Medium	High	None
12	Bacterial	B6SJL	Low	High	None
13	Bacterial	B6SJL	Low	High	None
14	Bacterial	B6SJL	High	High	None
D2	Bacterial	D2SJL	Low	High	Male germ cells only
26	Mammalian	B6SJL	Low	Low	Expressed, low
27	Mammalian	B6SJL	High	High	Expressed, low
34	Mammalian	B6SJL	Medium	Low	Ubiquitous, high
43	Mammalian	B6SJL	Low	Low	Ubiquitous, high
46	Mammalian	B6SJL	Low	Low	Ubiquitous, low
47	Mammalian	B6SJL	Low	ND	Expressed, low
51	Mammalian	B6SJL	Low	Low	Ubiquitous, low

[a] Copy number was estimated from genomic blots of *Eco*RI-digested DNA hybridized sequentially with probes to *lacI* and a single copy sequence (MyoD or p53). High ≤ 100; medium = 11–100; low = 1–10.

[b] Extent of methylation was determined by comparison of the sizes of *lacI* fragments detected in *Hpa*II- and *Msp*I-digested DNA. High, no digestion at sites within the *lacI* coding region; low, digestion of *lacI* sites; ND, not determined (material not available for analysis).

[c] Expression in various tissues from juvenile or adult mice was determined either by Northern blot (nos. 01–14; 27; 34; 43; 46; 51) or RT-PCR (nos. 26; 47) analysis of total RNA or by immunohistochemistry with anti-lacI antibody (D2).

RT-PCR: Following reverse transcription with avian myeloblastosis virus-reverse transcriptase of 1 μg total RNA in a reaction volume of 20 μl, 5 μl was immediately subjected to 35 rounds of amplification (95° 30 sec; 63° 15 sec; 72° 1 min) using primers homologous to the mammalian coding sequence. The sequence of the primers was 5′-tgggagcctctgtggtggtgtcaa-3′ (sense) and 5′-gcactccagtcaccttctctttca-3′ (antisense). The 373-bp PCR product was identified on 2% agarose or 10% polyacrylamide gels.

Immunohistochemistry: We generated a polyclonal antibody to the *lac* repressor in rabbits by the trpE fusion protein method (KOERNER *et al.* 1991). The 1.1-kilobase (kb) wild-type *lacI* gene was subcloned in all three reading frames into p*ATH* vectors, and the plasmid containing the fragment in the correct reading frame was identified as the one producing a 75-kD fusion protein that represents 38 kD of *lac* repressor and 37 kD of trpE. Sera from two different animals were found by ELISA to be immunoreactive against the fusion protein. The sera were affinity purified by chromatography through Affi-Gel (Bio-Rad) coupled to the fusion protein and depleted of antibodies recognizing trpE epitopes by passage through a similar column of Affi-Gel coupled to trpE. The affinity-purified, trpE-depleted anti-lacI antibody recognizes a protein of the correct size (38.6 kD) on Western blots. For immunohistochemistry, a sexually mature male mouse was given a heparin injection (0.13 U/g) 15 min prior to euthanasia, was flushed with heparinized saline, and was perfused with 4% paraformaldehyde for 30 min. The testes were removed, cryoprotected by bathing in 20% buffered sucrose overnight, and frozen. Cryostat sections (30 μm) were incubated with primary antibody (1:100) in 1% BSA/0.3% Triton X-100 in a humidified chamber at room temperature overnight. The next day, biotinylated goat anti-rabbit secondary (1:200) and ABC-Elite reagent (Vector Laboratories, Burlingame, CA) with DAB containing nickel ammonium sulfate and hydrogen peroxide were used to detect the immunocomplexes.

RESULTS

Hypermethylation and restriction of the *lacI* sequence from *E. coli* in the transgenic mouse: The *lac* repressor transgene (Figure 2) consisted of the wild-type *lacI* gene from *E. coli* driven by the 4.3-kb human β-actin promoter (LEAVITT *et al.* 1984) and the polyadenylation signal from the bovine growth hormone gene (WOYCHIK *et al.* 1982). Southern blots were prepared from *Eco*RI-restricted DNA

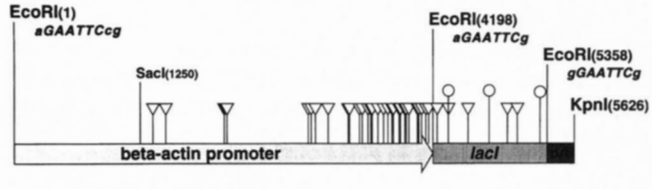

FIGURE 2.—*lacI* transgene. Position of restriction sites for various enzymes is indicated. Sequences flanking *Eco*RI sites in the bacterial *lacI* transgene are given above each *Eco*RI site. In the mammalian version, there is no *Eco*RI cloning site at the 5′ end of *lacI*. *Msp*I/*Hpa*II sites are indicated by stemmed triangles. The three remaining *Msp*I/*Hpa*II sites in the mammalian version of *lacI* are indicated by circles.

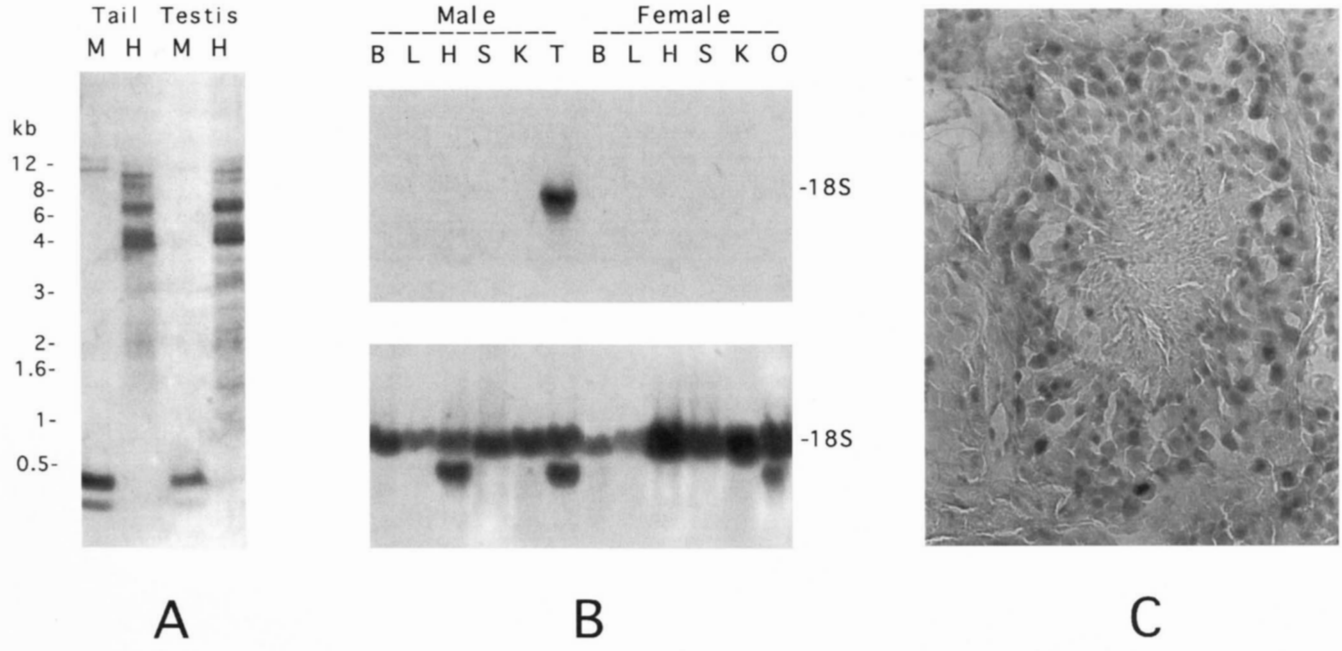

FIGURE 3.—Methylation and expression of bacterial *lacI* transgenes. (A) Southern blot of *Hpa*II (H)- and *Msp*I (M)-digested DNA from tail or testis hybridized to *lacI*. (B) Northern blot of 20 μg total RNA hybridized sequentially to *lacI* (top) and β-actin (bottom). The β-actin cDNA probe hybridizes to several actin species, including cardiac actin in heart and skeletal actin in muscle, in addition to cytoplasmic actin in all tissues. The location of the 18S ribosomal RNA species, which in the mouse is 1950 bases in size, was taken from the ethydium bromide-stained membrane following RNA transfer. B, brain; L, lung; H, heart; S, spleen; K, kidney; T, testis; O, ovary. (C) Section from testis of adult male stained with an antibody against the *lac* repressor.

from founder animals and hybridized to a probe homologous to *lacI*. As indicated in Figure 2, *Eco*RI sites flank the *lacI* fragment, which should hybridize to the probe as a 1.2-kb band. Instead of the predicted fragment, however, there were ladders of high molecular weight fragments hybridizing to the probe, suggesting that transgene DNA was completely inaccessible to the enzyme (data not shown). To determine if and to what extent methylation was the source of *Eco*RI inaccessibility, founder DNA was digested with *Msp*I or *Hpa*II, which recognize the same sequence (CCGG) but are differentially sensitive to cytosine methylation. An identical pattern was observed in the DNA of 14 independently derived founder animals carrying the bacterial *lacI* transgene. In all 14, the *lacI* probe hybridized to several small fragments in DNA digested with *Msp*I, which is not affected by methylation, but only to ladders of high-molecular-weight DNA when digested with *Hpa*II, which is methylation sensitive, as shown for a representative animal in Figure 3A (tail). This pattern was maintained in animals from subsequent generations in lines derived from eight individual founders (02, 03, 04, 05, 06, 07, 09, and D2; Table 1).

To determine if bacterial *lacI* mRNA was present in transgenic animals, Northern blots of total RNA were hybridized to a probe complementary to the *lacI* coding sequence. Expression was undetectable in all tissues derived from both founders and their offspring in all 13 lines on a B6SJL background (data not shown). In animals in the D2 line, which was derived on a D2SJL hybrid background and maintained on a DBA/2J inbred background, expression was undetectable in all tissues except the testis, as shown for a representative animal in Figure 3B. In the testis, expression was induced from undetectable levels in prepubertal animals (2-wk-old males; data not shown) to very high levels in sexually mature animals. Expression was detected in spermatogonia and primary spermatocytes, but not in any of the somatic tissues of the testis, as determined by immunohistochemistry using an antirepressor antibody (Figure 3C). There was no concomitant expression of the transgene in the ovary at any stage (Figure 3B).

The methylation state of *lacI* in testis DNA from sexually mature animals was analyzed by comparing *Msp*I and *Hpa*II digestion patterns. There was no detectable difference in the overall degree of methylation in testis DNA, as the pattern of *Hpa*II fragments resembled that of tail DNA (Figure 3A). A similar result was obtained using purified sperm DNA in place of whole testis (data not shown).

As summarized in Table 1, bacterial *lacI* sequences resident at 14 unique transgene loci that differed by copy number and insertion site, in two different hybrid genetic backgrounds, were uniformly hypermethylated and not expressed in any somatic tissues in the mouse.

Variable methylation and ubiquitous expression of a mammalian homologue of the *lacI* gene in the transgenic mouse: We modified the wild-type *lacI* DNA sequence such that the same amino acids were re-encoded with mammalian codons and replaced the bacte-

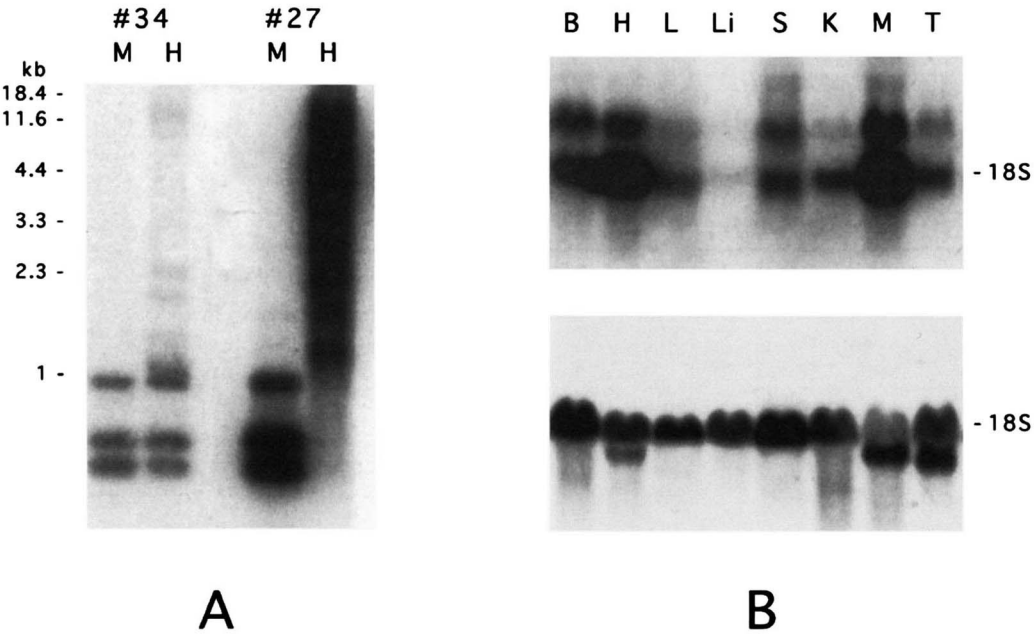

FIGURE 4.—Methylation and expression of mammalian *lacI* transgenes. (A) Southern blot of *Hpa*II (H)- and *Msp*I (M)-digested DNA from tail of low (no. 34) or high (no. 27) copy number transgenes hybridized to *lacI*. (B) Northern blot of total RNA hybridized to a probe complementary to the mammalian *lacI* sequence (top) and rehybridized to the β-actin probe (bottom). The upper band represents the unprocessed transcript that includes the intervening sequence in the proximal β-actin promoter. B, brain; H, heart; L, lung; Li, liver; S, spleen; K, kidney; M, muscle; T, testis.

rial sequence with the mammalian *lacI* sequence in the transgene construct used previously (Figure 2). The methylation status of seven individual mammalian *lacI* transgene loci was assayed by comparison of *Hpa*II and *Msp*I patterns on Southern blots of restricted DNA hybridized to a mammalian *lacI* fragment labeled with $\alpha[^{32}P]$-dCTP. As a result of removing CpG dinucleotides from the bacterial *lacI* coding sequence, the number of *Msp*I/*Hpa*II sites in the mammalian sequence was reduced from five to three. Six of the seven mammalian transgenes had a methylation pattern similar to no. 34, shown in Figure 4A. In addition to the same small fragments seen in *Msp*I-digested DNA, there was some hybridization to the higher-molecular-weight *Hpa*II fragment corresponding to the entire transgene. The remaining animal (no. 27; Figure 4A) had a hypermethylated transgene locus resembling those with the bacterial sequence.

In contrast to the bacterial repressor, all (7/7) mammalian transgenes were expressed (Table 1). When we examined *lacI* expression in greater detail by Northern blot analysis of RNA extracted from various tissues from adult animals with four different transgenes (nos. 34, 43, 46, and 51), we detected *lacI* expression in all tissues tested. Levels of expression of transgene no. 34, as shown in Figure 4B, ranged from very high (heart and muscle), to moderate (brain, lung, and spleen), to low (kidney and gonads). Lowest expression was seen in the liver. The expression pattern of transgene no. 43 was virtually identical (data not shown). Expression of nos. 46 and 51 was also ubiquitous, but at a lower level overall (Table 1).

As summarized in Table 1, transgenes encoding the *lac* repressor with a mammalian version of *lacI* from which 90/95 (95%) of the CpG dinucleotides were removed were all expressed. This includes transgene loci with only a few to those with very many copies of the resident transgene, as well as transgenes in both high and low methylation states.

DISCUSSION

Changing the bacterial coding sequence of the *lac* repressor to reflect the more AT-rich mammalian genome was the single most important factor affecting transgene expression. The bacterial sequence is high in G + C (65%) and encodes amino acids with GC-rich codons that are relatively rare in mammals (ZHANG *et al.* 1991). In Table 2, codon usage is compared between the wild-type (bacterial) and our modified (mammalian) sequences for each degenerate triplet set encoding a single amino acid. Particular attention was paid to low usage codons (indicated in boldface type), which were replaced with more common mammalian codons wherever possible. As a consequence, the number of CpG dinucleotides was reduced from 95 to five. Thus, not only did the re-encoded sequence have access to more abundant transfer RNAs for each of the amino acids, but also most of the methyltransferase substrates had been removed. As a result, the overall methylation state changed from being hypermethylated in the bacterially encoded *lacI* transgenes to mostly unmethylated in all but the highest copy number mammalian *lacI* transgenes, and the expression pattern changed from being restricted to ubiquitous, respectively.

The possibility that the restricted expression pattern of the bacterial *lacI* sequence results from the decay of highly unstable RNA transcripts rather than a block to transcription has not been formally excluded. That is, the failure to detect transcripts on Northern blots of total RNA derived from the various somatic tissues of the animal could reflect differential processing of mRNA in so-

TABLE 2

Codon usage in wild-type and modified *lacI* sequences

Amino acid[a]	Codon[b]	No. in wild-type	No. in synthetic	Percent of total	
				Wild-type	Synthetic
Leu[c]	**UUA**	**5**	**0**	**12.5**	**0**
	UUG	5	6	12.5	14.6
	CUU	2	3	5	7.3
	CUC	3	5	7.5	12.2
	CUA	**0**	**1**	**0**	**2.4**
	CUG	25	26	62.5	63.4
Ile	AUU	10	12	55.5	66.7
	AUC	7	5	38.9	27.8
	AUA	**1**	**1**	**5.6**	**5.6**
Val	GUU	8	1	22.9	31.4
	GUC	9	5	25.7	14.3
	GUA	**3**	**3**	**8.6**	**8.6**
	GUG	15	16	42.8	45.7
Ser	UCU	7	11	21.2	33.3
	UCC	6	4	18.2	12.1
	UCA	4	7	12.1	21.2
	UCG	**5**	**0**	**15.1**	**0**
	AGU	5	8	15.1	24.2
	AGC	6	3	18.2	9.1
Pro[c]	CCU	0	0	0	0
	CCC	6	7	42.9	46.7
	CCA	2	8	14.3	53.3
	CCG	**6**	**0**	**42.9**	**0**
Thr	ACU	3	7	15.8	36.8
	ACC	11	8	57.9	42.1
	ACA	1	4	5.3	21.0
	ACG	**4**	**0**	**21.0**	**0**
Ala	GCU	3	7	6.8	15.2
	GCC	14	16	31.8	34.8
	GCA	5	22	11.4	47.8
	GCG	**22**	**1**	**50.0**	**2.2**
Arg	**CGU**	**2**	**0**	**10.5**	**0**
	CGC	**10**	**0**	**52.6**	**0**
	CGA	**4**	**1**	**21.1**	**5**
	CGG	**2**	**1**	**10.5**	**5**
	AGA	1	13	5.3	65.0
	AGG	0	5	0	25
Tyr	**UAC**	**3**	**1**	**37.5**	**12.5**
	UAU	**5**	**7**	**62.5**	**87.5**

[a] The remaining amino acids (Lys, Gly, Gln, Glu, Asn, His, Asp, Cys, Phe, Met, and Trp) are not included because they do not contain any low usage codons and were not significantly affected by the sequence changes.

[b] Low usage codons are indicated in boldface type.

[c] There are discrepancies in the total numbers of Leu and Pro (as well as Lys) because the synthetic sequence includes the SV40 nuclear localization signal and a linker sequence connecting it to the heptad repeat region important for dimerization of repressor monomers.

matic cells and male germ cells that leaves the RNA in all but the spermatogonial cells vulnerable to decay. This is unlikely, however, for two reasons. First, *lacI* mRNA is stable in male germ cells, where it is translated into repressor protein that is recognized by antirepressor antibody in these cells. Second, cultured somatic cells of mammalian origin (including mouse) transfected with the identical β-actin/bacterial *lacI* transgene used in the animals described here express both the message and the protein encoded by it, as we have previously described (LIU *et al.* 1992). In these cultured cells, the transgene also remains hypomethylated. In aggregate, these observations favor a mechanism in which methylation differences lead to alterations in transcriptional activity rather than one in which transcripts produced in somatic and germ cells are rendered differentially unstable. Support for the former mechanism has been described for the testis-specific expression of another transgene driven by the ubiquitous metallothionein promoter (SALEHI-ASHTIANI *et al.* 1993).

The difference in the way the highly divergent bacterial sequence is methylated and expressed in the context of the transgenic mouse might reflect a normal process that protects the genome from acquiring new material. By necessity, a system that recognizes and marks DNA as foreign as a way of protecting the endogenous genome must be relaxed during the process of fertilization, when the oocyte is invaded by a sperm carrying a completely different haploid genome. Indeed, it may be this window of acceptance that allows the artificial introduction of new genetic material by transgenesis to occur. Once introduced, this material becomes subject to the same mechanisms that impose methylation states on endogenous genes (FRANK *et al.* 1991). Structures similar to that of the *lac* repressor transgenes used in this study, in which the CpG island in the 3′ region of the β-actin promoter is juxtaposed with the *lacI* coding sequence, have been shown to be the targets of coordinate demethylation and *de novo* methylation during early development (FRANK *et al.* 1991) and spermatogenesis (KAFRI *et al.* 1992) and to be associated with stage-specific transcription (SINGER-SAM *et al.* 1990). This could explain why the bacterial *lacI* sequence can be expressed during spermatogenesis from a hypermethylated transgene that is silenced in all somatic tissues. However, there may be limits on what material is allowed to be introduced, either in terms of how closely the new genetic material resembles the host DNA or how big it is. Hypermethylation of the highly divergent bacterial sequence and of the mammalian sequence at high copy number could reflect what happens when those limits are exceeded. The existence of a mechanism that senses aberrant stretches of sequence could explain the hypermethylation of CpG islands adjacent to triplet repeat regions that accompanies expansion at the fragile X loci in humans (OBERLÉ *et al.* 1991; HANSEN *et al.* 1992; HORNSTRA *et al.* 1993; KNIGHT *et al.* 1993). If methylation is part of a system that monitors the fate of newly acquired DNA in mice and humans (see BESTOR and TYCHO 1996 for review), then, by analogy to restriction-modification systems in bacteria, we should also expect to identify an activity that removes illegitimate sequences from the genome. One way of effectively removing such a sequence would be by silencing its expression, and we have demonstrated an example of that. The recent observation that the instability of both myotonic dystrophy- (MONCKTON *et al.* 1997; GOURDON *et al.* 1997) and Huntington disease- (MANGIARINI *et al.* 1997) like triplet repeats in transgenic mice that can undergo both deletions and expansions are under genetic control supports the idea of a dynamic process that maintains the integrity of the mammalian genome.

The authors gratefully acknowledge the help of their co-workers and colleagues, especially W. GLUBA of the Department of Neuroscience, University of Virginia, and Z. SEVER, M. MCFARLAND, G. COLWELL, D. FORREST, and W. BI of the Department of Cell Biology, Neurobiology, and Anatomy at the University of Cincinnati, for technical assistance; and, J. YUN and T. WAGNER of the Edison BioTechnology Institute for constructing *lacI* Founder animals. This work was supported in the initial stages by a postdoctoral fellowship to H.S. from the Jane Coffin Childs Memorial Fund for Medical Research and is currently supported by National Institutes of Health grants R24RR11102 to H.S. and T32 ES07250 and P01 ES05652 to P.J.S.

LITERATURE CITED

BEDDINGTON, R. S. P., J. MORGENSTERN, H. LAND and A. HOGAN, 1989 An *in situ* transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner-cell mass clones during early organogenesis. Development **106:** 37–46.

BESTOR, T. H., and B. TYCHO, 1996 Creation of genomic methylation patterns. Nature Gen. **12:** 363–367.

BOWMAN, S., J. A. TISCHFIELD and P. J. STAMBROOK, 1990 An efficient and simplified method for producing site-directed mutations by PCR. Technique **2:** 254–260.

BUCKINGHAM, M., 1985 Actin and myosin multigene families. Essays Biochem. **20:** 77–109.

CUI, C., M. A. WANI, D. WIGHT, J. KOPCHICK and P. J. STAMBROOK, 1994 Reporter genes in transgenic mice. Transgenic Res. **3:** 182–194.

DAVIS, R. L., H. WEINTRAUB and A. B. LASSAR, 1987 Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51:** 987–1000.

ENGLER, P., D. HAASCH, C. A. PINKERT, L. DOGLIO, M. GLYMOUR *et al.*, 1991 A strain-specific modifier on mouse chromosome 4 controls the methylation of independent transgene loci. Cell **65:** 939–947.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Ann. Biochem. **132:** 6–13.

FRANK, D., L. KESHET, M. SHANI, A. LEVINE, A. RAZIN *et al.*, 1991 Demethylation of CpG islands in embryonic cells. Nature **351:** 239–241.

GOURDON, G. F., F. RADVANYI, A. S. LIA, C. DUROS, M. BLANCHE *et al.*, 1997 Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. Nature Gen. **15:** 190–192.

HANSEN, R. S., S. M. GARTLER, C. R. SCOTT, S.-H. CHEN and C. D. LAIRD, 1992 Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. Hum. Mol. Genet. **1:** 571–578.

HORNSTRA, I. K., D. L. NELSON, S. T. WARREN and T. P. YANG, 1993 High resolution methylation analysis of the FMR1 gene tricucleotide repeat region in Fragile X syndrome. Hum. Mol. Genet. **2:** 1659–1665.

JACOB, F., and J. MONOD, 1961 Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3:** 318–356.

KAFRI, T., M. ARIEL, M. BRANDEIS, R. SHEMER, L. URVEN *et al.*, 1992 Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev. **6:** 705–714.

KNIGHT, S. J. L., A. V. FLANNERY, M. C. HIRST, L. CAMPBELL, Z. CHRISTODOULOU *et al.*, 1993 Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. Cell **74:** 127–134.

KOERNER, T. J., J. E. HILL, A. M. MYERS and A. TZAGOLOFF, 1991 High-expression vectors with multiple cloning sites for construction of trpE-fusion genes: pATH vectors. Methods Enzymol. **194:** 477–490.

LEAVITT, J., P. GUNNING, P. PORRECA, S. Y. NG, C. S. LIN *et al.*, 1984 Molecular cloning and characterization of mutant and wild-type human beta-actin genes. Mol. Cell. Biol. **4:** 1961–1969.

LIU, H.-S., H. SCRABLE, D. B. VILLARET, M. A. LIEBERMAN and P. J. STAMBROOK, 1992 Control of Ha-*ras*-mediated mammalian cell transformation by *Escherichia coli* regulatory elements. Cancer Res. **52:** 983–989.

MANGIARINI, L., K. SATHASIVAM, A. MAHAL, R. MOTT, M. SELLER *et al.*, 1997 Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nature Gen. **15:** 197–200.

MONCKTON, D. G., M. I. COOLBAUGH, K. T. ASHIZAWA, M. J. SICILIANO and C. T. CASKEY, 1997 Hypermutable myotonic dystrophy CTG repeats in transgenic mice. Nature Gen. **15:** 193–196.

NILSSON, E., and U. LENDAHL, 1993 Transient expression of a human β-actin promoter/lacZ gene introduced into mouse embryos correlates with a low degree of methylation. Mol. Reprod. Dev. **34:** 149–157.

OBERLÉ, I., F. ROUSSEAU, D. HEITZ, C. KRETZ, D. DEVYS *et al.*, 1991 Instability of a 550-base pair DNA segment and abnormal methylation in Fragile X syndrome. Science **252:** 1097–1102.

ROVINSKI, B., D. MUNROE, J. PEACOCK, M. MOWAT, A. BERNSTEIN *et al.*, 1987 Deletion of 5′-coding sequences of the cellular p53 gene in mouse erythroleukemia: a novel mechanism of oncogene regulation. Mol. Cell. Biol. **7:** 847–853.

SALEHI-ASHTIANI, K., R. J. WIDROW, C. L. MARKERT, and E. GOLDBERG, 1993 Testis-specific expression of a metallothionein I-driven transgene correlates with undermethylation of the locus in testicular DNA. Proc. Natl. Acad. Sci. USA **90:** 8886–8890.

SINGER-SAM, J., M. ROBINSON, A. R. BELLVE, M. I. SIMON and A. D. RIGGS, 1990 Measurement by quantitative PCR of changes in HPRT, PGK-1, PGK-2, APRT, MTase, and Zfy gene transcripts during mouse spermatogenesis. Nucleic Acids Res. **18:** 1255–1259.

WOYCHIK, R. P., S. A. CAMPER, R. H. LYONS, S. HOROWITZ, E. C. GOODWIN *et al.*, 1982 Cloning and nucleotide sequencing of the bovine growth hormone gene. Nucleic Acids Res. **10:** 7197–7210.

ZHANG, S., G. ZUBAY, and E. GOLDMAN, 1991 Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. Gene **105:** 61–72.

Communicating editor: N. A. JENKINS