MOLECULAR AND CELLULAR BIOLOGY, Oct. 1991, p. 5154–5163
0270-7306/91/105154-10$02.00/0

Vol. 11, No. 10

Retrovirus-Induced Insertional Mutagenesis: Mechanism of Collagen Mutation in Mov13 Mice

DOUGLAS D. BARKER,[1] HONG WU,[1] STEFAN HARTUNG,[2] MICHAEL BREINDL,[2] AND RUDOLPH JAENISCH[1,3]*

Whitehead Institute for Biomedical Research, Nine Cambridge Center,[1] and Department of Biology, Massachusetts Institute of Technology,[3] Cambridge, Massachusetts 02142, and Department of Biology, College of Sciences, San Diego State University, San Diego, California 92182[2]

Received 14 February 1991/Accepted 15 July 1991

The Mov13 mouse strain carries a mutation in the α1(I) procollagen gene which is due to the insertion of a Moloney murine leukemia provirus into the first intron. This insertion results in the de novo methylation of the provirus and flanking DNA, the alteration of chromatin structure, and the transcriptional inactivity of the collagen promoter. To address the mechanism of mutagenesis, we reintroduced a cloned and therefore demethylated version of the Mov13 mutant allele into mouse fibroblasts. The transfected gene was not transcribed, indicating that the transcriptional defect was not due to the hypermethylation. Rather, this result strongly suggests that the mutation is due to the displacement or disruption of *cis*-acting regulatory DNA sequences within the first intron. We also constructed a Mov13 variant allele containing a single long terminal repeat instead of the whole provirus. This construct also failed to express mRNA, indicating that the Mov13 mutation does not revert by provirus excision as has been observed for other retrovirus-induced mutations.

Murine retroviruses have extensive structural and functional similarities to retrotransposons such as the Ty elements of yeast and the copia-like elements of *Drosophila melanogaster* and share the ability to induce insertional mutations upon integration into the host genome. The analysis of retrotransposon-induced mutations in yeasts and *Drosophila melanogaster* has revealed a wide variety of different mechanisms responsible for these mutations, including both dominant activating mutations and recessive inactivating mutations. Dominant activating mutations induced by Ty and copia-like elements frequently result from the insertion of a retrotransposon enhancer element in proximity to a promoter, resulting in the inappropriate transcription of the target gene (2, 3). Similar mutations can be induced by murine retroviruses and have been identified upon activation of proto-oncogenes (52).

Recessive inactivating mutations induced by Ty and copia-like elements have been observed upon integration into either the coding or noncoding regions of the target gene. Mutations caused by integration into the coding sequence of a gene rarely revert, reflecting the low frequency of precise excision of these elements. Inactivating mutations caused by insertions into noncoding regions, such as introns and 5′ flanking regions, can be the result of the insertion of a premature polyadenylation signal (24, 57) or caused by the disruption or displacement of essential promoter elements (2, 3). Some mutations in this latter class can revert by homologous recombination between the two long terminal repeats (LTRs), thus eliminating the element. In addition, some insertional mutants can be suppressed by mutations in unlinked genes (2, 3, 31, 33, 34, 43, 54, 57).

In mice, the insertion of retroviral proviruses has been shown to be the cause of the spontaneous mutations dilute (20) and hairless (49). In both mutant strains, the proviruses have apparently integrated into noncoding sequences since no open reading frames have been identified in the flanking DNA sequences. Furthermore, the mutant phenotype has reverted spontaneously by homologous recombination between the LTRs, leaving a solo LTR in the revertant allele (8, 41, 49). Retrovirus-induced mutations have also been shown to result from the activation of an endogenous retrovirus during oogenesis and subsequent germ line integration (46). Similarly, experimental infection of mouse embryos with virus has resulted in the induction of a number of recessive lethal mutations leading to either embryonic (44) or adult (53) lethality.

The Mov13 mutation was generated by retrovirus infection of midgestation mouse embryos, and the molecular parameters associated with this virus-induced insertional mutation have been studied in detail. The provirus has integrated into the first intron of the α1(I) procollagen gene with its transcriptional orientation opposite that of the collagen gene (40). This insertion results in a recessive lethal mutation arresting development between days 11 and 14 of gestation (11, 13, 28) due to the failure to initiate transcription of collagen mRNA (12). The Mov13 provirus is transcriptionally repressed throughout most of embryonic development (14, 15) and induces de novo methylation of both proviral and flanking collagen DNA sequences (16–18). Furthermore, the chromatin structure of the collagen (*Col1a1*) gene is altered. The provirus prevents the developmentally regulated formation of a DNase I-hypersensitive site in the 5′ flanking sequences which correlates with active collagen gene transcription (6).

Several molecular mechanisms may explain the failure of the Mov13 collagen gene to be expressed. For example, the transcriptional defect of the collagen gene may be a consequence of altered DNA methylation induced upon integration of the Moloney murine leukemia virus (MoMLV) provirus. Transcriptional inactivity and de novo methylation of proviruses are characteristic for retrovirus infection of preimplantation embryos (19) and embryonal carcinoma cells (10, 27, 32, 48) and are seen in proviruses which are transmitted through the germ line (50). It is possible that the insertion of the provirus displaces *cis*-acting regulatory

* Corresponding author.

sequences, leading to de novo methylation and chromatin alterations of the collagen gene. These alterations may constitute the primary cause of the transcriptional inactivation of the Mov13 mutant gene by preventing assembly of collagen gene sequences into a transcriptionally competent chromatin structure or by precluding the binding of transcription factors.

Another possibility is that the insertion of the provirus may disrupt or displace DNA sequence elements essential for the transcription of the collagen gene. The mouse α1(I) procollagen gene, like other mouse and human collagen genes, contains a complex combination of *cis*-acting DNA sequence elements which have positive and negative effects on transcription (4, 5, 26, 35–38). These regulatory elements have been located in both the 5′ flanking sequences and within the first intron. The insertion of a 9-kb provirus therefore may physically disrupt one of these elements or displace it from the promoter and thus prevent transcription. In this case, the methylation of the collagen gene may be a consequence of its transcriptional inactivity rather than its cause.

If collagen transcription in Mov13 mice is prevented by the methylation of important DNA sequences, the removal of this modification should restore transcriptional activity. If, however, the proviral insertion had displaced or disrupted a DNA sequence element essential for transcription, demethylation should have no effect on the transcriptional defect. This study attempts to distinguish between these alternatives by introducing a cloned version of the Mov13 mutant allele gene into fibroblasts. We found that removal of the Mov13 methylation pattern did not restore transcription upon reintroduction into mouse fibroblast cells. Likewise, the Mov13 variant gene carrying a single proviral LTR instead of the provirus was transcriptionally inactive. These results strongly suggest that the MoMLV provirus insertion disrupts DNA sequences in the first intron of the collagen gene which are essential for its transcription.

MATERIALS AND METHODS

Cell lines and culture conditions. The homozygous Mov13 mouse embryo fibroblast cell lines M13-5 and 4Cl4 were derived from homozygous day 12 embryos and immortalized by infection with simian virus 40 as previously described (39). All cell lines were maintained as monolayers and grown in Dulbecco's modified Eagle's medium (GIBCO-BRL) supplemented with 10% fetal calf serum. DNA transfections were performed according to the method of van der Eb and Graham (51), and G418 (GIBCO-BRL)-resistant cultures were selected as described by Southern and Berg (45).

Construction of plasmids pWTC-Mov and pWTC-LTR. The construction of pWTC has been described previously (47, 56). pWTC contains the wild-type mouse α1(I) procollagen gene which has been marked by the insertion of a 21-bp *Xba*I-*Bam*HI-*Xba*I linker in the 5′ untranslated region of the procollagen transcript in order to distinguish it from the endogenous collagen gene.

pWTC-Mov was constructed by a stepwise construction of several plasmids, beginning with pWTC-14. pWTC-14 contains the 5′-most 14-kb *Eco*RI fragment of pWTC and was constructed by digesting pWTC with *Eco*RI followed by circularization. pWTC-14-Mov was then constructed by the ligation of two fragments: one, an *Xba*I (partial)-*Sna*BI fragment, derived from pMov13 (7), extending from the *Xba*I site in the first exon to the only *Sna*BI site in the gene, about 3.55 kb 3′ of the transcriptional start site; the second, another *Xba*I (partial)-*Sna*BI fragment derived from pWTC-14 extending from the same unique *Sna*BI site 3′ into the plasmid vector and through the 5′ portion of gene up to the marked *Xba*I site within the first exon. pWTC-Mov was then constructed by replacing a fragment extending from the *Sal*I site in the vector polylinker 5′ to the procollagen gene of pWTC to the unique *Sna*BI site with the equivalent *Sal*I (partial)-*Sna*BI fragment of pWTC-14-Mov (Fig. 1). Unique *Not*I restriction sites were then introduced at the position of the *Sal*I site in the vector polylinker or at the position of the *Sal*I site within the MoMLV provirus by linearization with a *Sal*I partial digest, fill-in of the *Sal*I sites with DNA polymerase Klenow fragment, and ligation of *Not*I linkers (New England Biolabs). The two resulting plasmids used in the experiments described here are pWTC-Mov, which contains a unique *Not*I site in the polylinker of the plasmid, and pWTC-Mov-N, which contains a unique *Not*I site in the MoMLV virus sequences. The insertion of this *Not*I linker into the MoMLV sequences results in an in-frame insertion of four amino acids into the *env* gene.

To construct pWTC-LTR, plasmid pWTC-14-Mov was digested with *Nhe*I and recircularized by ligation to form pWTC-14-LTR. *Nhe*I cuts once within each LTR of the MoMLV provirus; therefore, ligation of these two sites results in the formation of a single LTR. To form pWTC-LTR, the *Sal*I-*Sna*BI fragment containing the LTR was used to replace the equivalent fragment of pWTC (Fig. 1).

Preparation and analysis of nucleic acids from transfected cells. Cell cultures were scraped from their plates in *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES)-buffered saline and pelleted by centrifugation at 1,000 rpm for 5 min at 4°C. The cells were washed once with HEPES-buffered saline and lysed by resuspension in Nonidet P-40 (NP-40) lysis buffer (0.15 M NaCl, 10 mM Tris [pH 7.5], 1.5 mM $MgCl_2$, 0.65% NP-40) at 0°C. The nuclei were pelleted by centrifugation at 1,000 rpm for 10 min at 0°C. RNA was prepared by mixing the cytoplasmic supernatant with an equal volume of 7 M urea–0.35 N NaCl–0.01 M Tris (pH 7.4)–0.01 M EDTA–1% sodium dodecyl sulfate (SDS), phenol-chloroform extracting twice, and precipitating with ethanol. The ethanol pellet was washed twice with 95% ethanol–0.4 M $NaCH_3CO_2$ (pH 5.2), 2.5:1, and resuspended in H_2O. Genomic DNA was prepared from the nuclear pellet. The nuclei were washed once with NP-40 lysis buffer, pelleted, resuspended in 0.15 M NaCl–10 mM Tris (pH 8.0)–1 mM EDTA, and disrupted by the addition of 0.15 M NaCl–10 mM Tris (pH 8.0)–1 mM EDTA–0.4% SDS–200 μg of proteinase K per ml. After digestion at 65°C, this solution was extracted with phenol-chloroform twice, precipitated with ethanol, and resuspended in 10 mM Tris (pH 8.0)–1 mM EDTA.

Expression of cytoplasmic mRNA was assayed by S1 nuclease analysis performed essentially as described by Berk and Sharp (1). Hybridizations were performed in 50 μl of 80% formamide–0.4 M NaCl–40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4)–1 mM EDTA, and digestions were initiated with the addition of 450 μl of 0.25 M NaCl–30 mM $NaCH_3CO_2$ (pH 4.6)–1 mM $ZnCl_2$–200 U of S1 nuclease (GIBCO-BRL) at 0°C and incubated at 25°C for 60 min. Expression of mRNA from the transfected mouse α1(I) procollagen plasmid was detected with a 600-bp *Ava*I fragment, derived from pWTC, which was labelled with ^{32}P at its 3′ end. Expression of mRNA from the endogenous mouse α2(I) procollagen was detected with either of two fragments derived from pAZ1002 (25), a 5′-end-labelled 227-bp *Acc*I-*Xho*I or a 3′-end-labelled 504-bp *Bst*EII-*Stu*I

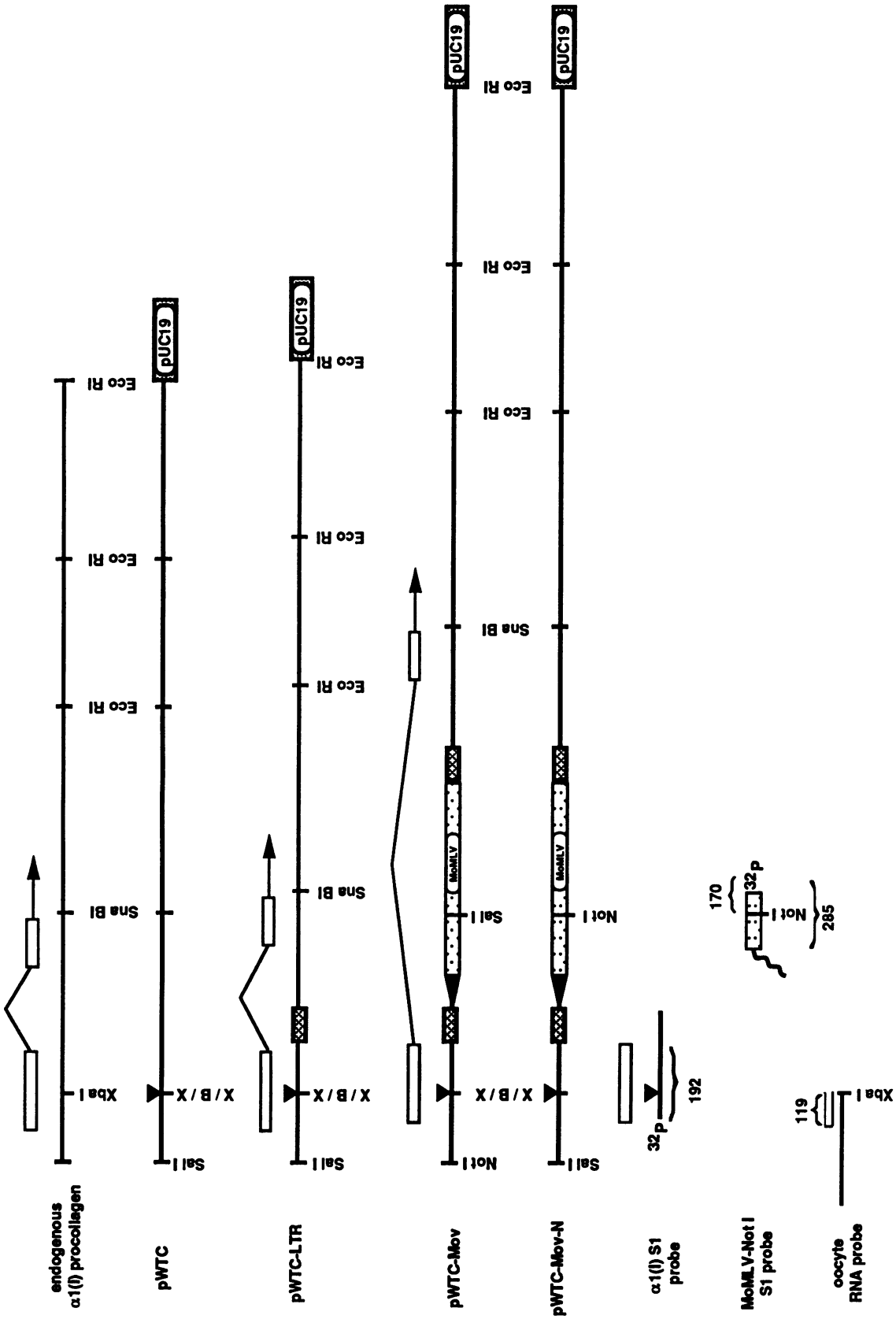

FIG. 1. Map of α1(I) procollagen plasmids and structure of the S1 and RNase protection probes. Procollagen DNA sequences are indicated by the horizontal lines. Positions of the first two mRNA exons are indicated by open boxes. Each arrowhead marked X/B/X indicates the position of the insertion of the *Bam*HI linker within an *Xba*I site. This insertion is within the 5′ untranslated portion of the mRNA. The position of the MoMLV insertion is indicated for pWTC-Mov and pWTC-Mov-N by stippled boxes, and the LTRs are indicated by hatched boxes. The position of the 600-bp *Ava*I fragment used in S1 nuclease assays to detect α1(I) procollagen mRNA expressed from the transfected plasmids is shown. A 192-base protected fragment is generated extending from the *Ava*I site within the first exon to the 5′ splice site of the first intron. The position of the subcloned MoMLV *Bam*HI-*Pvu*II fragment used in S1 nuclease assays to detect expression of RNA from the transfected provirus is shown. A 285-base fragment is protected by RNA containing the inserted *Not*I site, and a 170-base fragment is seen for MoMLV RNAs without the *Not*I site. The position of the fragment used to transcribe a uniformly labelled RNA probe used to detect procollagen RNA expressed in injected oocytes is shown. Also shown is a 119-base protected fragment representing transcripts initiated at the α1(I) procollagen promoter.

fragment. Expression of mRNA from the provirus inserted into the transfected Mov13 gene was detected by using a subcloned fragment encompassing the inserted *Not*I site of plasmid pWTC-Mov-N. This fragment was 3′ end labelled with ^{32}P at the *Bam*HI site at nucleotide 3535 of the MoMLV sequence (42) and extended beyond the *Pvu*II site at nucleotide 3805 into plasmid sequences.

Estimation of the relative copy number of transfected plasmids was performed by Southern blots performed as described previously (29). Genomic DNAs were digested with restriction endonuclease *Pst*I, transferred to nylon membranes (Biotrans; ICN), and probed with a 325-bp *Pvu*II fragment derived from the *lac* region of pUC19. This fragment detects only the transfected procollagen plasmids and not the cotransfected pSV2-neo, which is derived from pBR322.

To perform in vitro nuclear transcription assays, washed cells were lysed in NP-40 lysis buffer at 0°C. Nuclei were washed once with NP-40 lysis buffer and once with nuclear transcription buffer (20 mM Tris [pH 8.0], 140 mM KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, 20% glycerol) and resuspended in 0.2 ml of nuclear transcription buffer plus 1 mM ATP, 1 mM CTP, 1 mM GTP, 1 mM UTP, and 250 mCi of [α-^{32}P]UTP (3,000 Ci/mmol). After incubation at 30°C for 15 min, the nuclei were pelleted by centrifugation at 1,500 rpm for 5 min, resuspended in 0.4 ml of 20 mM Tris (pH 7.5)–10 mM NaCl–10 mM $MgCl_2$–10 mM $CaCl_2$ with 100 U of DNase I, and incubated at 30°C for a further 30 min. Then SDS was added to 1% and proteinase K was added to 100 μg/ml, and digestion continued for 30 min at 37°C. After phenol-chloroform extraction, the RNA was precipitated with ethanol, resuspended, and separated from unincorporated nucleoside triphosphates by Sephadex G-50 chromatography. Radioactive incorporation was determined by Cerenkov counts after spotting RNA onto DE81 paper, washing with 0.5 M $NaPO_4$, and drying. Equal counts of RNA were then hybridized to identical Southern blot filters containing bacteriophage λ DNA digested with *Hin*dIII; rat α-tubulin DNA (pILaT1) (22) digested with *Pst*I; mouse α2(I) procollagen DNA (pAZ1002) (25) digested with *Xho*I; and two regions of the mouse α1(I) procollagen gene, p13-3, a subcloned *Bgl*II-*Hpa*I fragment which extends from 5′ of the promoter to the first intron (11), digested with *Eco*RI and *Pst*I, and pTZ19-3.5, which contains sequences from an *Eco*RI site about 10 kb downstream of the transcriptional start site to an *Xba*I site 3.5 kb further downstream (55), digested with *Eco*RI, *Xba*I, and *Bam*HI. After hybridization for 48 h at 65°C in 20 mM Tris (pH 7.4)–20 mM EDTA–0.6 M NaCl–0.4% SDS, the filters were washed twice in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 65°C, once in 2× SSC–2.5 μg of RNase A per ml at 37°C, and twice in 2× SSC–0.2% SDS at 65°C.

Injection of plasmids into *Xenopus* oocytes and analysis of RNA by RNase protection. Collagen plasmids were microinjected into *Xenopus laevis* oocytes, and RNAs were analyzed as previously described (7).

RESULTS

Introduction of wild-type and mutant collagen genes into fibroblast cell lines. The mouse α1(I) procollagen gene, disrupted by the Mov13 provirus, was cloned into *Escherichia coli*, thereby removing the provirus-induced methylation pattern. Plasmids containing either the Mov13 collagen gene (pWTC-Mov and pWTC-Mov-N) or the collagen gene in its wild-type form (pWTC) (Fig. 1) were introduced into

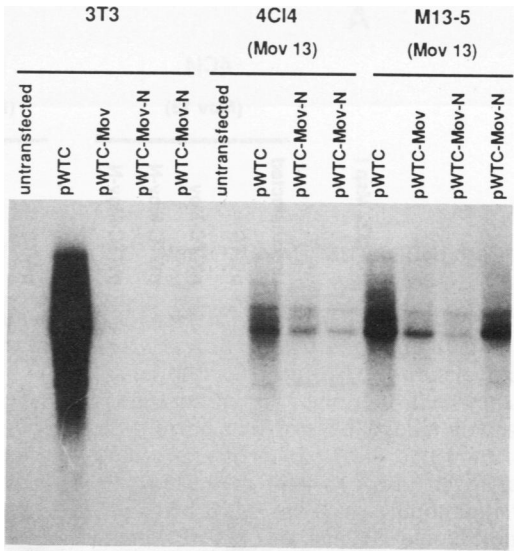

FIG. 2. Estimation of the relative copy number of transfected procollagen plasmids. Genomic DNAs (10 μg) prepared from the nuclei of 3T3, 4Cl4, and M13-5 cells transfected with the indicated α1(I) procollagen plasmids were digested with *Pst*I and analyzed by Southern blot analysis. The probe used was derived from pUC19 and specifically detects sequences from the transfected procollagen plasmids.

NIH 3T3 cells and two different Mov13 homozygous cell lines, M13-5 and 4Cl4, by cotransfection with the selectable plasmid pSV2-neo. Consistently, the recovery of G418-resistant colonies upon transfection of MoMLV-containing plasmids was significantly lower than when plasmids containing no MoMLV provirus were transfected (data not shown). Resistant colonies were pooled and grown as populations. After several passages, a portion of the cells was harvested and used to prepare total cytoplasmic RNA and genomic DNA, and the remainder was frozen for later analysis.

To determine whether the number of transfected collagen plasmids integrated into these G418-resistant populations was equivalent, Southern blot analyses were performed. Equal amounts of *Pst*I-digested genomic DNA were probed with a DNA fragment derived from the lactose operon sequences of pUC19. This fragment specifically detects the transfected collagen plasmid sequences without hybridizing to the transfected pSV2-neo. Figure 2 shows that 3T3 cells transfected with the wild-type gene carried a large number of pWTC plasmids, whereas few or no plasmids were detected in cells transfected with the Mov13 mutant gene pWTC-Mov or pWTC-Mov-N. The only difference between these constructs is the absence or presence of the provirus. This result, together with the reduced number of G418-resistant colonies obtained in pWTC-Mov transfections, suggests that proviral DNA sequences have a negative effect on the survival or growth of the transfected cells. Since transfection results in the introduction of a large number of provirus-containing plasmids, this negative effect may be due to the overexpression of MoMLV proteins. When the same constructs were introduced into homozygous Mov13 cells, a two- to fivefold excess of pWTC copies over pWTC-Mov was observed, in contrast to 3T3 cells, in which a much larger difference was observed. Mov13 cells express MoMLV viral proteins due to their superinfection with

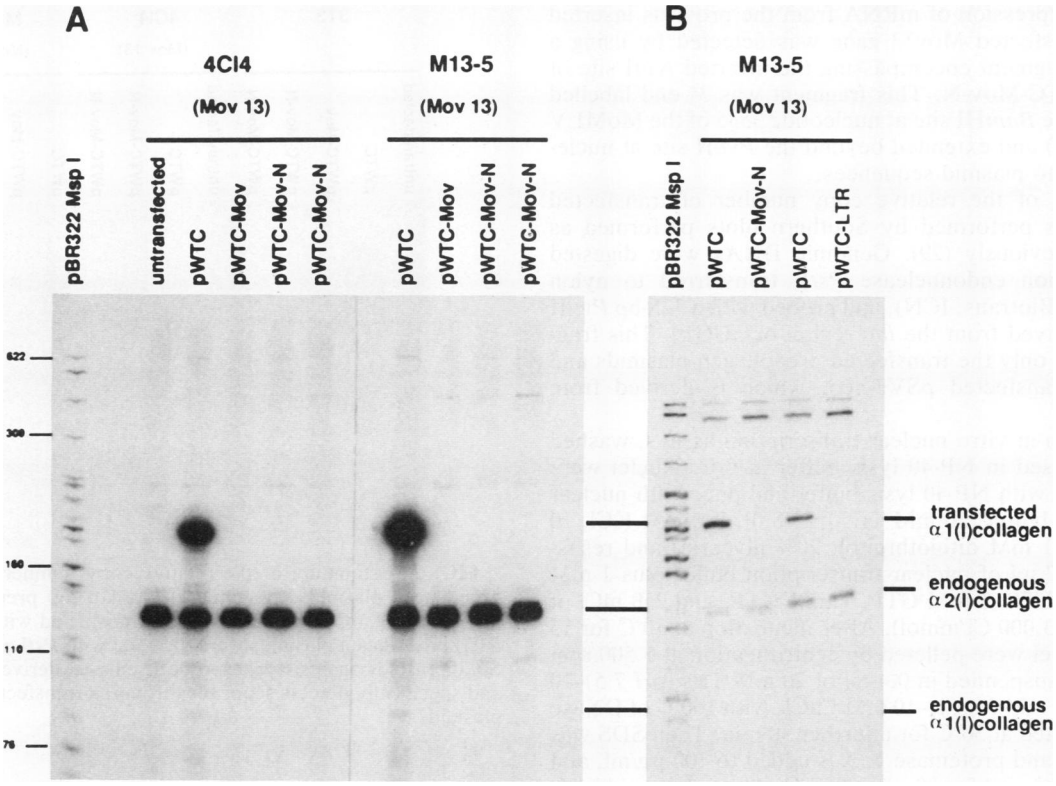

FIG. 3. S1 nuclease analysis of procollagen mRNA expression from transfected plasmids. Total cytoplasmic RNA (10 μg) from 4Cl4 and M13-5 cells transfected with the indicated procollagen plasmids were hybridized with two end-labelled DNA probes for 16 h at 52°C. The α1(I) procollagen gene probe yields a 196-base protected fragment when hybridized to RNA from the transfected plasmid and a 95-base fragment when hybridized to RNA derived from the endogenous gene. The α2(I) procollagen probe yields a 122-base protected fragment.

infectious virions during their prolonged culture. This is consistent with the notion that the prior expression of viral proteins may select for cells that can tolerate MoMLV provirus expression.

Expression of the collagen and proviral genes. To determine whether the cloned Mov13 collagen gene is expressed in transfected cells, the level of collagen mRNA was measured by S1 nuclease analysis (Fig. 3A). An end-labelled DNA probe capable of distinguishing between RNAs derived from the endogenous and transfected genes was used [α1(I) S1 probe; Fig. 1]. This probe spans the site of a linker insertion present within the 5′ untranslated region of all collagen plasmids used. The level of α2(I) procollagen RNA which was determined as an internal control was equal in all RNA samples analyzed. Expression of α1(I) procollagen RNA, however, was detected only in cells that had been transfected with the wild-type collagen gene. This result indicated that demethylation of the Mov13 mutant allele did not lead to the expression of collagen mRNA. Although there was a two- to fivefold difference in the number of the wild-type and Mov13 collagen plasmids integrated into the transfected cells, this difference is clearly not sufficient to account for the difference in mRNA levels.

In Mov13 animals, the provirus is also transcriptionally inactive at early stages of development (14). We therefore studied whether the provirus was expressed after transfection of the cloned Mov13 allele into fibroblasts. A *Not*I linker was introduced into the *env* gene of the Mov13 clone (pWTC-Mov-N) to distinguish transcription of the transfected Mov13 allele from that of the MoMLV proviral copies carried in the Mov13 cell lines. Using an end-labelled DNA probe derived from the altered provirus (MoMLV-*Not*I S1 probe; Fig. 1), an S1 nuclease analysis was performed with RNA derived from untransfected M13-5 cells or M13-5 cells transfected with the Mov13 gene (Fig. 4). In each lane a strong band of 170 nucleotides was detected, which corresponds to the signal expected for the high level of expression from the proviruses carried in the parental cell line. In addition, a band of 285 nucleotides corresponding to transcripts from pWTC-Mov-N was detected only with RNA derived from cells transfected with the Mov13 constructs. The presence of this RNA demonstrates that the provirus was actively transcribed from the transfected Mov13 mutant allele.

Initiation of transcription. Because the Mov13 provirus insertion disrupts the first intron of the α1(I) procollagen gene, the blockage of expression could also be due to disruption of the splicing and posttranscriptional processing of the collagen mRNA. To test this, the level of transcription across the transfected collagen genes was measured in isolated nuclei. Transfected 4Cl4 cells, frozen at the time RNA and DNA were isolated, were replated, passaged once, and used to prepare nuclei. ^{32}P-labelled RNA was synthesized in vitro and hybridized to Southern blots containing collagen and control DNAs (Fig. 5). As previously observed, nuclei from untransfected 4Cl4 cells actively transcribed the α-tubulin gene and the α2(I) collagen gene but did not transcribe sequences derived from the 5′ region of the α1(I) collagen gene (12). The cells did, however, transcribe sequences from a downstream region of the collagen gene as

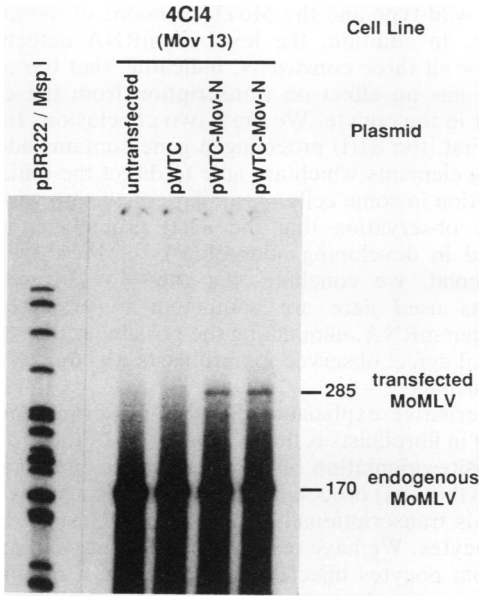

FIG. 4. S1 nuclease analysis of MoMLV mRNA expression from transfected plasmids. Total cytoplasmic RNA (10 μg) from 4Cl4 and M13-5 cells transfected with the indicated procollagen plasmids was hybridized with an end-labelled DNA probe spanning the position of the inserted *Not*I site in the MoMLV *env* gene for 16 h at 55°C. RNA derived from the endogenous MoMLV proviruses and the transfected proviruses without the inserted *Not*I site protects a 170-base fragment, while RNA derived from the transfected proviruses containing the inserted *Not*I site protects a fragment of 285 bases.

has been observed previously, likely as a result of transcription originating at a cryptic promoter in this region of the genome (12). Nuclei derived from pWTC-transfected cells transcribed in addition the 5′ region of the transfected collagen gene and showed an increased level of downstream transcription. Transcripts from the transfected vector sequences were also observed, likely as a result of tandem integration of the cloned DNA and transcription initiating in the collagen promoter. In contrast, nuclei isolated from 4Cl4 cells transfected with two different Mov13 mutant collagen constructs failed to transcribe the 5′ α1(I) collagen region but did transcribe the transfected vector sequences and the more 3′ α1(I) collagen sequences. The level of transcription from the vector sequences in these cells served as a good control for the reduced copy number of the transfected collagen genes and indicated that a 5′ collagen signal would have been visible if the Mov13 mutant allele was transcribed normally. This result indicates that the initiation of transcription from the transfected Mov13 mutant genes was defective. It demonstrates that demethylation of the Mov13 mutant allele is not sufficient to restore collagen transcription and strongly suggests that the transcriptional block is due to displacement or disruption of DNA elements required for α1(I) procollagen transcription.

A single LTR causes inhibition of collagen transcription. Retrotransposon-induced mutations have been reported to revert after excision of the provirus as a consequence of recombination between the LTRs, which leaves a single LTR behind (2, 3, 8, 49). To determine whether a similar recombination could revert the Mov13 mutation, we have constructed an α1(I) procollagen gene disrupted by a single LTR in place of the Moloney provirus. This plasmid, pWTC-LTR, was cotransfected into M13-5 cells along with pSV2-

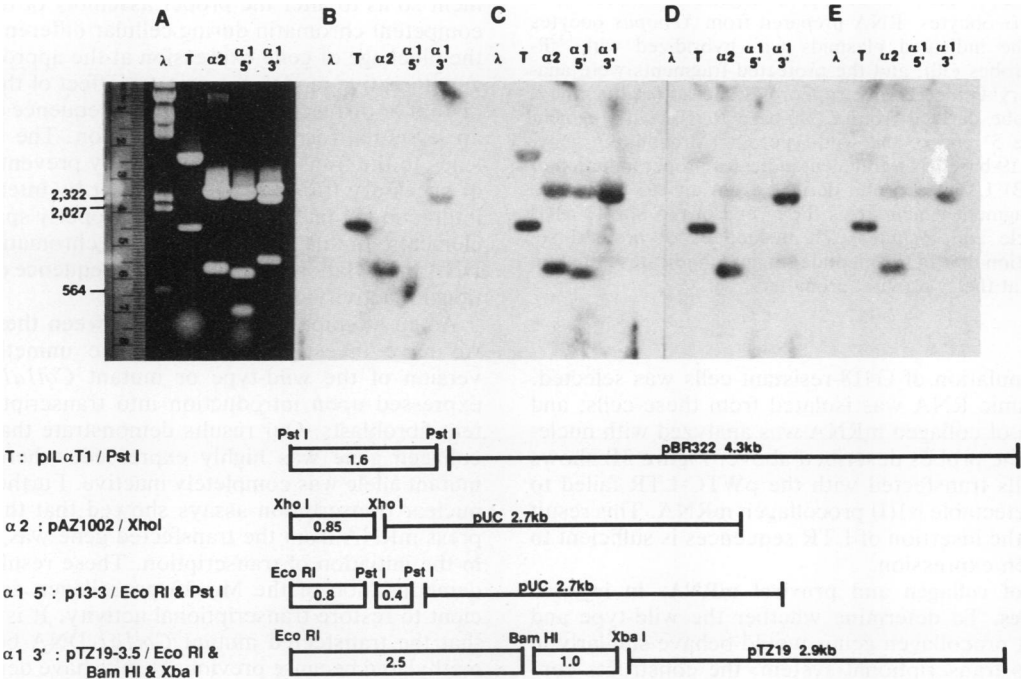

FIG. 5. Nuclear transcription assay of 4Cl4 cells transfected with procollagen plasmids. Nuclei isolated from untransfected 4Cl4 cells (B) or 4Cl4 cells transfected with pWTC (C), pWTC-Mov (D), and pWTC-Mov-N (E) were labelled in vitro with [^{32}P]UTP. RNA isolated from these nuclei was hybridized to Southern blot filters containing bacteriophage λ DNA (λ), rat α-tubulin DNA (T), mouse α2(I) procollagen DNA (α2), and two regions of the mouse α1(I) procollagen gene (α1 5′ and α1 3′). An ethidium bromide-stained gel used for this blot is shown in panel A.

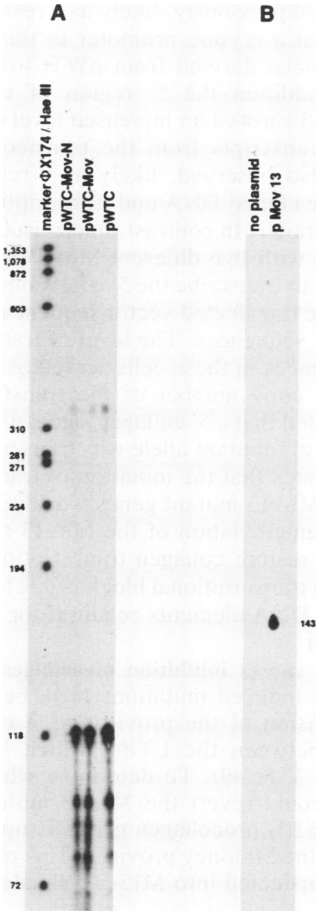

FIG. 6. RNase protection of α1(I) procollagen RNA expressed in injected *X. laevis* oocytes. RNA prepared from *Xenopus* oocytes injected with the indicated plasmids was hybridized with ^{32}P-labelled RNA probes (30), and the protected fragments were analyzed by polyacrylamide gel electrophoresis and autoradiography. (A) Using a probe derived from a 340-base *Bgl*II-*Xba*I fragment which spans the 5′ end of the wild-type α1(I) procollagen gene, protection of a 119-base RNA fragment indicates proper initiation of transcription. (B) Using a probe derived from an 803-nucleotide *Bgl*II-*Eco*RV fragment which spans the 5′ end of the Mov13 α1(I) procollagen allele and includes 371 nucleotides of proviral sequences, protection of a 143-nucleotide fragment indicates initiation of transcription at the 5′ proviral promoter.

neo, and a population of G418-resistant cells was selected. Total cytoplasmic RNA was isolated from these cells, and the expression of collagen mRNA was analyzed with nuclease S1, using the probes described above. Figure 3B shows that M13-5 cells transfected with the pWTC-LTR failed to express any detectable α1(I) procollagen mRNA. This result indicates that the insertion of LTR sequences is sufficient to disrupt collagen expression.

Expression of collagen and proviral mRNAs in injected *Xenopus* oocytes. To determine whether the wild-type and Mov13 mutant procollagen genes would behave similarly in a heterologous transcriptional system, the constructs containing these genes were injected into *Xenopus* oocytes. After 24 h, RNA was isolated and expression of procollagen and proviral mRNAs was measured with an RNase protection assay (Fig. 6A). By using a probe which spans the 5′ end of the procollagen gene, specific initiation was detected from both the wild-type and the Mov13 versions of the procollagen gene. In addition, the level of mRNA detected was similar for all three constructs, indicating that the provirus insertion has no effect on transcription from the collagen promoter in the oocyte. We draw two conclusions from this result. First, the α1(I) procollagen gene contains additional *cis*-acting elements which are able to direct the initiation of transcription in some cells. A similar conclusion was drawn from the observation that the α1(I) procollagen gene is expressed in developing odontoblasts of Mov13 embryos (21). Second, we conclude that the Mov13 procollagen constructs used here are competent to produce stable procollagen mRNA, eliminating the possibility that the transcriptional defect observed in fibroblasts was due to a trivial cloning artifact.

An alternative explanation for the transcriptional defect observed in fibroblasts is that transcription of the provirus in the opposite orientation results in suppression of transcription from the α1(I) procollagen promoter because the Mov13 provirus is transcriptionally active in fibroblasts but not *X. laevis* oocytes. We have tested this possibility and analyzed RNA from oocytes injected with pMov13, a plasmid containing a truncated version of the Mov13 α1(I) procollagen allele (7), for transcripts initiated in the proviral 5′ LTR. Figure 6B shows that proviral transcripts are readily detectable, ruling out the possibility that proviral transcription per se is sufficient to inactivate α1(I) procollagen transcription.

DISCUSSION

Two types of models can be envisaged to explain the transcriptional defect induced by the Mov13 mutation. In one model, the primary effect of the provirus insertion is either to induce de novo methylation of the *Col1a1* gene or to disrupt the local chromatin structure during early development so as to alter the proper assembly of transcriptionally competent chromatin during cellular differentiation, causing the blockage of gene expression at the appropriate stage. In an alternative model, the primary effect of the provirus is to disrupt or displace *cis*-acting DNA sequence elements which are essential for *Col1a1* transcription. The disruption then leads to the transcriptional defect by preventing the binding of necessary transcription factors or by interfering with the interaction of factors bound to improperly spaced regulatory elements. In this model, the altered chromatin structure and DNA methylation might be the consequence of the transcriptional inactivity of the gene.

As an attempt to distinguish between these alternatives, we have investigated whether the unmethylated cloned version of the wild-type or mutant *Col1a1* genes can be expressed upon introduction into transcriptionally competent fibroblasts. Our results demonstrate that the wild-type collagen gene was highly expressed, whereas the Mov13 mutant allele was completely inactive. Furthermore, in vitro nuclear transcription assays showed that the failure to express mRNA from the transfected gene was due to a defect in the initiation of transcription. These results indicate that demethylation of the Mov13 procollagen gene is not sufficient to restore transcriptional activity. It is highly unlikely that the transfected mutant *Col1a1* DNA became de novo methylated because previous results have demonstrated that proviral DNA introduced into fibroblasts remains unmethylated and is expressed, in contrast to transfection into embryonic carcinoma cells (48). Similarly, Chan et al. (7) found that treatment of Mov13 fibroblasts with 5-azacytidine resulted in the demethylation of the *Col1a1* gene but failed to

induce its transcription. Our results, therefore, support a model in which the provirus displaces or disrupts essential *cis*-acting DNA sequence elements.

Possible candidate sequence elements have been identified within the 3′ half of the first intron. Rippe et al. (35) found that the presence of first-intron sequences from the mouse α1(I) procollagen gene exerts a negative effect on the expression of choramphenicol acetyltransferase transcribed from the procollagen promoter. The first intron of the human α1(I) procollagen gene has been analyzed in much more detail (4, 5, 26, 37), and several positive and negative elements have been identified. One of these sequences, termed A275 (5), has been found to inhibit transcription when inverted or placed 5′ of the human α1(I) procollagen promoter. An element within this sequence has been proposed to regulate collagen transcription by promoter-intron interactions mediated by DNA-binding proteins. A second element, which has been narrowed down to a 29-bp sequence called S1 (26), exerts a positive, orientation-dependent effect on transcription. This element binds nuclear proteins and contains an AP1 consensus binding site which is essential for its stimulatory activity. Although the first introns of the human and mouse α1(I) procollagen genes share little sequence homology, both the A275 and S1 elements are highly conserved. A 29-bp portion of the A275 element and a 32-bp sequence containing the S1 element share 90 and 94% identity, respectively, with sequences in the same relative positions of the first intron of the mouse gene. In addition to these sequences, one other region of the introns shares a high degree of sequence identity; a functional role for this sequence, however, has not been reported. All of these conserved sequence elements lie downstream from the provirus insertion site and would be displaced by about 9 kb in the mutant gene. Alternatively, direct disruption of a crucial *cis*-acting control element by proviral insertion instead of displacement may be the cause of the mutation. We note that the sequences flanking the proviral insertion at the 5′ end of the intron are highly homologous to the NFκB binding site (8 of 10 nucleotide identities) (23). While a functional role of these sequences in collagen transcription has not been established as yet, the observation that a single LTR has the same effect on collagen transcription as the full-length provirus is consistent with disruption of a control element rather than displacement of far-downstream elements. Our results are most consistent with the notion that the transcriptional defect in the Mov13 allele is due to disruption or displacement of essential sequence elements within the first intron. Site-directed mutagenesis of these elements will be necessary to define their function in the control of collagen gene expression.

Our observation that the mutated Mov13 procollagen gene is transcribed normally when injected into *Xenopus* oocytes indicates that it contains additional *cis*-acting regulatory elements which are capable of directing the initiation of transcription in a cell type other than fibroblasts. This result is consistent with the observations of Kratochwil et al., who demonstrated that the Mov13 procollagen gene is expressed in odontoblasts and in some osteoblasts (21). This finding suggested that these tissues, which exclusively and abundantly express type I collagen, do so under the control of *cis*-acting regulatory elements distinct from those required for transcription in fibroblasts. The most likely position of these alternative tissue-specific elements would be upstream of the provirus and within the 5′ flanking region and would therefore not be disrupted by the viral insertion. Preliminary results indeed are consistent with such a notion (9).

Finally, it might be argued that transcription from the proviral promoter may interfere with transcription from the opposing procollagen promoter, causing the Mov13 mutant phenotype. If this were true, one would expect that the provirus would be expressed in fibroblasts transfected with the mutant procollagen gene but not in oocytes injected with it. To the contrary, our results show that the provirus is actively transcribed in oocytes injected with pMov13, indicating that transcription from the opposing proviral promoter is not sufficient to prevent *Coll a1* transcription. In addition, nuclear transcription assays reported here and previously (12) failed to detect transcription in Mov13 cells across the first exon of the procollagen gene in either direction. This indicates either that the level of transcription from the proviral promoter is much lower than the normal level of transcription from the procollagen promoter or that proviral transcription is efficiently terminated. In either case, there does not appear to be a significant level of transcription opposing the procollagen promoter in the transfected fibroblasts.

Excision of a provirus by homologous recombination leaves a single LTR at the site of the proviral insertion and can result in the reversion of the mutant phenotype. This has been observed for the two recessive mutations dilute (8) and hairless (49). In contrast to these spontaneous mutations, the Mov13 mutant phenotype is not reverted by excision of the provirus because a single LTR at the insertion site blocks collagen expression as efficiently as the complete provirus. While the molecular mechanisms of the dilute and hairless mutations are not understood as yet, our results indicate that retroviruses mutate cellular genes by different mechanisms.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants PO1 HL41484 and (OIG) 5R35-CA44339 and by grant CD-458 from the American Cancer Society. D.D.B. was supported by American Cancer Society postdoctoral fellowship grant PF-3005.

REFERENCES

1. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
2. **Bingham, P. M., and Z. Zachar.** 1989. Retrotransposons and the FB transposon from *Drosophila melanogaster*, p. 485–502. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
3. **Boeke, J. D.** 1989. Transposable elements in *Saccharomyces cerevisiae*, p. 335–374. *In* D. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
4. **Bornstein, P., and J. McKay.** 1988. The first intron of the alpha 1(I) collagen gene contains several transcriptional regulatory elements. J. Biol. Chem. **263**:1603–1606.
5. **Bornstein, P., J. McKay, D. J. Liska, S. Apone, and S. Devarayalu.** 1988. Interactions between the promoter and first intron are involved in transcriptional control of alpha 1(I) collagen gene expression. Mol. Cell. Biol. **8**:4851–4857.
6. **Breindl, M., K. Harbers, and R. Jaenisch.** 1984. Retrovirus-induced lethal mutation in collagen I gene of mice is associated with altered chromatin structure. Cell **38**:9–16.
7. **Chan, H., S. Hartung, and M. Breindl.** 1991. Retrovirus-induced interference with collagen I gene expression in Mov13 fibroblasts is maintained in absence of DNA methylation. Mol. Cell. Biol. **11**:47–54.
8. **Copeland, N. G., K. W. Hutchison, and N. A. Jenkins.** 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. Cell **33**:379–387.
9. **Faessler, R., and R. Jaenisch.** Unpublished data.

10. **Gautsch, J., and M. Wilson.** 1983. Restriction of Moloney murine leukemia virus growth in teratocarcinoma: involvement of factors other than DNA methylation. Cold Spring Harbor Conf. Cell Prolifer. **10**:363–378.
11. **Harbers, K., M. Kuehn, H. Delius, and R. Jaenisch.** 1984. Insertion of retrovirus into the first intron of the α1(I) collagen gene leads to embryonic lethal mutation in mice. Proc. Natl. Acad. Sci. USA **81**:1504–1508.
12. **Hartung, S., R. Jaenisch, and M. Breindl.** 1986. Retrovirus insertion inactivates mouse α1(I) collagen gene by blocking initiation of transcription. Nature (London) **320**:365–367.
13. **Jaenisch, R., K. Harbers, A. Schnieke, J. Löhler, I. Chumakov, D. Jähner, D. Grotkopp, and E. Hoffman.** 1983. Germline integration of Moloney murine leukemia virus at the Mov 13 locus leads to recessive lethal mutation and early embryonic death. Cell **32**:209–216.
14. **Jaenisch, R., D. Jähner, P. Nobis, I. Simon, J. Löhler, K. Harbers, and D. Grotkopp.** 1981. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell **24**:519–529.
15. **Jähner, D., and R. Jaenisch.** 1980. Integration of Moloney leukemia virus into the germ line of mice: correlation between genotype and virus activation. Nature (London) **287**:456–458.
16. **Jähner, D., and R. Jaenisch.** 1984. DNA methylation in early mammalian development, p. 189–219. *In* A. Razin, H. Cedar, and A. D. Riggs (ed.), DNA methylation. Springer-Verlag, New York.
17. **Jähner, D., and R. Jaenisch.** 1985. Chromosomal position and specific demethylation in enhancer sequences of germ line-transmitted retroviral genomes during mouse development. Mol. Cell. Biol. **5**:2212–2220.
18. **Jähner, D., and R. Jaenisch.** 1985. Retrovirus induced *de novo* methylation of flanking host sequences correlates with gene inactivity. Nature (London) **315**:594–597.
19. **Jähner, D., H. Stuhlmann, C. L. Stewart, K. Harbers, J. Löhler, I. Simon, and R. Jaenisch.** 1982. *De novo* methylation and expression of retroviral genomes during mouse embryogenesis. Nature (London) **298**:623–628.
20. **Jenkins, N. A., N. G. Copeland, B. A. Taylor, and B. K. Lee.** 1981. Dilute (d) coat color mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature (London) **293**:370–374.
21. **Kratochwil, K., K. von der Mark, E. J. Kollar, R. Jaenisch, K. Mooslehner, M. Schwarz, K. Haase, I. Gmachl, and K. Harbers.** 1989. Retrovirus-induced insertional mutation in Mov 13 mice affects collagen I expression in a tissue-specific manner. Cell **57**:807–816.
22. **Lemischka, I., and P. A. Sharp.** 1982. The sequences of an expressed rat alpha-tubulin gene and a pseudogene with an inserted repetitive element. Nature (London) **300**:330–335.
23. **Lenardo, M. J., and D. Baltimore.** 1989. NF-kappa B: a pleiotropic mediator of inducible and tissue-specific gene control. Cell **58**:227–229.
24. **Levis, R., K. O'Hare, and G. M. Rubin.** 1984. Effects of transposable element insertions on RNA encoded by the white gene of *Drosophila*. Cell **36**:471–481.
25. **Liau, G., Y. Yamada, and B. de Crombrugghe.** 1985. Coordinate regulation of the levels of type III and type I collagen mRNA in most but not all mouse fibroblasts. J. Biol. Chem. **260**:531–536.
26. **Liska, D. J., J. L. Slack, and P. Bornstein.** 1990. A highly conserved intronic sequence is involved in transcriptional regulation of the α1(I) collagen gene. Cell Regul. **1**:487–498.
27. **Loh, T. P., L. L. Sievert, and R. W. Scott.** 1987. Proviral sequences that restrict expression in mouse embryonal carcinoma cells. Mol. Cell. Biol. **7**:3775–3784.
28. **Löhler, J., R. Timpl, and R. Jaenisch.** 1984. Embryonic lethal mutation in mouse collagen I gene causes rupture of blood vessels and is associated with erythropoietic and mesenchymal cell death. Cell **38**:597–607.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
31. **Modelell, J., W. Bender, and M. Meselson.** 1983. *Drosophila melanogaster* mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3kb mobile element. Proc. Natl. Acad. Sci. USA **80**:1678–1682.
32. **Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara.** 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell **32**:1105–1113.
33. **Parkhurst, S. M., and V. G. Corces.** 1985. Forked, gypsys and suppressors in *Drosophila*. Cell **41**:429–437.
34. **Paro, R., M. L. Goldberg, and W. J. Gehring.** 1983. Molecular analysis of large transposable elements carrying the white locus of *Drosophila melanogaster*. EMBO J. **2**:853–860.
35. **Rippe, R. A., S. Lorenzen, D. A. Brenner, and M. Breindl.** 1989. Regulatory elements in the 5′-flanking region of the first intron contribute to transcriptional control of the mouse alpha 1 type I collagen gene. Mol. Cell. Biol. **9**:2224–2227.
36. **Rossi, P., and B. de Combrugghe.** 1987. Identification of a cell-specific transcriptional enhancer in the first intron of the mouse a2 (type I) collagen gene. Proc. Natl. Acad. Sci. USA **84**:5590–5594.
37. **Rossouw, C. M. S., W. P. Vergeer, S. J. du Plooy, M. P. Bernard, F. Ramirez, and W. J. de Wet.** 1987. DNA sequences in the first intron of the human pro-α1(I) collagen gene enhance transcription. J. Biol. Chem. **262**:15151–15157.
38. **Schmidt, A., P. Rossi, and B. de Combrugghe.** 1986. Transcriptional control of the mouse α2 (I) collagen gene: functional deletion analysis of the promoter and evidence for cell specific expression. Mol. Cell. Biol. **6**:347–354.
39. **Schnieke, A., M. Dziadek, J. Bateman, T. Mascara, K. Harbers, R. Gelinas, and R. Jaenisch.** 1987. Introduction of the human pro-alpha 1 (I) collagen gene into pro-alpha 1 (I)-deficient Mov 13 mouse cells leads to formation of functional mouse-human hybrid type I collagen. Proc. Natl. Acad. Sci. USA **84**:764–768.
40. **Schnieke, A., K. Harbers, and R. Jaenisch.** 1983. Embryonic lethal mutation in mice induced by retrovirus insertion into the α1(I) collagen gene. Nature (London) **304**:315–320.
41. **Sererack, P. K., M. C. Strobel, D. J. Corrow, N. A. Jenkins, and N. G. Copeland.** 1988. Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. Proc. Natl. Acad. Sci. USA **85**:189–192.
42. **Shinnick, T., R. Lerner, and J. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
43. **Simchen, G., F. Winston, C. A. Styles, and G. R. Fink.** 1984. Ty-mediated gene expression of the LYS2 and HIS4 genes of Saccharomyces cerevisiae is controlled by the same SPT genes. Proc. Natl. Acad. Sci. USA **81**:2431–2434.
44. **Soriano, P., T. Gridley, and R. Jaenisch.** 1987. Retroviruses and Insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. Genes Dev. **1**:366–375.
45. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.
46. **Spence, S. E., D. J. Gilbert, D. A. Swing, N. G. Copeland, and N. A. Jenkins.** 1989. Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. Mol. Cell. Biol. **9**:177–184.
47. **Stacey, A., J. Bateman, T. Choi, T. Mascara, W. Cole, and R. Jaenisch.** 1988. Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered mutant pro-α1(I) collagen gene. Nature (London) **332**:131–136.
48. **Stewart, C., H. Stuhlman, D. Jähner, and R. Jaenisch.** 1982. *De novo* methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA **79**:4098–4192.
49. **Stoye, J. P., S. Fenner, G. E. Greenoak, C. Moran, and J. M. Coffin.** 1988. Role of endogenous retroviruses as mutagens: the

hairless mutation of mice. Cell **54**:383–391.

50. **Stuhlmann, H., D. Jahner, and R. Jaenisch.** 1981. Infectivity and methylation of retroviral genomes is correlated with expression in the animal. Cell **2**:221–232.
51. **van der Eb, A. J., and F. L. Graham.** 1980. Assay of transforming activity of tumor virus DNA. Methods Enzymol. **65**:826–839.
52. **Varmus, H., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
53. **Weiher, H., T. Noda, D. A. Gray, A. H. Sharpe, and R. Jaenisch.** 1990. Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. Cell **62**:425–434.
54. **Winston, F., D. T. Chaleff, B. Valent, and G. R. Fink.** 1984. Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. Genetics **107**:179–197.
55. **Wu, H., and R. Jaenisch.** Unpublished data.
56. **Wu, H., J. Bateman, A. Schnieke, A. Sharpe, D. Barker, T. Mascara, D. Eyre, R. Bruns, P. Krimpenfort, A. Berns, and R. Jaenisch.** 1990. Human-mouse interspecies collagen I heterotrimer is functional during embryonic development of Mov 13 mutant mouse embryos. Mol. Cell. Biol. **10**:1452–1460.
57. **Zachar, Z., D. Davison, D. Garza, and P. M. Bingham.** 1985. A detailed developmental and structural study of the transcriptional effects of insertion of the copia transposon into the white locus of *Drosophila melanogaster*. Genetics **111**:495–515.