MOLECULAR AND CELLULAR BIOLOGY, Oct. 1983, p. 1834–1845
0270-7306/83/101834-12$02.00/0

Vol. 3, No. 10

Identification of Transcriptional Elements Within the Long Terminal Repeat of Rous Sarcoma Virus

GREGORY M. GILMARTIN† AND J. THOMAS PARSONS*

Department of Microbiology, University of Virginia Medical School, Charlottesville, Virginia 22908

Received 15 April 1983/Accepted 2 August 1983

Transcriptional regulatory elements within the Rous sarcoma virus long terminal repeat were examined by the construction of a series of deletions and small insertions within the U3 region of the long terminal repeat. The analysis of these mutations in chicken embryo cells and COS cells permitted the identification of important transcriptional regulatory elements. Sequences within the region 31 to 18 base pairs upstream of the RNA cap site (−31 to −18), encompassing a TATA box-like sequence, function in the selection of the correct site of transcription initiation and, in addition, augment the efficiency of transcription. These sequences are essential for virus replication. Sequences within the region −79 to −59, overlapping a CAAT box-like sequence, are not required for virus replication and have no obvious effect on viral RNA transcription in the presence of an intact TATA box. However, in mutants lacking a functional TATA sequence, mutations in this region serve to decrease the efficiency of correct transcriptional initiation events.

Transcription of Rous sarcoma virus (RSV) proviral DNA by RNA polymerase II in chicken embryo cells is a highly efficient event. Viral RNA transcribed from between 1 and 20 copies of proviral DNA may comprise as much as 20% of the polyadenylated RNA of infected cells (45). The DNA sequences responsible for the promotion of RSV transcription are presumed to reside primarily within the U3 region of the long terminal repeat (LTR) that flanks the provirus. In vitro transcription experiments have demonstrated the ability of this region to promote the specific initiation of RSV transcripts by RNA polymerase II (47). The RSV LTR has also been shown to function as a strong promoter upon transfection of a variety of eucaryotic cells (18). DNA sequence analysis of this region revealed the presence of putative eucaryotic RNA polymerase II transcriptional regulatory sequence elements. A Goldberg-Hogness sequence, or TATA box, of the form 5′-TATTTAA-3′ resides 31 to 23 base pairs (bp) upstream from the viral RNA cap site, whereas a CAAT box-like sequence, 5′-GGTCTAAGA-3′, is found 81 to 73 bp upstream. Such sequence elements have been demonstrated in several systems to function in the modulation of RNA polymerase II transcription (3, 10, 20, 22, 33, 35). The presence of nucleotide sequences within the LTR that may serve to enhance or activate RNA transcription has been implicated by the identification of an enhancer element in the LTR of Moloney murine sarcoma virus (25). Although avian retrovirus LTRs lack the directly repeated sequences characteristic of some enhancer elements (2, 25), evidence indicates that they may exhibit such a function in the transcriptional activation of cellular oncogenes (37).

We used site-directed mutagenesis techniques to investigate the function of the U3 region in the modulation of RSV transcription. The analysis of a series of deletions and small insertions at a restriction site 53 bp upstream from the viral RNA cap site enabled us to identify DNA sequences involved in the regulation of viral transcription in vivo. We conclude that the RSV TATA box functions primarily as a selector element, specifying the correct site for the initiation of RSV transcription, in addition to augmenting transcription efficiency. The −79 to −59 region, encompassing most of the CAAT box, is dispensable in the presence of an intact TATA box, yet appears to serve a quantitative function in its absence.

MATERIALS AND METHODS

Cells, viruses, and plasmids. Primary cultures of chicken embryo cells were prepared from gs-negative, chf-negative embryos (SPAFAS, Norwich, Conn.) and maintained as previously described (36). Endogenous virus-negative (ev⁻) chicken cells were provided by R. Smith, Colorado State University. The recombinant plasmid pSL103, containing a biologically active

† Present address: Institut für Molekularbiologie II der Universität Zürich, Hönggerberg, CH-8093 Zurich, Switzerland.

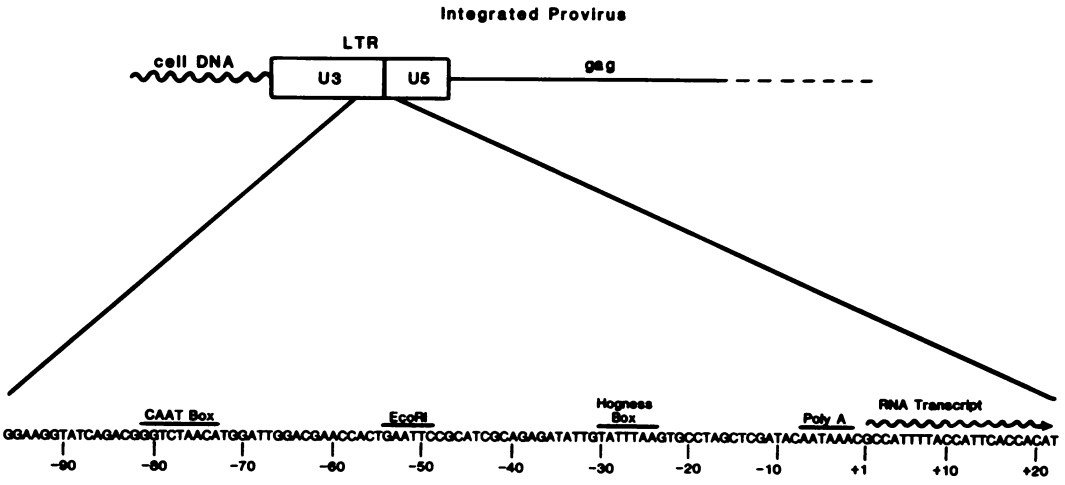

FIG. 1. Structure of the putative promoter region of the RSV provirus. The nucleotide sequence flanking the RNA cap site, denoted +1, is illustrated. The locations of transcriptional consensus sequences are indicated as well as the *Eco*RI restriction site utilized for the construction of mutations in this region. The region 605 bp 5′ and 157 bp 3′ to the RNA cap site, as well as the region between the 3′ end of *env* and the 5′ end of *src* of Prague A RSV, have been sequenced and are virtually identical to the sequence of Prague C RSV as published by Schwartz et al. (40).

Prague A RSV insert, was subcloned from λRPA103 (24) into the *Sal*I site of pBR322. Plasmid pSV010, provided by R. Tjian (University of California, Berkeley), is a pBR322 derivative lacking the "poison sequences" that inhibit replication in animal cells (26) and possessing a 228-bp fragment encompassing the simian virus 40 (SV40) origin of replication. The herpes simplex virus thymidine kinase gene was derived from the plasmid pHSV-106 described by McKnight and Gravis (31).

Restriction endonucleases. Restriction endonucleases were obtained from New England Biolabs, Boston, Mass., and Bethesda Research Laboratories, Gaithersburg, Md., and used according to the specifications of the supplier unless stated otherwise.

Site-directed mutagenesis. Plasmid pSL103 DNA was incubated with *Eco*RI under conditions for partial digestion (50 mM NaCl, 100 mM Tris-hydrochloride, pH 7.5, 3 mM $MgCl_2$, 25 μg of DNA, 2.5 U of *Eco*RI) for 30 min at 37°C, and the reaction was terminated by the addition of EDTA to 15 mM. Linear molecules of 13.7 kilobases (kb) were resolved by agarose gel electrophoresis and eluted. The DNA was digested with exonuclease *Bal* 31 (Bethesda Research Laboratories) in 200 mM NaCl–12 mM $CaCl_2$–12 mM $MgCl_2$–20 mM Tris-hydrochloride (pH 8.0)–1 mM EDTA–0.2 U of *Bal* 31 at 15°C for 45 or 120 s. The reaction was terminated by the addition of EDTA to 50 mM, followed by ethanol precipitation. DNA containing protruding 5′ ends was repaired by incubation with 18 mM deoxynucleotide phosphates–50 mM Tris-hydrochloride (pH 7.8)–5 mM $MgCl_2$–10 mM 2-mercaptoethanol–50 μg of bovine serum albumin per ml–10 U of DNA polymerase I (Bethesda Research Laboratories) for 15 min at 15°C. The DNA was ethanol precipitated, and the residual triphosphates were removed by Sephadex G-100 chromatography. The DNA was ligated with T4 DNA ligase (New England Biolabs) and used to transform *Escherichia coli* HB101 (9). Plasmid DNA was isolated from individual ampicillin-resistant transformants and screened for the absence of the U3 *Eco*RI site by *Eco*RI digestion and agarose gel electrophoresis.

DNA sequence analysis. The nucleotide sequence of the U3 region was determined by the method of Maxam and Gilbert (28). A 918-bp *Taq*I fragment was 5′ end labeled with ^{32}P and subsequently digested with *Bgl*I. An 826-bp fragment was isolated by agarose gel electrophoresis and sequenced, and the products were resolved in 8 and 20% polyacrylamide gels.

Transfection of chicken embryo cells. Plasmid pSL103 DNA and its derivatives were digested with *Sal*I, and the 9.3-kb viral insert was purified by agarose gel electrophoresis. Purified viral DNA was ligated with T4 DNA ligase, calcium precipitated, and applied to chicken embryo cells as previously described (24). Cultures were monitored for cellular transformation up to 21 days. Culture supernatants harvested from cultures which exhibited no cellular transformation after 21 days were tested for the presence of reverse transcriptase activity (39).

Southern blot analysis. High-molecular-weight DNA was isolated from transfected ev^- chicken embryo cells as previously described (6). The DNA was digested with either *Pvu*I or *Eco*RI and electrophoresed in 0.75% agarose gels. DNA was transferred from agarose gels to nitrocellulose filter paper and hybridized with ^{32}P-labeled probe DNA as described by Southern (43). Probes were ^{32}P labeled by nick translation (27) (specific activity, 10^8 cpm/μg) of the 9.3-kb *Sal*I fragment of pSL103 representative of the entire RSV genome.

Quantitation of intracellular viral RNA and virion RNA. RNA levels were quantitated by hybridization of in vivo ^3H-labeled RNA with excess filter-bound viral DNA. Plasmid pSL103 DNA was bound to nitrocellulose filter paper as previously described (15). An equivalent amount of pBR322 DNA bound to nitrocel-

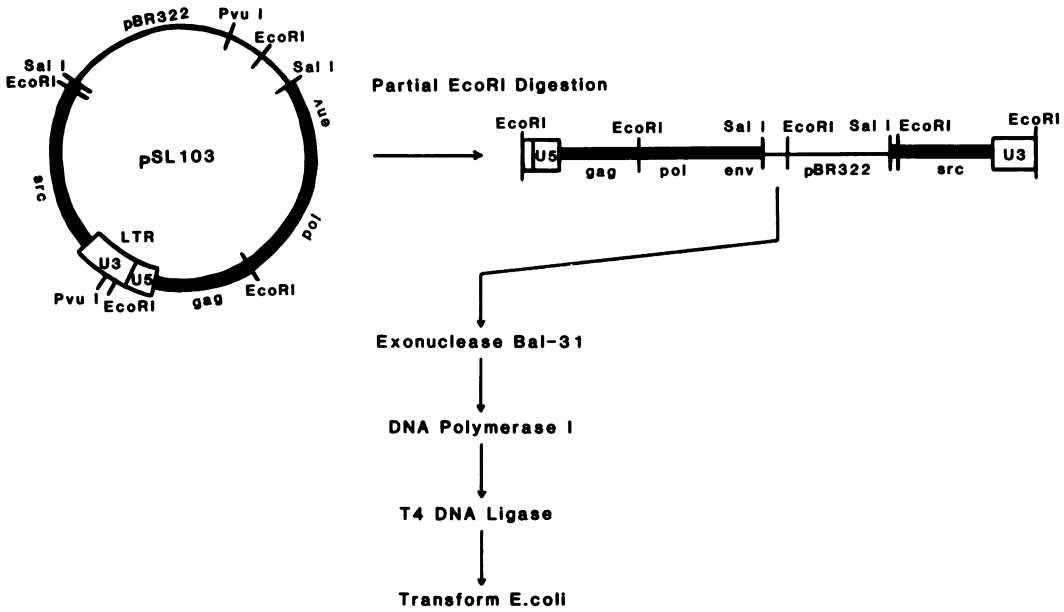

FIG. 2. Construction of mutations in the RSV U3 region. The plasmid pSL103 was linearized by partial *Eco*RI digestion, and the 13.7-kb linear molecules were isolated by agarose gel electrophoresis. The linear molecules were subjected to exonuclease *Bal* 31 treatment to generate terminal deletions, followed by incubation with DNA polymerase I to repair protruding 5′ ends. These molecules were then circularized with T4 DNA ligase and used to transform *E. coli*.

lulose was used as a control for nonspecific hybridization. Chicken embryo cells were labeled with [^{3}H]uridine or $^{32}P_i$, and cellular and virion RNA were isolated as previously described (36). The RNAs were incubated with duplicate sets of nitrocellulose filters containing either viral DNA or pBR322 DNA. To each hybridization was added a small amount of 70S ^{32}P-labeled viral RNA as an internal control. The amount of RNase-resistant counts per minute bound to DNA was then determined as described previously (44).

Nuclease S1 analysis of 5′ termini of RNA transcripts. Cells were lysed in 6 M urea–3 M LiCl–10 mM sodium acetate, pH 5.0–0.1% sodium dodecyl sulfate, and total cellular RNA was prepared as described previously (1). The RSV DNA probes were constructed by ^{32}P 5′ end labeling (28) a 10.3-kb *Bst*EII fragment of pSL103 and digestion with either *Pvu*I or *Eco*RI and isolation of 218- or 157-bp fragments, respectively. The *tk* DNA probe was constructed by ^{32}P 5′ end labeling *Bgl*II-digested pHSV-106 DNA, followed by *Eco*RI digestion and isolation of a 131-bp fragment.

Hybridization of 1 to 10 μg of total cellular RNA and 0.01 to 0.03 pmol of probe DNA (specific activity, 10^6 cpm/pmol) was carried out in 80% formamide–400 mM NaCl–40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.4)–1 mM EDTA in a total volume of 10 μl. Hybridization reactions were incubated at 60°C for 10 min and transferred to 45°C for the RSV probe and 51°C for the *tk* probe. After annealing for 10 to 12 h, hybridizations were added to 110 μl of S1 nuclease buffer containing 250 mM NaCl–30 mM sodium acetate (pH 4.5)–1 mM $ZnSO_2$–20 μg of single-stranded calf thymus DNA per ml–200 U of S1 nuclease per ml (P-L Biochemicals, Milwaukee, Wis.) and incubated for 40 min at 25°C. Nuclease digestion was terminated by extraction with an equal volume of $CHCl_3$-phenol (1:1), followed by ethanol precipitation. Reaction products were resolved on 8% polyacrylamide gels.

COS cell transfection. COS-7 cells (17) were maintained in Dulbecco minimal essential medium with 10% fetal calf serum. COS cells were transfected as previously described (34). Each 100-mm plate of cells received 1 ml of calcium phosphate-DNA coprecipitate (19), containing 15 to 30 μg of plasmid DNA. After incubation for 30 min at room temperature, fresh growth medium was added, and the plates were incubated for 5 to 7 h at 37°C. The cells were then shocked with 1 ml of 25% glycerol in growth medium for 1 min at room temperature, rinsed with medium, and refed with fresh growth medium. Cells were harvested 48 h posttransfection.

RESULTS

Construction and characterization of mutations in the RSV U3 region. The structure of the putative RSV promoter region is shown in Fig. 1; the RNA cap site, denoted +1, and sequence elements potentially involved in the regulation of transcription are indicated. The *Eco*RI site located in the U3 region at position −53 was utilized for the construction of mutations in this region. Mutations at the U3 *Eco*RI site were generated in the plasmid pSL103 (Fig. 2). Plasmid pSL103 was linearized by partial digestion with *Eco*RI and subjected to exonuclease *Bal* 31 digestion for various lengths of time. After incu-

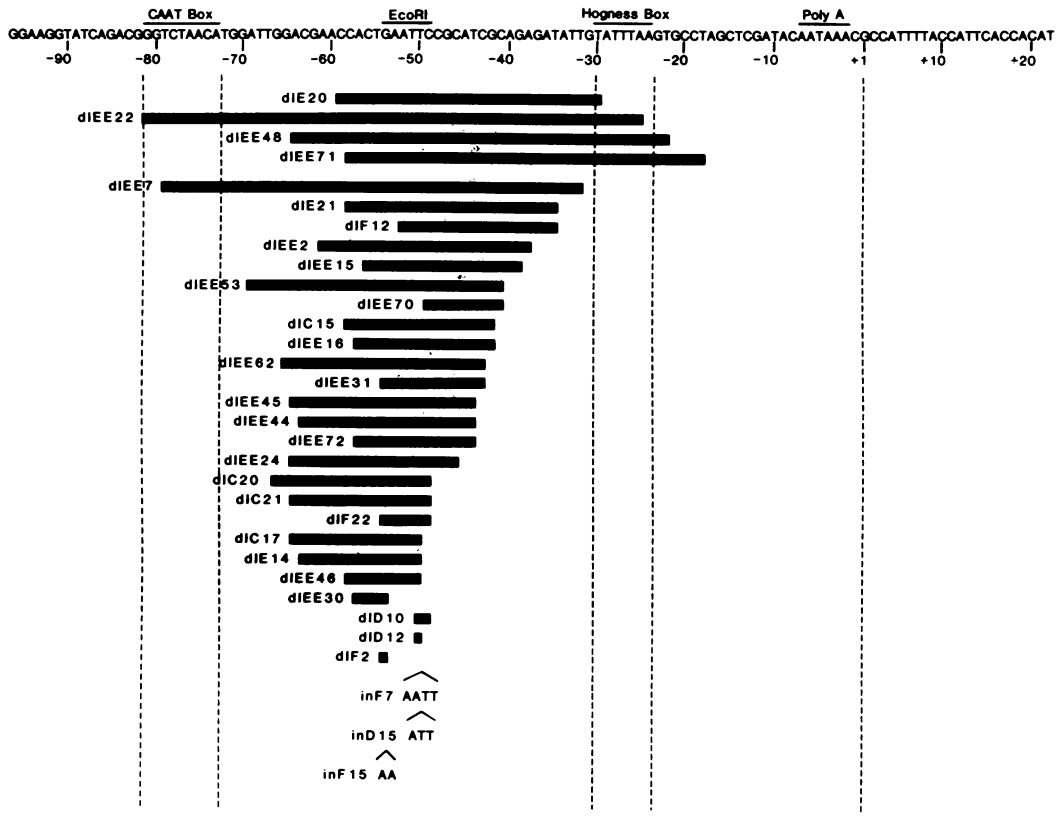

FIG. 3. Structure of mutations constructed at the U3 *Eco*RI site of pSL103. The nucleotide sequence flanking the RNA cap site is illustrated. The extent of each deletion is depicted by a solid bar preceded by the name of the mutant. Insertion mutations are denoted by a $\wedge$ over the inserted nucleotides at the point of insertion.

bation with DNA polymerase I to repair protruding 5′ ends, flush-ended molecules were circularized with T4 DNA ligase and subsequently used to transform *E. coli* HB101. Plasmid DNA was isolated from individual transformants and screened for the absence of the U3 *Eco*RI site. Sequence analysis of plasmids lacking the U3 *Eco*RI site revealed a spectrum of deletions as well as a set of small insertions (Fig. 3).

The effect of mutations in the U3 region upon virus replication was ascertained by transfection of chicken embryo cells. The viral DNA insert was excised from each of the mutagenized pSL103 plasmids by *Sal*I digestion and incubated with T4 DNA ligase. Chicken embryo cells were transfected with viral DNA and monitored for the appearance of transformed foci. Successful transfection of chicken embryo cells requires the efficient transcription of input DNA and subsequent formation of infectious particles (7). Therefore, cellular transformation serves as a sensitive indicator of functional DNA expression in the initial transfection assay. Culture supernatants from transfected cell cultures which failed to exhibit cellular transformation after 21 days were then assayed for the presence of reverse transcriptase activity (39). Four mutants were replication defective: *dl*E20, *dl*EE22, *dl*EE48, and *dl*EE71. All the remaining mutants were replication competent.

To obviate the possibility of wild-type virus contamination of mutant viral DNA-transfected cells, cellular DNA was isolated from transfected ev$^-$ chicken embryo cells 14 to 21 days posttransfection, digested with either *Eco*RI or *Pvu*I, and subjected to Southern blot analysis. *Eco*RI digestion of cellular DNA from wild-type viral DNA-transfected cells yielded three fragments of 3.8, 3.1, and 2.4 kb. *Eco*RI digestion of cellular DNA isolated from replication-competent mutant viral DNA-transfected cells (Fig. 4A) exhibited a single fragment of 3.8 kb. The deletion of the U3 *Eco*RI site resulted in the fusion of both the 3.1- and 2.4-kb proviral fragments with the virus-cell juncture fragments. In each case, *Pvu*I digestion generated a genome-length fragment of 9.3 kb. Southern blot analysis of cellular DNA isolated from cells transfected with replication-defective mutant viral DNA (Fig. 4B) failed to detect any of the expected

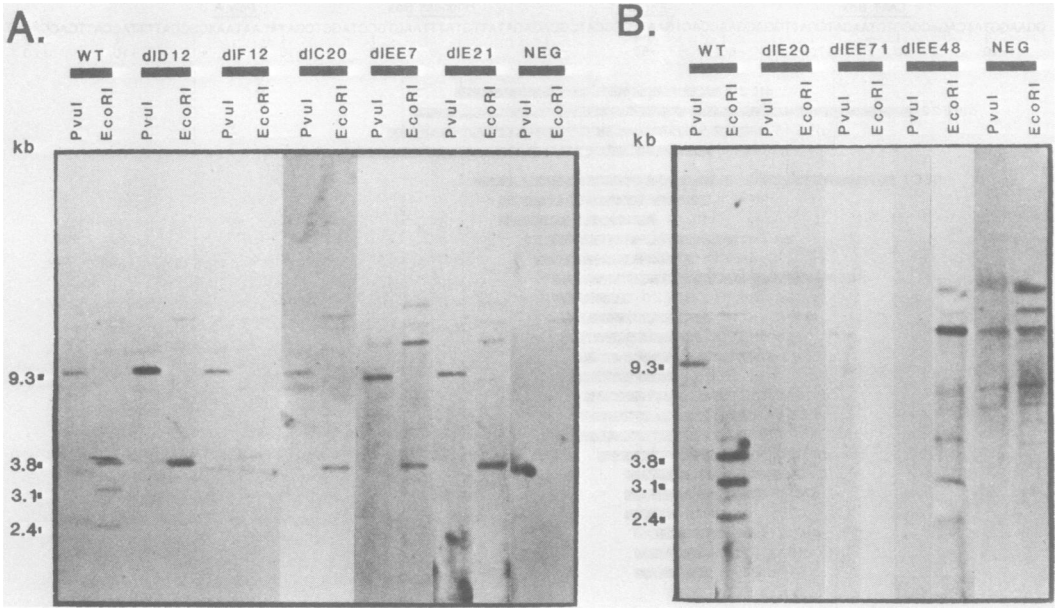

FIG. 4. Southern blot analysis of transfected ev⁻ chicken embryo cells. Cell DNA was isolated from transfected ev⁻ chicken embryo cells 14 to 21 days posttransfection, digested with either *Eco*RI or *Pvu*I, electrophoresed in 0.75% agarose gels, transferred to nitrocellulose filters, and hybridized with a ^{32}P-labeled RSV genomic probe as described in the text. Fragment sizes were determined by parallel electrophoresis of *Hin*dIII-digested DNA. (A) DNA from cells transfected with five representative replication-competent mutants. (B) DNA from cells transfected with three representative replication-defective mutants. Abbreviations: WT, DNA from pSL103-transfected cells; NEG, DNA from mock-transfected cells.

virus-specific fragments, indicating the absence of integrated proviral DNA. The fragments in the digests of cellular DNA from *dl*EE48-transfected cells, as well as those from the negative control, did not comigrate with any of the wild-type virus-specific fragments and are of undetermined origin. The 20-kb fragment in the *Eco*RI digests may represent c-*src* sequences (29).

These initial experiments therefore affirmed the altered genotype of the proviral DNA in replication-competent mutant viral DNA-transfected cells. Furthermore, we found no evidence for the presence of intact proviral DNA in cells transfected with the replication-defective mutants *dl*E20, *dl*EE22, *dl*EE48, and *dl*EE71.

Analysis of viral RNA transcripts. The influence of mutations in the U3 region upon viral transcription was examined by quantitation of the level of intracellular virus-specific RNA and the level of virion RNA production in transfected cell cultures. Table 1 shows the relative amounts of virus-specific RNA from cell cultures transformed with four individual replication-competent mutants. Analysis of either cloned (16) or uncloned transformed chicken embryo cells revealed no significant alteration of viral RNA levels (Table 1). Similar analysis of additional replication-competent mutants supported the conclusion that these cells infected with replication-competent mutants did not contain significantly altered levels of viral transcripts (data not shown).

Evidence indicates that retrovirus RNA transcription is initiated with the capped nucleotide (5, 47), designated +1 in Fig. 1. To examine the fidelity of mutant virus transcription initiation, we used the S1 nuclease protection technique of Weaver and Weissmann (46) to map the 5′ ends of the wild-type and mutant viral RNA transcripts. RNA from virus-infected chicken embryo cells was isolated and hybridized with a 5′ end-labeled restriction fragment that overlaps the 5′ end of the RNA transcripts (Fig. 5B). The RNA-DNA hybrids were digested with S1 nuclease, and the protected probe fragments were sized on an 8% polyacrylamide gel in parallel with the four Maxam and Gilbert sequencing reactions (28) of the RSV probe DNA (Fig. 5A). In each case, the major DNA fragment protected by mutant viral RNA comigrated with that protected by wild-type viral transcripts. We therefore conclude that within the constraints of the S1 nuclease assay, the 5′ ends of the mutant viral transcripts examined are identical to those of wild-type virus, and hence the site of transcription initiation appears unaffected by the range of mutations examined. One must note, however, that the alignment of the protected

TABLE 1. Quantitation of intracellular viral RNA and virion RNA

DNA	Virus-specific RNA[a]			
	Total cell RNA from:		Virion RNA from:	
	Cloned virus-transformed cells	Uncloned cells	Cloned virus-transformed cells	Uncloned cells
*dl*C17	0.48	1.19	1.04	0.78
*dl*C20	0.44	1.06	ND[b]	0.58
*dl*C21	1.06	1.26	0.60	0.65
*dl*F22	1.08	0.70	0.36	0.48
Prague A RSV	1.00	1.00	1.00	1.00
Mock infected	ND	0.01	ND	0.01

[a] Determined by hybridization of ^{3}H-labeled RNA with excess filter-bound DNA as described in the text. Values represent the average of duplicate analyses expressed as a fraction of the Prague A RSV value.
[b] ND, Not determined.

probe fragments with the DNA sequencing ladder requires a correction of at least one nucleotide. The sequencing reactions eliminate the modified base (28), whereas S1 nuclease digestion preserves the terminal nucleotide of the hybrid. Fragments generated by the sequencing reactions therefore migrated at least one nucleotide faster than the corresponding fragments generated by S1 nuclease. In addition, the chemical cleavage reactions yield fragments terminated with a 3′ phosphate group, which migrate slightly faster than the corresponding fragments generated by S1 nuclease digestion that possess a 3′ hydroxyl group. An inherent heterogeneity of S1-resistant DNA 3′ ends (23, 42, 46), due in part to steric hindrance by the $m^{7}G^{5'}ppp^{5'}Gm$ cap group of the hybridized RNA and the use of less than stringent S1 digestion conditions, may account for the set of fragments of slower than expected mobility.

In addition to the major S1 nuclease-resistant fragments of approximately 108 bp in length, we observed a set of protected fragments of significantly slower mobility (Fig. 5A). These fragments represented the products of S1 nuclease digestion of the ^{32}P-labeled probe annealed with two noncontiguous RNA sequences (data not shown). These fragments might be predicted from the composition of the probe, which possesses U5 sequences that reside at the 5′ end of the viral transcripts as well as U3 sequences present at the 3′ end of the same transcripts. In addition, a 17- to 21-bp directly repeated sequence is present at both termini (41). Hence, S1 nuclease digestion of such hybrids yields a DNA fragment protected to the 3′ endpoint of the deletion in the U3 region. These unique S1 products serve as an internal control for the absence of wild-type virus contamination of mutant virus-infected cells. An apparent exception, *dl*D12, may have resulted from the inability of S1 nuclease to efficiently recognize a single nucleotide loop-out under the S1 nuclease digestion conditions used.

RNA isolated from chicken embryo cells transfected with replication-defective viral DNA was also examined by S1 nuclease analysis. No hybridizable RNA was detectable in any of the replication-defective viral DNA-transfected cells harvested 21 days posttransfection (data not shown), thus confirming the Southern blot analysis.

Analysis of replication-defective viral transcription in COS cells. The further analysis of replication-defective viral RNA transcription required the examination of the transient expression of the transfected mutant viral DNA. Our first approach was to investigate the transient expression in chicken embryo cells of a wild-type viral DNA colinear with the RSV genome (data not shown). With this assay, the level of viral transcripts as detected by S1 analysis 30 to 48 h posttransfection was insufficient to warrant further investigation. To circumvent this problem, we employed an SV40 origin fragment (SV-ORI) vector (34) in COS cells, which offered sufficient amplification of the transfected DNA to permit transcriptional analysis. COS cells are replication origin-defective SV40-transformed CV-1 monkey cell lines that constitutively express large T antigen (17) and therefore support the efficient replication of bacterial plasmids containing a 228-bp SV40 origin fragment (34). The 228-bp SV40 origin fragment lacks sequences essential for efficient early region gene transcription (3), including the enhancer function of the 72-bp repeats (2) as well as the initiation sites of the major late mRNAs (14). The COS cell–SV-ORI vector system therefore serves as an efficient system for the amplification and transcriptional analysis of cloned eucaryotic genes.

The SV-ORI vector constructed for the analysis of RSV transcription, pTEB/103, was derived from the plasmid pSV010 and is diagrammed in Fig. 6. In addition to the SV40 origin, pTEB/103 contains a 3-kb *Xba*I/*Bam*HI fragment from pSL103 possessing 2,430 bp of viral sequences

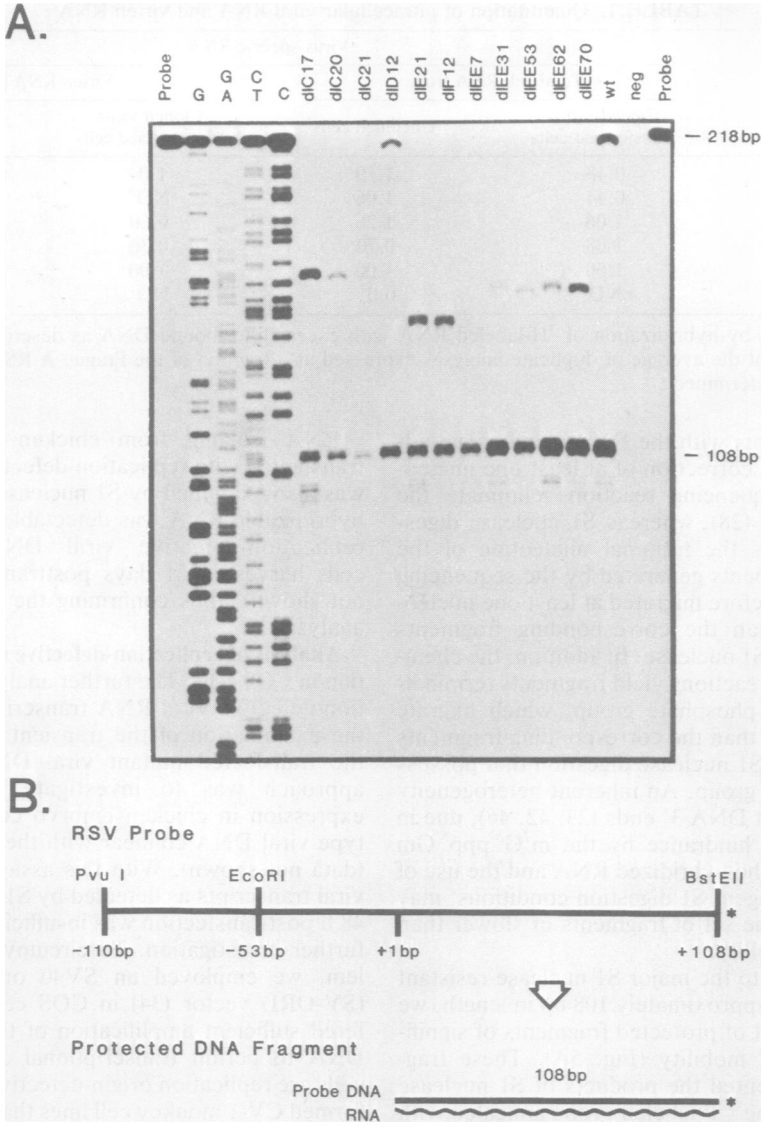

FIG. 5. S1 nuclease analysis of RSV transcripts isolated from chicken embryo cells 7 to 14 days posttransfection. RNA from cells transfected with individual replication-competent mutants (A) were hybridized with a 218-bp 5′ end-labeled *Bst*EII/*Pvu*I probe (B). The RNA-DNA hybrids were digested with S1 nuclease, and the S1-resistant fragments were sized on 8% polyacrylamide gels in parallel with the four Maxam and Gilbert sequencing reactions of the RSV probe DNA. Wild-type Prague A RSV transcripts protected a 108-bp fragment. Abbreviations: wt, RNA from Prague A RSV-transfected cells; neg, RNA from mock-transfected cells. The asterisks in (B) denote the position of the ^{32}P label. The variation in the S1 signal represents a variation in RNA inputs and does not represent a quantitative difference in virus-specific transcript.

upstream of the RNA cap site and 532 bp downstream, as well as a 3.4-kb *Bam*HI fragment encoding the herpes simplex virus thymidine kinase (*tk*) gene, which served as an internal transcription control. Correctly initiated transcripts of both the RSV and *tk* sequences were synthesized in COS cells upon transfection with pTEB/103 DNA. The S1 nuclease mapping experiment (Fig. 7A) showed that a 108-bp DNA fragment of the RSV probe (Fig. 7B) and a 56-bp DNA fragment of the *tk* probe (Fig. 7C) were protected by RNA synthesized in COS cells 48 h after transfection with pTEB/103. These labeled DNA fragments were of the size predicted after S1 nuclease digestion of hybrids formed between correctly initiated RSV and *tk* RNAs and their respective probes (32). In addition to the expected *tk* transcripts, we observed an even

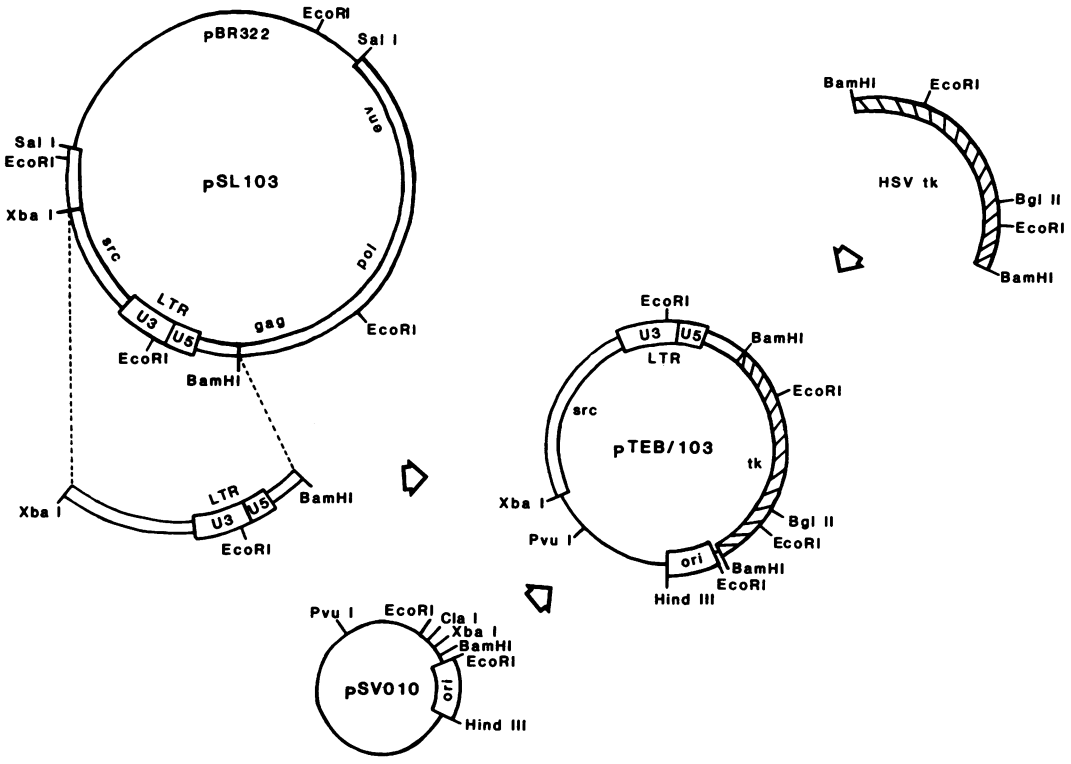

FIG. 6. Construction of the plasmid pTEB/103. The 3-kb *Xba*I/*Bam*HI fragment of pSL103 containing the LTR was inserted into the *Xba*I/*Bam*HI-digested SV-ORI vector, pSV010 (kindly provided by R. Tjian, University of California, Berkeley). A 3.4-kb *Bam*HI fragment containing the HSV *tk* gene (31) was then inserted into the *Bam*HI site at the junction of the pSL103 and pSV010 sequences. Transcription initiated within the LTR proceeds clockwise, whereas transcription of the *tk* gene is counterclockwise in this construct.

more abundant set of transcripts originating upstream of the expected 5′ end of the *tk* gene. These transcripts have been consistently observed in the COS cell system as well as in vitro in HeLa cell extracts (38). In the plasmid pTEB/103, as in all subsequent SV-ORI constructions, the LTR and *tk* gene were oriented in opposite directions with respect to transcription. Neither qualitative nor quantitative transcriptional differences could be detected with the *tk* gene inserted in the opposite orientation (data not shown).

Derivatives of the plasmid pTEB/103 were constructed by replacing the virus-specific 3-kb *Xba*I/*Bam*HI insert with the cognate *Xba*I/*Bam*HI fragment from each of the four replication-defective mutants: *dl*E20, *dl*EE22, *dl*EE48, and *dl*EE71. The transient expression of these plasmids was then examined by S1 nuclease analysis of the RNA synthesized in COS cells 48 h posttransfection. S1 mapping of mutant viral DNA-transfected COS cell RNA with an RSV probe (Fig. 7B) revealed several new transcripts not detected in cells transfected with pTEB/103 DNA (Fig. 7A). In addition, cells transfected with pTEB/*dl*E20, pTEB/*dl*EE48, and pTEB/*dl*EE71 DNAs still yielded low, but detectable, levels of correctly initiated RSV transcripts.

Three major new RSV transcripts were detected in COS cells transfected with the plasmids pTEB/*dl*E20, pTEB/*dl*EE48, and pTEB/*dl*EE71. The 5′ ends of these transcripts appeared to map to A residues at +16, +19, and +21. Cells transfected with the plasmid pTEB/*dl*E20 also yielded an additional transcript mapping to an A residue at position +12. A minor set of transcripts mapped to a stretch of A and T residues between +4 and +9. These transcripts were also generated in wild-type virus-infected chicken embryo cells.

The level of RSV-specific transcripts in COS cells transfected with each of the mutant viral DNAs was quantitated relative to the level of RSV-specific transcripts in pTEB/103-transfected cells by using densitometry of the autoradiogram in Fig. 7A. The level of *tk* gene transcripts served as an internal control. The ratios of RSV transcripts to *tk* transcripts in each S1 assay were used to determine the relative abundance of mutant and wild-type RSV transcripts. The

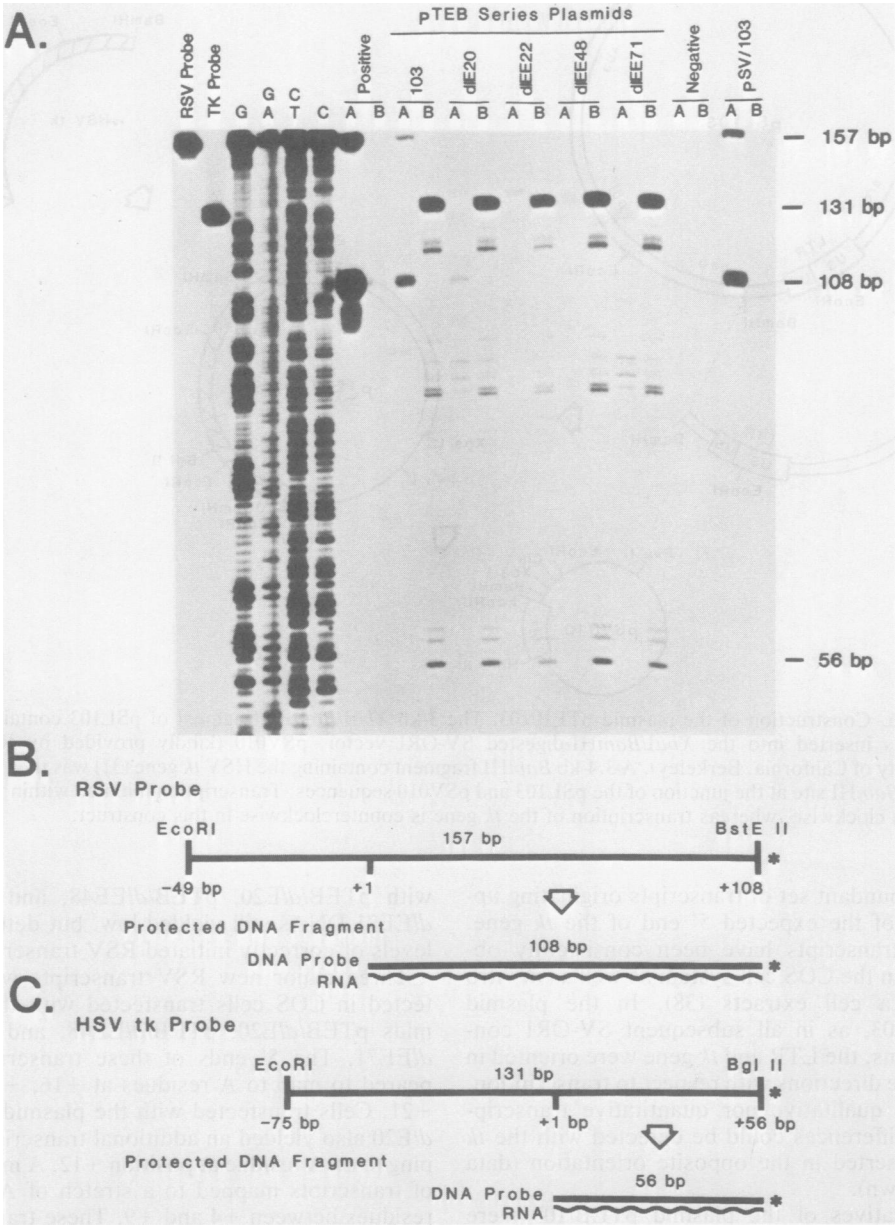

FIG. 7. S1 nuclease analysis of RSV and *tk* transcripts synthesized in COS cells, 48 h posttransfection. RNAs from COS cells transfected with plasmids pTEB/103, pTEB/*dl*E20, pTEB/*dl*EE22, pTEB/*dl*EE48, pTEB/*dl*EE71, and pSV/103 were annealed with each of two probes: (i) an RSV probe (B) and (ii) a *tk* probe (C). The RNA-DNA hybrids were digested with S1 nuclease, and the S1-resistant fragments were sized on an 8% polyacrylamide gel in parallel with the four Maxam and Gilbert sequencing reactions of the RSV probe DNA (A). The A tracks indicate hybridization with the RSV probe, and the B tracks indicate hybridization with the *tk* probe. The positive control was RNA isolated from Prague A RSV-infected chicken embryo cells, and the negative control was RNA isolated from mock-transfected COS cells. The plasmid pSV/103 is identical to pTEB/103 except that the 3.4-kb *Bam*HI fragment containing the *tk* gene has been removed. The asterisks in (B) and (C) denote the position of the ^{32}P label.

TABLE 2. Quantitation of relative levels of RSV transcripts synthesized in COS cells 48 h posttransfection

Plasmid DNA	bp deleted	Deletion endpoints	RSV-specific transfection efficiency[a]	
			Correct transcripts[b]	Total transcripts[c]
pTEB/103	0		1.00	1.00
pTEB/*dl*E20	30	−59/−30	0.10	0.25
pTEB/*dl*EE22	57	−81/−25	NT[d]	0.06
pTEB/*dl*EE48	43	−64/−22	0.01	0.04
pTEB/*dl*EE71	41	−58/−18	0.02	0.42

[a] Transcription efficiency is expressed as the ratio of the corrected level of mutant transcripts to the corrected level of pTEB/103 transcripts. The corrected level of transcripts is determined by the ratio of RSV transcripts to *tk* transcripts transcribed from a particular construct. The level of transcripts was quantitated by laser densitometry of the autoradiogram in Fig. 7.

[b] Transcripts with 5′ ends mapping to position +1.

[c] Excludes read-through transcripts initiated upstream of the 3′ end of the RSV probe.

30-bp deletion in *dl*E20 reduced the level of correctly initiated RSV transcripts by 90% (Table 2). Correctly initiated RSV transcripts from pTEB/*dl*EE48 and pTEB/*dl*EE71 DNA were present at 1 and 2% of the level of correct transcripts from pTEB/103 DNA, respectively, whereas no correctly initiated transcripts were observed with pTEB/*dl*EE22 DNA. Interestingly, the levels of total RSV-specific transcripts (both correctly and incorrectly initiated) from pTEB/*dl*E20 and pTEB/*dl*EE71 DNA were reduced only 4- and 2.5-fold, respectively, whereas the level of RSV-specific transcripts from pTEB/*dl*EE22 and pTEB/*dl*EE48 were reduced about 20-fold. These data therefore confirm earlier suggestions (18, 47) that the TATA-like sequence between −18 and −31 plays an essential role in viral RNA transcription and hence virus replication.

DISCUSSION

The structure of the RSV LTR and its presence at the termini of the integrated provirus suggest that this region may serve not only an essential replicative role in the processes of reverse transcription and integration but may function as the primary determinant of retrovirus gene expression as well. Putative eucaryotic RNA polymerase II transcriptional regulatory elements are found to reside within the LTR U3 region. A sequence closely related to the consensus form of the TATA box, 5′-TATA$^{A}_{T}$ A$^{T}_{A}$-3′ (8), is located at position −30 to −24, whereas a sequence similar in form to the CAAT box consensus sequence, 5′-GG$^{C}_{T}$CAATCT-3′ (4), is found at position −81 to −73. A putative eucaryotic polyadenylation signal, 5′-AATAAA-3′ (11), is also found at the 3′ end of the U3 region. Consistent with these observations is the ability of the LTR region to direct the synthesis of correctly initiated RSV transcripts in vitro (47) as well as to function as a strong promoter upon transfection of a variety of eucaryotic cells (18). The RSV LTR also appears to enhance the frequency of DNA-mediated cell transformation (12), a function shared by the SV40 72-bp directly repeated sequences.

To identify the transcriptional regulatory elements residing within the RSV LTR, we constructed a series of deletions and small insertions in the U3 region at a site 53 bp 5′ to the viral RNA cap site. In vivo analysis of the transcriptional activity of these mutants has allowed us to conclude that sequences within the −31 to −18 region, encompassing the TATA box, function in the selection of the correct site of transcription initiation and, in addition, augment the efficiency of transcription. Sequences within the −79 to −59 region, overlapping the putative CAAT box sequence, are dispensable in the presence of an intact TATA box, yet they appear to influence transcription in a quantitative fashion in its absence.

The role of the TATA box as a selector element in the specification of the site of transcription initiation has been well documented for a number of eucaryotic and viral genes transcribed by RNA polymerase II (3, 10, 13, 20, 22, 32, 33, 35), although its role in the modulation of transcriptional efficiency is less clear. In agreement with the results from several systems, deletions extending into the −30 to −18 sequences of the RSV U3 region invariably lead to an alteration in both the specificity and efficiency of transcription initiation. Notably, the mutant *dl*E20, possessing a deletion extending only through the first T residue of the TATA box, yields transcripts initiated at four new sites, a subset of which are found to be common to mutants *dl*EE48 and *dl*EE71, and a 10-fold reduction in the synthesis of correctly initiated transcripts.

The mutants *dl*E20 and *dl*EE71 possess deletions with nearly identical 5′ deletion termini, whereas their 3′ deletion termini differ by 12 bp. Cells transfected with these mutants exhibited only a 4- and 2.5-fold reduction in total RSV-specific transcripts, respectively, and the transcripts possessed a nearly identical complement of new initiation sites. Therefore, with the exception of a higher level of correctly initiated transcripts synthesized in cells transfected with *dl*E20, these two mutations induce very similar transcriptional phenotypes, both quantitatively and qualitatively, despite the extent of their

deletions. A similar phenotype is exhibited by a deletion mutant of the sea urchin histone *H2A* gene. The deletion of 54 bp immediately preceding the RNA cap site of the histone *H2A* gene, which includes the TATA box, results in both the appearance of new downstream transcripts as well as a fivefold lower transcriptional efficiency (20). It has also been shown that the replacement of the histone *H2A* gene TATA box with an equivalent sequence containing a point mutation within this region (TATA→TAGA) also results in a fivefold decrease in specific transcripts, although without a concomitant change in the specificity of initiation (21).

Although the transcriptional phenotypes of mutants *dl*E20 and *dl*EE71 were not at all surprising, the reduction in the transcriptional efficiency of mutants *dl*EE22 and *dl*EE48 was quite unexpected. Experiments conducted with the replication-competent mutants indicated that a range of deletions and small insertions in the region −79 to −32 had no detectable quantitative (Table 1, unpublished data) or qualitative (Fig. 5) effect upon transcriptional regulation. Our initial conclusion based on these observations was that the −79 to −32 region was of little functional significance in the promotion of RNA transcription. However, the examination of mutants *dl*EE22 and *dl*EE48 provided evidence to the contrary. The 3′ deletion termini of mutants *dl*EE22 and *dl*EE48 fall between those of *dl*E20 and *dl*EE71 (Fig. 2), two mutants with transcriptional efficiencies significantly higher than *dl*EE22 and *dl*EE48 (Table 2). Therefore, the basis of the transcriptional phenotypes exhibited by *dl*EE22 and *dl*EE48 is presumed to reside in the sequences altered by the 5′ deletion termini. Whereas the 5′ deletion termini of *dl*E20 and *dl*EE71 are at positions −59 and −58, respectively, *dl*EE48 deletes through position −64 and *dl*EE22 deletes through position −81. These mutations therefore appear to define a region from −59 to −81 which includes a CAAT box-like sequence that is required for the efficient synthesis (in COS cells) of RSV-specific transcripts in the absence of an intact TATA box. Alternatively, the juxtaposition of upstream sequences as a consequence of deletion may influence transcriptional initiation events. Transcriptional analysis of several eucaryotic and viral genes reveals that, in most cases, the −100 to −50 region encodes a quantitative regulatory function; mutations in this region most often result in a decrease in transcriptional efficiency without a concomitant change in specificity (10, 22, 33, 38). A notable exception to this trend is a sea urchin histone *H2A* gene deletion mutation extending from −55 to −110, leaving the TATA box intact, that results in a twofold increase in specific transcription (20). This deletion may be functionally analogous to *dl*EE7, which deletes 48 bp, from −32 to −79, yet appears transcriptionally wild type.

The construction of replication-competent mutants possessing deletions of 1 to 48 bp and insertions of 2 to 4 bp in the −79 to −32 region reveals the flexibility in the spatial relationship between the TATA box and possible upstream regulatory sequences. Flexibility in the spacing of transcriptional elements has also been observed in the herpes simplex virus *tk* gene promoter region (30). Mutants *dl*EE48 and *dl*EE71 possess deletions of 43 and 41 bp, respectively, such that their sequences 3′ to −18 are identical, and the sequences 5′ to −23, although common to both, differ in spatial alignment by only 2 bp. We do not believe that such spatial differences alone account for the 10-fold difference in transcriptional efficiency between these two mutants. However, the juxtaposition of upstream sequences as a consequence of deletion must be viewed cautiously in lieu of a clear definition of the boundaries of regulatory sequence elements.

ACKNOWLEDGMENTS

We thank Betty Creasy for excellent technical assistance and M. Mitchell Smith for critical reading of the manuscript.

This work was supported by Public Health Services grants CA29243 and CA27578 from the National Cancer Institute and grant MV-29E from the American Cancer Society. J.T.P. is a recipient of a faculty research award from the American Cancer Society. G.M.G. was supported by training grant 5T32 GM07082-03 from the National Institutes of Health.

LITERATURE CITED

1. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
2. **Banerji, J., S. Rosconi, and W. Schaffner.** 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell **27**:299–308.
3. **Benoist, C., and P. Chambon.** 1981. *In vivo* sequence requirements of the SV40 early promoter region. Nature (London) **290**:304–310.
4. **Benoist, C., K. O'Hare, R. Brethnach, and P. Chambon.** 1980. The ovalbumin gene—sequence of putative control regions. Nucleic Acids Res. **8**:127–142.
5. **Benz, E. W., Jr., R. W. Wydro, B. Nadal-Ginard, and D. Dina.** 1980. Moloney murine sarcoma proviral DNA is a transcriptional unit. Nature (London) **288**:665–669.
6. **Collins, C. J., and J. T. Parsons.** 1977. Integration of avian sarcoma virus DNA sequences in transformed mammalian cells. Proc. Natl. Acad. Sci. U.S.A. **74**:4301–4305.
7. **Cooper, G. M., and S. Okenquist.** 1978. Mechanism of transformation of chicken embryo fibroblasts by Rous sarcoma virus DNA. J. Virol. **28**:45–52.
8. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and C. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
9. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene **6**:23–28.
10. **Dierks, P., A. Van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann.** 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6

cells. Cell 32:695–706.
11. **Fitzgerald, M., and T. Shenk.** 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24:251–260.
12. **Folger, K. R., E. A. Wong, G. Wahl, and M. R. Capecchi.** 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. 2:1372–1387.
13. **Ghosh, P. K., P. Lebowitz, R. J. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. U.S.A. 78:100–104.
14. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. Heterogeneity and 5'-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. 126:813–846.
15. **Gillespie, D., and S. Spiegelman.** 1965. A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. J. Mol. Biol. 12:829–842.
16. **Gilmer, T. M., and J. T. Parsons.** 1979. Analysis of cellular integration sites in avian sarcoma virus-infected duck embryo cells. J. Virol. 32:762–769.
17. **Gluzman, Y.** 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23:175–182.
18. **Gorman, C. M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. Howard.** 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. U.S.A. 79:6777–6781.
19. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467.
20. **Grosschedl, R., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. 77:1432–1436.
21. **Grosschedl, R., B. Wasylyk, P. Cambon, and M. L. Birnstiel.** 1981. Point mutation in the TATA box curtails expression of sea urchin H2A histone gene *in vivo*. Nature (London) 294:178–180.
22. **Grosveld, G. C., E. deBoer, C. K. Shewmaker, and R. A. Flavell.** 1982. DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. Nature (London) 295:120–126.
23. **Hentschel, C., J. C. Irminger, P. Bucher, and M. L. Birnstiel.** 1980. Sea urchin histone mRNA termini are located in gene regions downstream from putative regulatory sequences. Nature (London) 285:147–151.
24. **Highfield, P. E., L. E. Rafield, T. M. Gilmer, and J. T. Parsons.** 1980. Molecular cloning of avian sarcoma virus closed circular DNA: structural and biological characterization of three recombinant clones. J. Virol. 36:271–279.
25. **Levinson, B., G. Khoury, G. Vande Woude, and P. Gruss.** 1982. Activation of SV40 genome by 72-base pair tandem repeats of Moloney sarcoma virus. Nature (London) 295:568–572.
26. **Lusky, M., and M. Botchan.** 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature (London) 293:79–81.
27. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequences of the rightward operator of phage. Proc. Natl. Acad. Sci. U.S.A. 72:1184–1188.
28. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
29. **McClements, W., H. Hanafusa, S. Tilghman, and A. Skalka.** 1979. Structural studies on oncornavirus-related sequences in chicken genome DNA: two-step analysis of *Eco*RI and *Bgl*I restriction digests and tentative mapping of a ubiquitous provirus. Proc. Natl. Acad. Sci. U.S.A. 76:2165–2169.
30. **McKnight, S. L.** 1982. Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell 31:355–365.
31. **McKnight, S. L., and E. R. Gravis.** 1980. Expression of the herpes thymidine kinase gene in Xenopus laevis oocytes: an assay for the study of deletion mutants constructed *in vitro*. Nucleic Acids Res. 8:5931–5948.
32. **McKnight, S. L., E. R. Gavis, R. Kingsbury, and R. Axel.** 1981. Analysis of the transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell 25:385–398.
33. **McKnight, S. L., and R. Kingsbury.** 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science 217:316–324.
34. **Mellon, P., V. Parker, Y. Gluzman, and T. Maniatis.** 1981. Identification of DNA sequences required for transcription of the human 1-globin gene in a new SV40-vector system. Cell 27:279–288.
35. **Osborne, T. F., R. B. Gaynor, and A. J. Berk.** 1982. The TATA homology and the mRNA 5' untranslated sequence are not required for expression of essential adenovirus E1A functions. Cell 29:139–148.
36. **Parsons, J. T., P. Lewis, and P. Dierks.** 1978. Purification of virus-specific RNA from chicken cells infected with avian sarcoma virus: identification of genome-length and subgenome-length viral RNAs. J. Virol 27:227–238.
37. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature (London) 295:209–214.
38. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the Drosophila Hsp 70 heat-shock gene. Cell 30:517–528.
39. **Peters, G. G., and J. Hu.** 1980. Reverse transcriptase as the major determinant for selective packaging of tRNAs into avian sarcoma virus particles. J. Virol. 36:692–700.
40. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869.
41. **Schwartz, D. E., P. C. Zamecnik, and H. L. Weith.** 1977. Rous sarcoma virus is terminally redundant: the 3' sequence. Proc. Natl. Acad. Sci. U.S.A. 74:994–998.
42. **Sollner-Webb, B., and R. H. Reeder.** 1979. The nucleotide sequence of the initiation and termination sites for ribosomal RNA transcription in X. laevis. Cell 18:485–499.
43. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
44. **Tsai, S. Y., D. R. Roop, W. E. Stumph, M. Tsai, and B. O'Malley.** 1980. Evidence that deoxyribonucleic acid sequences flanking the ovalbumin gene are not transcribed. Biochemistry 19:1755–1761.
45. **Varmus, H., and R. Swanstrom.** 1982. Replication of retroviruses, p. 450. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses: molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
46. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15 S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7:1175–1193.
47. **Yamamoto, T., B. de Crombrugghe, and I. Pastan.** 1980. Identification of a functional promoter in the long terminal repeat of Rous sarcoma virus. Cell 22:787–797.