JOURNAL OF VIROLOGY, Apr. 1992, p. 1959–1970
0022-538X/92/041959-12$02.00/0

Vol. 66, No. 4

CArG, CCAAT, and CCAAT-Like Protein Binding Sites in Avian Retrovirus Long Terminal Repeat Enhancers

KAREN R. ZACHOW AND KATHLEEN F. CONKLIN*

Institute of Human Genetics and Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455

Received 3 October 1991/Accepted 18 December 1991

A strong enhancer element is located within the long terminal repeats (LTRs) of exogenous, oncogenic avian retroviruses, such as Rous sarcoma virus (RSV) and the avian leukosis viruses. The LTRs of a second class of avian retroviruses, the endogenous viruses (evs), lack detectable enhancer function, a property that correlates with major sequence differences between the LTRs of these two virus groups. Despite this lack of independent enhancer activity, we previously identified sequences in ev LTRs that were able to functionally replace essential enhancer domains from the RSV enhancer with which they share limited sequence similarity. To identify candidate enhancer domains in ev LTRs that are functionally equivalent to those in RSV LTRs, we analyzed and compared ev and RSV LTR-specific DNA-protein interactions. Using this approach, we identified two candidate enhancer domains and one deficiency in ev LTRs. One of the proposed ev enhancer domains was identified as a CArG box, a motif also found upstream of several muscle-specific genes, and as the core sequence of the c-*fos* serum response element. The RSV LTR contains two CArG motifs, one at a previously identified site and one identified in this report at the same relative location as the ev CArG motif. A second factor binding site that interacts with a heat-stable protein was also identified in ev LTRs and, contrary to previous suggestions, appears to be different from previously described exogenous virus enhancer binding proteins. Finally, a deficiency in factor binding was found within the one inverted CCAAT box in ev LTRs, affirming the importance of sequences that flank CCAAT motifs in factor binding and providing a candidate defect in the ev enhancer.

Detailed studies of several eukaryotic transcriptional enhancers indicate that these elements share a modular organization, being composed of multiple interacting domains, each of which binds specific *trans*-acting proteins (reviewed in reference 2). In some cases, disruption of individual domains can have little effect on the activity of the enhancer, while mutation of other regions can abrogate enhancer function. Multimerization of some enhancer domains can also generate a functional enhancer. Together, these data suggest that while the function(s) of some enhancer domains and their cognate proteins is unique and essential, the function(s) of others may be redundant. We have analyzed the functional relatedness and interactions of discrete enhancer domains by studying sequences within the long terminal repeats (LTRs) of different types of avian retroviruses that exhibit either high or low enhancer function.

The avian exogenous viruses, which include Rous sarcoma virus (RSV) and the avian leukosis viruses (ALVs), contain a strong enhancer in the U3 portion of the viral LTR that is required for high-level expression from the viral promoter and that also is able to augment transcription from various heterologous promoters (6, 7, 22, 24, 30, 53). In ALVs, the LTR-associated enhancer is also an important contributor to cellular transformation, since it is responsible for the increased transcription of cellular oncogenes adjacent to sites of provirus integration (19, 32). An understanding of factors that mediate the activity of this enhancer is therefore important for elucidating the cellular components required for efficient retroviral transcription and retrovirus-mediated oncogenesis.

At least four proposed enhancer domains that interact with cellular proteins have been identified in exogenous virus LTRs. The first two include sequences located between the 5′ end of the viral LTR and an *Sph*I site at −141 from the start site of transcription. Cellular proteins that bind to sequences within these two domains have been identified and are called EFII (42), a1 (35, 36), c/EBP-like (38) or FIII (14), and a3 (35, 36) or FI (14). Two additional domains located downstream of the *Sph*I site at −141 include an inverted CCAAT box that is centered at −131 and that binds a cellular protein(s) called EFI (11, 17, 31, 42) or FII (14) and a region that is located between positions −112 and −87 and that binds the EFIII protein (3). The EFIII binding site includes a CArG box, defined as 5′-CC(A/T)$_6$GG-3′. CArG motifs are found upstream of several cellular genes, including the cardiac and skeletal actin genes (46), and are required for their tissue-specific expression (26, 29, 52). A CArG motif is also included in the c-*fos* serum response element (SRE; 12, 16, 28, 50), which mediates both cycloheximide-inducible expression and serum-inducible expression of c-*fos* (12, 16, 28, 49, 50) as well as its subsequent repression (34). Sealy (42a) and coworkers have found that the EFIII binding site, like other CArG motifs (48, 52), can mediate a modest transcriptional stimulation of linked promoters in response to serum. It has been proposed (3) that EFIII may be the avian homolog of the serum response factor (SRF), a positive regulatory protein that interacts with the c-*fos* SRE (12, 16, 33, 40, 50, 51). Promoter-proximal transcriptional control sequences also present in exogenous retroviral LTRs (13, 25, 54) include a second inverted CCAAT box that is centered at position −67 and that is a weak binding site for EFI (11, 17) and a TATA box that is located at position −24. At least some of the proteins that interact with exogenous virus LTRs show cell-type-specific differences in abundance and half-life that may be related to the oncogenic spectrum of virus isolates (35, 36).

The LTRs of the second class of avian retroviruses, the

* Corresponding author.

avian endogenous viruses (evs), are distinct from those of the exogenous viruses in that they lack a detectable associated enhancer (6–8, 30, 53). The absence of a strong enhancer in ev LTRs has been correlated with the low growth rate and low oncogenic potential of evs relative to exogenous viruses (53) as well as with major sequence differences between the U3 regions of these two virus groups (20, 21, 39, 41). In particular, the ev U3 region is shorter than that of exogenous viruses, containing gaps in sequence as well as regions of sequence divergence that include sequences that correspond to the two 5′-most exogenous virus enhancer domains (the EFII, FIII, or a1 site and the FI or a3 binding site) and to the RSV EFIII binding site. Although a detailed characterization of proteins that interact with ev LTRs has not been conducted, Ryden and Beemon (38) have demonstrated that ev LTRs do contain one protein binding site for a c/EBP-like, heat-stable protein between positions −57 and −77 and that this site is not found in exogenous virus LTRs at the corresponding positions. It has been suggested that this protein may be related to that which binds to one or both of the 5′-most proximal enhancer domains of RSV (38).

In a previous study on the function of retroviral LTR-associated enhancers, ev-RSV hybrid LTRs that contained a portion of the ev U3 region (from the 5′ end of the LTR through an *Acc*I site at −51) in place of essential enhancer sequences in the RSV LTR upstream of the *Sph*I site at −141 were constructed (5). Surprisingly, it was found that the ev U3 fragments, which lack independent enhancer function, were able to functionally replace essential enhancer sequences from RSV in an orientation-independent manner. The activity of the ev U3 region in these ev-RSV hybrid LTRs was not explained by differences in the ability of ev and RSV promoters to respond to enhancers placed in *cis*. In agreement with previous data (6–8, 30, 54), it was also found that the same ev U3 fragment that was able to restore transcription to the enhancer-deficient RSV LTR was unable to enhance transcription from an unrelated promoter, that of the herpesvirus thymidine kinase gene. From these data, it was concluded that ev LTRs contain a partial or weak enhancer element that can be complemented by sequences from the RSV LTR downstream of the *Sph*I site at −141 to generate an intact, efficient enhancer element.

The data cited above indicate that the 5′ portion of the ev U3 region must contain sequences that are functionally equivalent to sequences upstream of the *Sph*I site in the RSV LTR. Because of the limited degree of sequence similarity between ev and exogenous virus LTRs, it seems unlikely that the U3 regions of these two virus groups bind the same *trans*-acting proteins that mediate enhancer function. However, there are short (up to 9-bp) regions of similarity between the ev and exogenous virus LTRs that may contain sufficient information for binding of the relevant proteins. In addition, it is possible that ev and RSV U3 regions contain related sequences that are organized differently in their respective LTRs and that are therefore difficult to identify by simple sequence alignments. Alternatively, the ev U3 region may recognize a unique protein(s) that is functionally equivalent to those utilized by exogenous virus LTRs. Such a finding would provide an opportunity to compare functional domains of distinct, yet functionally related, cellular proteins important in enhancer function.

The experiments described in this report were designed to identify the relevant sequences from both the ev and the RSV U3 regions that contribute to the composite enhancer found in the ev-RSV hybrid LTRs and to compare the potential enhancer domains identified in ev LTRs with those previously described in the exogenous virus U3 region.

MATERIALS AND METHODS

Preparation of nuclear extracts. S13, a chicken B-cell line derived from an ALV-induced B-cell lymphoma (23), and MSB, a chicken cell line derived from a Marek's disease virus-induced T-cell lymphoma (1), were grown in Dulbecco's modified Eagle's medium (GIBCO) supplemented with 5% defined bovine serum (Hyclone), 1% chicken serum (GIBCO), and 10% tryptose phosphate broth (Difco). Nuclear protein extracts were prepared by a modification of a previously published procedure (9) as described below. Approximately 1×10^9 to 2×10^9 exponentially growing cells were centrifuged in a Beckman GPR centrifuge at 2,500 rpm for 5 min at 4°C, the cell pellet was washed one time with a balanced salt solution and resuspended in five cell pellet volumes of buffer A (10 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES] [pH 8], 1.5 mM $MgCl_2$, 10 mM NaCl, 2 mM dithiothreitol [DTT], 10 μg of leupeptin per ml, 10 μM phenylmethylsulfonyl fluoride [PMSF], 0.68 μg of pepstatin per ml), and the suspension was kept on ice for 5 to 10 min. After centrifugation at 3,500 rpm for 5 min at 4°C in a Beckman GPR centrifuge, the cell pellet was resuspended in two cell pellet volumes of buffer A and the suspension was homogenized on ice for 5 to 10 strokes with a Dounce homogenizer and a B pestle. After microscopic examination for lysis, nuclei were collected by centrifugation in a Beckman JS13 rotor at 8,000 rpm for 5 min. The nuclear pellet was resuspended in 1.5 ml of a buffer containing 20 mM HEPES [pH 8.0], 20% [vol/vol] glycerol, 0.4 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 2 mM DTT, 10 μg of leupeptin per ml, 10 μM PMSF, and 0.68 μg of pepstatin per ml per 10^9 cells. The suspension was kept on ice for 30 min, with vortexing every 5 min, and then centrifuged in a Beckman JS13 rotor at 8,000 rpm and 4°C for 10 min. The supernatant was dialyzed with a SPECTRA/POR 3 membrane against 200 volumes of a buffer containing 20 mM HEPES [pH 8.0], 20% [vol/vol] glycerol, 0.1 M NaCl, 0.2 mM EDTA, 1 mM DTT, 10 μg of leupeptin per ml, 10 μM PMSF, 0.68 μg of pepstatin per ml) for 60 to 90 min at 4°C. The dialyzed material was centrifuged at 8,000 rpm in a Beckman JA17 rotor for 10 min at 4°C, and supernatants were divided into aliquots, frozen in liquid nitrogen, and stored at −80°C. Aliquots were thawed immediately before use, and any unused portion was discarded.

DNA fragments and oligonucleotides. DNA fragments used for electrophoretic mobility shift assays (EMSA) were generated from plasmid DNA by use of the polymerase chain reaction (PCR) and a Perkin-Elmer Cetus thermal cycler. The oligonucleotide primers used in these reactions to delimit the DNA fragments are shown in Fig. 1, and the templates were plasmids that contained either the Rous-associated virus-0 (RAV-0) LTR (kindly provided by P. Tsichlis) or the RSV LTR (from pRSVNeo [15]; provided by C. Gorman). For generation of end-labeled DNA fragments, one of the oligonucleotide primers (20 pmol) was labeled on its 5′ end with T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ-^{32}P]ATP (NEN). After inactivation of the kinase, the labeled primer was added to 200 ng of template plasmid DNA and 20 pmol of the other oligonucleotide primer in a 100-μl reaction mixture containing 20 mM Tris-HCl (pH 8.4); 50 mM KCl; 1.5 mM $MgCl_2$; 0.2 mM each dATP, dCTP, dGTP, and dTTP; and 2.5 U of AmpliTaq

(Perkin-Elmer Cetus). DNA fragments were amplified for 25 cycles and purified by polyacrylamide gel electrophoresis.

Double-stranded oligonucleotides were generated by incubation of complementary oligonucleotides in 400 mM NaCl–10 mM Tris-HCl (pH 8.0)–1 mM EDTA at 65°C for 2 min and slow cooling to room temperature over 2 to 3 h.

Binding reactions. Nuclear extract (15 μg) was incubated in an 18-μl reaction mixture that contained a 1/10 volume of binding buffer (10× binding buffer is 50% glycerol, 10 mM EDTA, 500 mM NaCl, 100 mM Tris-HCl [pH 7.5], and 10 mM DTT) and various amounts of nonspecific competitor DNA for 10 min at room temperature. Both poly(dI) · poly(dC) and sonicated salmon sperm DNA were tested as nonspecific competitors; it was found that the salmon sperm DNA yielded a clearer banding pattern of DNA-protein complexes (data not shown). After this preincubation, 2 μl of labeled probe DNA (5,000 to 15,000 cpm and approximately 0.1 to 0.5 ng) was added, and the reaction mixture was incubated for a further 20 min at room temperature. Samples were applied to a 4.5% polyacrylamide gel in Tris-glycine buffer (40 mM Tris-HCl [pH 8.0], 0.0384 M glycine, 2 mM EDTA) and subjected to electrophoresis at 50 mA (100 V) for approximately 3 h. Gels were subsequently dried and exposed to Kodak RP film at −70°C with Dupont Cronex intensifying screens.

Methylation interference. Probes labeled on one strand were added to 200 μl of dimethyl sulfate (DMS) reaction mixture (50 mM sodium cacodylate [pH 8], 1 mM EDTA) with 10 μg of tRNA, and the mixture was incubated on ice for 5 min; 1 μl of concentrated DMS (Aldrich) was then added. Samples were mixed and incubated at room temperature for 5 to 10 min; 50 μl of DMS stop mixture (1.5 M sodium acetate [pH 7], 1.0 M β-mercaptoethanol, 100 μg of tRNA per ml) was then added, and DNA was ethanol precipitated. The resultant pellets were resuspended in 250 μl of 0.3 M sodium acetate, precipitated by the addition of ethanol, washed at least one time with 80% ethanol, and resuspended in sterile distilled water. The binding reactions with methylated probes were conducted in a manner similar to those with unmethylated probes, except that approximately 10^5 counts were added per 15 μg of extract. In each case, four to six individual reactions were conducted for each sample. After gel electrophoresis, bound and free fractions were localized by exposing the wet gel to X-ray film and were purified from the gel slice.

For generation of A/G ladders, probe DNA was mixed with 10 μg of tRNA in a final combined volume of 10 μl; 25 μl of concentrated formic acid was then added, and samples were incubated at room temperature for 10 to 15 min. Sodium acetate (0.3 M; 250 μl) and absolute ethanol (750 μl) were added, and samples were frozen and centrifuged. Samples were resuspended in 250 μl of 0.3 M sodium acetate, ethanol precipitated, and washed as described above.

The methylated and depurinated DNAs were cleaved by resuspension of the dried pellets in 100 μl of 1 M piperidine and incubation at 90°C for 30 min. The cleavage products were recovered by ethanol precipitation after the addition of 1 to 10 μg of tRNA and 100 μl of 0.3 M sodium acetate. The dried pellets were washed at least one time with 80% ethanol, resuspended in 90% formamide–1× TBE (54 g Trizma [Sigma], 27.5 g boric acid, 3.7 g disodium EDTA per liter) dye mixture, and counted, and approximately 2,000 to 5,000 cpm of each sample was applied to 12 to 18% denaturing polyacrylamide gels.

RESULTS

Detection of a protein that binds to the ev U3 region. Figure 1 shows a sequence comparison of the RSV (exogenous virus) and the ev-2–RAV-0 (endogenous virus) U3 regions aligned to show maximum homologies. To compare the protein binding regions in the ev U3 region with those in the exogenous virus enhancer, we generated DNA fragments from each LTR by PCR as described in Materials and Methods and used them in EMSA. The initial DNA fragments used in these studies were 106 bp (ev U3 fragment) and 143 bp (RSV U3 fragment) in length and extended from the 5′ end of the LTR to a region which, in exogenous virus LTRs, defines the 3′ boundary of the EFIII binding site (defined on the 5′ end by primers RSV-a and ev-a and on the 3′ end by primers RSV-c and ev-c; Fig. 1). The 106-bp fragment from the ev LTR did not contain the c/EBP-like protein binding site previously identified in the ev U3 region (38). Each of these fragments was incubated with avian nuclear extracts prepared from S13 cells (a chicken B-cell line generated from an ALV-induced B-cell lymphoma; 23) and from MSB cells (a Marek's disease virus-transformed chicken T-cell line; 1), and the resultant DNA-protein complexes were separated on polyacrylamide gels as described in Materials and Methods. In all cases tested, identical gel shift patterns were obtained with extracts generated from each of these cell types (data not shown).

As shown in Fig. 2, lane 1, the exogenous virus-derived U3 region fragment gave rise to two prominent bands and one diffuse band that migrated more slowly than the input probe DNA after incubation with extracts prepared from S13 cells. Methylation interference assays revealed that the faster-migrating of the DNA-protein complexes corresponded to binding at the upstream inverted CCAAT box (binding site for EFI [42] or FII [14]), while the slower-migrating band corresponded to binding at the EFIII binding site (20; data not shown). The diffuse band above the EFIII-DNA complex did not generate a consistent methylation interference pattern, and its origin has therefore not been determined. No evidence of binding over the two 5′-most RSV enhancer domains was detected in these experiments (corresponding to the EFII [42], FIII [14], or a1 [35, 36] binding site and the FI [14] or a3 [35, 36] binding site). This result is in agreement with those of previous studies that have demonstrated weak binding over the two 5′-most enhancer domains of exogenous virus LTRs with unfractionated extracts (3, 14, 35, 36, 42).

In contrast to data obtained with the RSV U3 fragment, the ev U3 fragment gave rise to only one shifted band (Fig. 2, lane 2) that migrated in a manner similar to that of the RSV-EFIII complex. To determine whether the ev DNA-protein complex was specific, we incubated the radiolabeled 106-bp ev U3 fragment with nuclear extracts in the presence of excess unlabeled DNAs. As shown in Fig. 3, the addition of a 50- to 200-fold molar excess of unlabeled 106-bp fragment resulted in a substantial decrease in the complex (compare lane 2 with lanes 9 to 11). In contrasts, the addition of up to a 200-fold molar excess of a 222-bp fragment from pUC had no effect on binding (data not shown), indicating that the ev LTR-protein complex constitutes a sequence-specific interaction.

An upstream inverted CCAAT box in ev LTRs does not efficiently bind EFI or FII protein(s). Exogenous virus LTRs contain two inverted CCAAT box motifs centered at positions −131 and −67 (Fig. 1). The upstream site includes the binding site for EFI or FII (11, 14, 17, 31, 42), while the

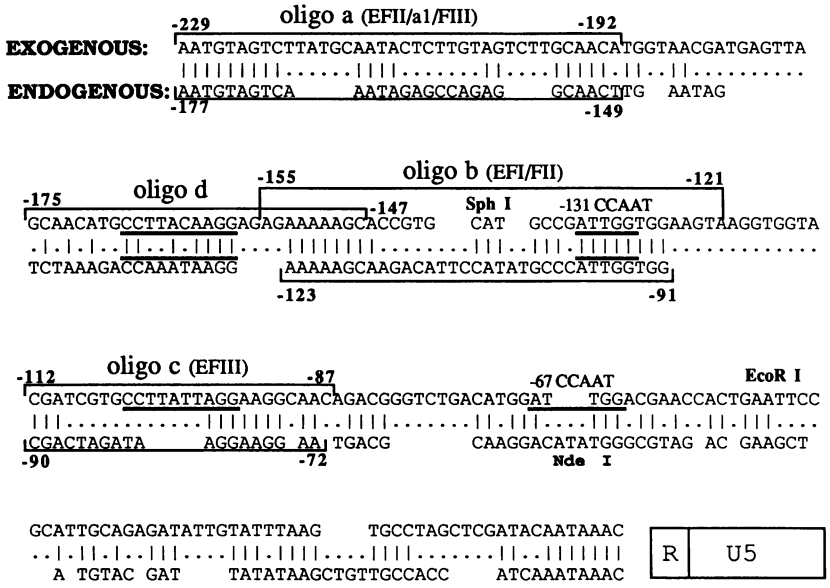

FIG. 1. Sequence comparison of exogenous virus and ev U3 regions. The sequences of the exogenous virus (RSV) and ev (ev-2–RAV-0) U3 regions are from previously published sequences (20, 21, 39, 41) and were independently confirmed. Vertical lines between the sequences represent regions of identity, while dots represent mismatches. Gaps were introduced to allow maximum alignment. The locations of selected restriction enzyme sites are shown. Letters above bracketed regions represent the oligonucleotides used in the experiments and correspond to binding sites as follows: RSV-a (−229 to −192) = EFII (42), FI (14), and a1 (35, 36); RSV-b (−155 to −121) = EFI (42), or FI (14); RSV-c (−112 to −87) = EFIII (3); and RSV-d (−175 to −145) = upstream CArG motif identified in this report. Brackets and nucleotide numbers in the ev sequence denote the corresponding ev-a, ev-b, and ev-c oligonucleotides. Identified motifs within the regions defined by the b, c, and d oligonucleotides are underlined (exogenous) or overlined (endogenous). Not shown is an additional factor binding site in the RSV LTR between positions −192 and −180 that includes the FI (14) or a3 (35, 36) binding site.

downstream site serves as a lower-affinity binding site for EFI (11, 17). Deletion and point mutation analyses have demonstrated that each of these motifs is important for high-level transcription from the exogenous virus promoter (6, 17) and may be important in mediating serum-responsive transcription from the exogenous virus LTR (10). ev LTRs lack the promoter-proximal inverted CCAAT box motif but contain the upstream motif centered in ev LTRs at position −96 (Fig. 1). In the experiments discussed above, however, the ev U3 fragment failed to generate a DNA-protein complex similar in size to the RSV-EFI or -FII complex, suggesting that the upstream inverted CCAAT box motif in ev LTRs either binds a different protein(s) than does RSV or is deficient in EFI or FII binding.

To address these questions, we generated two overlapping DNA fragments that included sequences from the 5′ and 3′ portions of the ev 106-bp probe and used them as both competitor DNAs in binding reactions with the 106-bp probe and probes in additional EMSA. The 5′-proximal fragment of 87 bp extended from the 5′ end of the ev LTR through position −91 (defined by primers ev-a and ev-b; Fig. 1) and therefore included the inverted CCAAT box as well as 3 bp of downstream DNA, while the 3′-proximal fragment of 52 bp extended from position −123 through position −72 (defined by primers ev-b and ev-c; Fig. 1) and therefore included 25 and 19 bp of DNA upstream and downstream of the CCAAT box motif, respectively.

Figure 3 shows data obtained when each of these unlabeled subfragments was added to binding reactions with the 106-bp ev DNA fragment as a probe. The 5′-proximal, 87-bp fragment was able to compete efficiently for binding to the 106-bp probe when added at a 50- to 200-fold molar excess, similar to the results obtained with competitor DNA homologous to the entire probe (compare lane 2 with lanes 3 to 5 and lanes 9 to 11). In contrast, the 52-bp, 3′-proximal fragment did not affect complex formation even when added at up to a 200-fold molar excess (lanes 6 to 8), suggesting that the intact protein binding region identified in the ev LTR is located in the 5′ portion of U3 and is not contained within the region that was present in both competitor subfragments and that included the inverted CCAAT box motif.

This suggestion was tested directly by using both the 87- and the 52-bp ev U3 subfragments as probes in EMSA, together with analogous fragments generated from the RSV U3 region as controls. As shown in Fig. 2, the 5′ portion of the RSV U3 region gave rise to one prominent DNA-protein complex (lane 4) that corresponded to binding of the EFI or FII factor to the upstream inverted CCAAT box (data not shown). A comparison of this complex with that obtained when the ev 87-bp fragment was used as a probe (lane 3) showed that the ev DNA-protein complex migrated significantly more slowly than did the RSV-EFI or -FII complex. When fragments from the 3′-proximal portion of the ev and RSV U3 regions were tested, it was found that the RSV probe (lane 5) gave rise to two DNA-protein complexes that corresponded to binding at the EFI or FII and the EFIII binding sites. When the corresponding ev U3 fragment was tested (lane 6), no significant binding was detected. Thus, while it is possible that the inverted CCAAT box in the ev LTR does serve as a binding site for a protein(s) not detectable in our extracts, it is not a high-affinity binding site for the same protein(s) that interacts with the exogenous virus LTR as assayed here.

Identification of the ev U3 binding site as a CArG box. The data discussed above indicate that the binding site for the ev-DNA binding protein was located in the 5′ portion of the

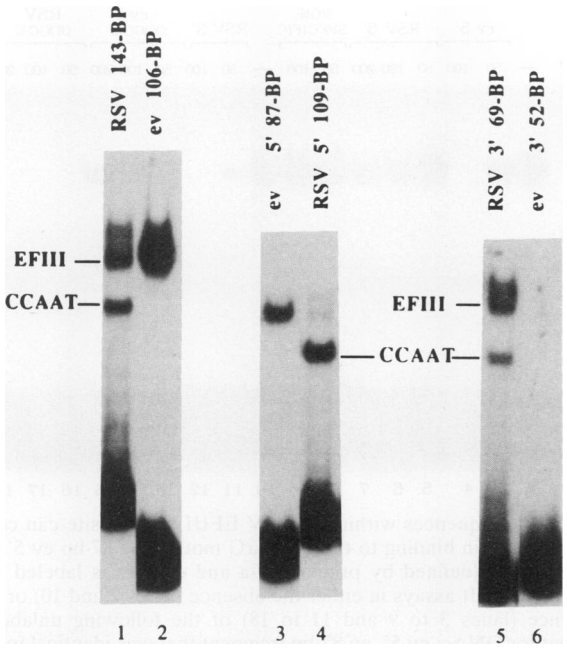

FIG. 2. Comparison of protein-DNA complexes in RSV and ev LTRs. The indicated DNA fragments from RSV or ev U3 regions were end labeled, incubated with 15 μg of nuclear protein extracted from S13 cells and nonspecific competitor salmon sperm DNA, and subjected to electrophoresis in polyacrylamide gels, which were subsequently dried and exposed to X-ray film. The CCAAT- and EFIII-specific complexes generated with RSV-derived DNA are marked and were identified by methylation interference assays (data not shown). The origin of, size of, and oligonucleotides used to generate each fragment by PCR as well as the micrograms of sonicated salmon sperm DNA added to each binding reaction were as follows: lane 1, RSV 143-bp fragment, primers RSV-a and RSV-c, 12 μg of salmon sperm DNA; lane 2, ev 106-bp fragment, primers ev-a and ev-c, 6 μg of salmon sperm DNA; lane 3, ev 87-bp fragment, primers ev-a and ev-b, 6 μg of salmon sperm DNA; lane 4, RSV 109-bp fragment, primers RSV-a and RSV-b, 6 μg of salmon sperm DNA; lane 5, RSV 69-bp fragment, primers RSV-b and RSV-c, 3 μg of salmon sperm DNA; and lane 6, ev 52-bp fragment, primers ev-b and ev-c, 1.5 μg of salmon sperm DNA.

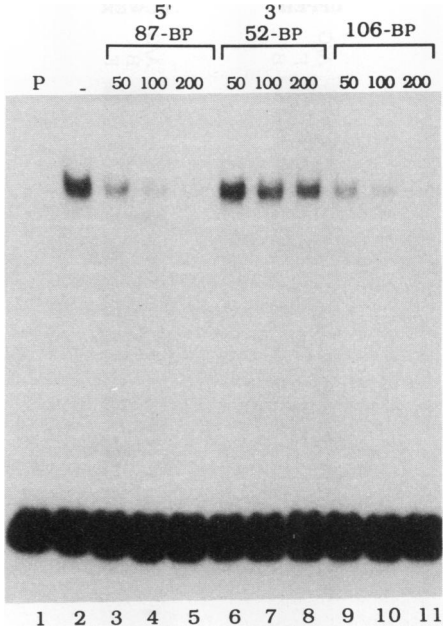

FIG. 3. The ev DNA-protein complex represents a sequence-specific interaction. The end-labeled 106-bp ev U3 fragment (generated with primers ev-a and ev-c) was incubated with 15 μg of protein extracted from S13 nuclei and 9 μg of sonicated salmon sperm DNA in the absence (lane 2) or presence (lanes 3 to 11) of a 50- to 200-fold molar excess of the following unlabeled competitor DNAs: 5′ 87-BP, a fragment that was from the 5′ end of the ev LTR and that was generated by PCR with the ev-a and ev-b primers; 3′ 52-BP, a fragment that was from the 3′ end of the ev LTR and that was generated by PCR with the ev-b and ev-c primers; and 106-BP, a fragment that was identical to the radiolabeled probe DNA used in this experiment and that was generated by PCR with the ev-a and ev-c primers. Lane P, input probe DNA that was not incubated with protein extracts.

ev U3 region, upstream of position −123. To verify this conclusion and to more precisely map the protein binding site, we conducted methylation interference assays with the 106-bp fragment labeled on either the upper or the lower strand as described in Materials and Methods. The results of this experiment are shown in Fig. 4. A comparison of the cleavage products generated from the free and bound fractions with the 106-bp DNA fragment labeled on the upper strand revealed depletion of two adjacent guanidine residues at positions −124 and −125 in the bound fraction, indicating that methylation of these residues interfered with protein binding. In support of data obtained with EMSA and discussed above, there was no evidence of binding over the inverted CCAAT box which, in exogenous virus LTRs, is affected by methylation at approximately seven residues within and flanking the inverted CCAAT box motif (14, 17; data not shown). Results obtained with the lower strand also revealed the depletion of two adjacent guanidine residues at positions −132 and −133 in the bound fraction; these residues were separated from those identified on the upper strand by six A-T base pairs. The proximity of affected guanidine residues identified on the upper and lower strands suggests that they identify contact points involved in the same protein binding site and therefore define a minimal recognition sequence of 10 bp composed of the sequence 5′-CCAAATAAGG-3′. A comparison of this protein binding site with known protein binding recognition sites revealed that it matched the consensus of a CArG box which is defined by the sequence 5′-CC(A/T)$_6$GG-3′. CArG boxes are required for the tissue-specific expression of cardiac and skeletal actin genes (26, 29, 52) and are a necessary component of the c-*fos* SRE, which is responsible for the serum-inducible transcription and subsequent repression of this gene (12, 16, 28, 34, 49, 50). The methylation interference pattern seen here for the ev CArG-protein complex identified the same protein contact points important for binding as have been identified for binding of the SRF to the CArG motif within the c-*fos* SRE (40, 50).

The RSV EFIII binding site competes for factor binding to the ev CArG box. Exogenous virus LTRs also contain a CArG box (5′-CCTTATTAGG-3′) at positions −95 to −104 within the EFIII binding site (3). The corresponding region in ev LTRs lacks this motif because of a deletion and base substitutions (Fig. 1). Published data (3), confirmed by us (data not shown), demonstrate that the RSV CArG-containing EFIII binding site generated the same methylation interference pattern as that seen in Fig. 4 for the ev CArG, suggesting that the ev and RSV sites, although present at

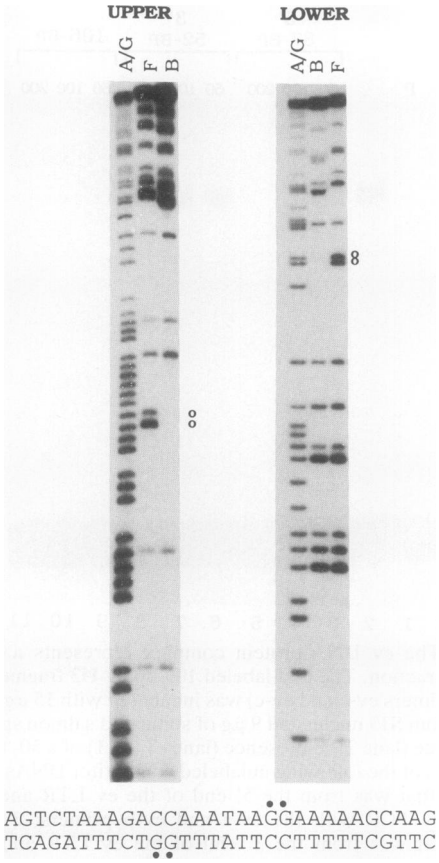

FIG. 4. Identification of the ev U3 binding site as a CArG box. The ev 106-bp fragment (generated with the ev-a and ev-c primers) was labeled individually on each strand with either radiolabeled primer ev-a (upper strand) or radiolabeled primer-c (lower strand) by PCR as described in Materials and Methods. The end-labeled probes were partially methylated with DMS and incubated with S13 nuclear extracts. After electrophoresis, the free (F) and bound (B) fractions were excised from the gel, purified, cleaved with piperidine, and subjected to electrophoresis in denaturing acrylamide gels. The A/G ladders were prepared in parallel as described in Materials and Methods. The sequence below the gel is the region from −143 to −114 of the ev LTR. The dots indicate the guanidine residues that were specifically depleted in the bound fraction.

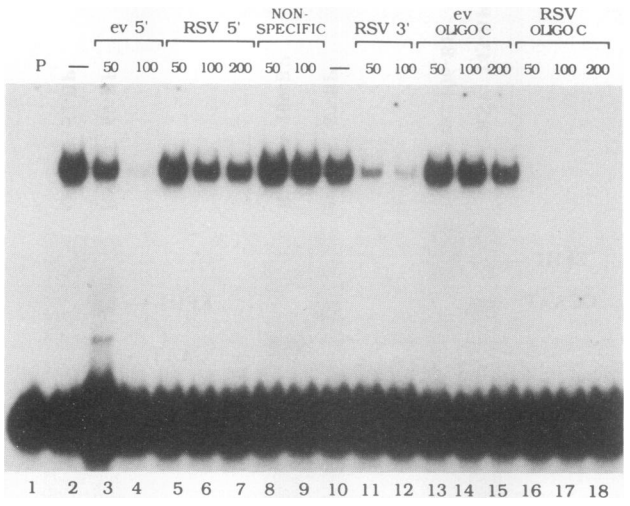

FIG. 5. Sequences within the RSV EFIII binding site can compete for protein binding to the ev CArG motif. The 87-bp ev 5′ U3 region probe (defined by primers ev-a and ev-b) was labeled and used in gel shift assays in either the absence (lanes 2 and 10) or the presence (lanes 3 to 9 and 11 to 18) of the following unlabeled competitor DNAs: ev 5′, an 87-bp fragment that was identical to the input probe DNA (lanes 3 and 4); RSV 5′, a 109-bp fragment defined by primers RSV-a and RSV-b (lanes 5 to 7); NON-SPECIFIC, a 222-bp fragment from pUC (lanes 8 and 9); RSV 3′, a 69-bp fragment defined by primers RSV-b and RSV-c (lanes 11 and 12); and ev OLIGO C (lanes 13 to 15) and RSV OLIGO C (lanes 16 to 18), double-stranded oligonucleotides (Fig. 1) that correspond to the EFIII binding site defined in RSV. Lane P, input probe DNA.

different locations within their respective LTRs, may interact with the same or a similar protein(s).

To investigate this question, we tested excess unlabeled RSV- and ev-derived sequences for their ability to compete for formation of the ev DNA-protein complex. As shown in Fig. 5, the addition of a 50- to 100-fold molar excess of unlabeled DNA identical to the input probe efficiently blocked complex formation (compare lanes 2 and 10 with lanes 3 and 4), while an equivalent amount of nonspecific DNA had no effect (lanes 8 and 9). The addition of the RSV 109-bp 5′ fragment, which does not contain the CArG box within the EFIII binding site, had no effect on the ev DNA-protein complex when added at a 50-fold molar excess (lane 5). A detectable decrease was seen, however, when this fragment was added at a 100-fold molar excess (lane 6), but there was no further decrease when this fragment was added at a 200-fold molar excess (lane 7). This relatively weak but detectable competition for binding may indicate that the 5′-most 109 bp of the RSV LTR contains a protein binding site partially related to the ev CArG binding site (see below). The results obtained when the RSV 69-bp fragment that included the CArG-containing EFIII binding site was added as a competitor are shown in Fig. 5, lanes 11 and 12. The addition of a 50- to 100-fold molar excess of this fragment resulted in a dramatic decrease in complex formation equivalent to that seen when the unlabeled 87-bp ev DNA was used as a competitor, suggesting that the CArG box located in the 3′ portion of the RSV LTR could compete for binding to the ev CArG sequence.

To address this question directly, a double-stranded oligonucleotide corresponding to the RSV EFIII binding site was used as a competitor (oligonucleotide RSV-c). As a control, a double-stranded oligonucleotide corresponding to this same region in the ev LTR, which lacks a CArG motif, was also synthesized and tested in parallel (oligonucleotide ev-c). As shown in Fig. 5, the addition of a 50-fold molar excess of oligonucleotide RSV-c resulted in the virtual elimination of the ev DNA-protein complex (lanes 16 to 18). This inhibition was not due to nonspecific competition by the oligonucleotide, since the same region in the ev LTR, which lacks a CArG motif, had no effect (lanes 13 to 15). We therefore conclude that the ev CArG motif binds a protein(s) that is also efficiently recognized by the RSV EFIII binding site.

RSV contains a second CArG motif at a position corresponding to that of the ev CArG box. After identification of the ev CArG box, we noted that RSV contains a similar motif at the corresponding location when the U3 regions are aligned as shown in Fig. 1. This motif, 5′-CCTTACAAGG-3′, which differs from the consensus CArG motif by substitution of the fourth internal A-T base pair with a C-G base pair, is located between positions −167 and −158 in the RSV

LTR. A suggestion that this second, upstream CArG box might be a binding site for a factor at least related to that which recognized the ev CArG binding site came from the competition experiments shown in Fig. 5, in which it is shown that the 5′-proximal, 109-bp fragment from RSV was able to partially compete for binding to the ev CArG box. EMSA, however, had failed to reveal binding to this upstream CArG motif in RSV when either the 143-bp (defined by primers RSV-a and RSV-c) or the 109-bp (defined by primers RSV-a and RSV-b) U3 fragment was tested. As shown in Fig. 1, the upstream CArG motif in RSV is separated by only 2 bp from sequences involved in EFI binding, as determined by DNase I footprinting (42), raising the possibility that binding to this CArG motif might be blocked by occupancy of the adjacent site. To test this possibility, we generated from RSV a DNA fragment that extended from the 5′ end of the LTR through position −147 (defined by primers RSV-a and RSV-d) and that therefore dissociated sequences required for EFI or FII binding from the upstream CArG motif. As shown in Fig. 6, this shortened, 83-bp fragment (lanes 4 to 6) gave rise to a DNA-protein complex that was not evident with the RSV 109-bp fragment, which contained the adjacent EFI or FII binding site (lanes 1 to 3). A similarly sized DNA-protein complex was also seen when a double-stranded oligonucleotide (oligonucleotide RSV-d; Fig. 1) that included the upstream CArG motif was used (Fig. 6, lanes 7 to 9). As shown in Fig. 7, methylation interference assays conducted with oligonucleotide RSV-d identified as contact points important for protein binding to this upstream CArG motif the same four guanidine residues that were seen previously for the ev CArG binding site (Fig. 4) and the RSV EFIII binding site (3; data not shown), with the possible involvement of an additional guanidine residue upstream of the CArG motif (marked by an asterisk in Fig. 7). Thus, RSV contains two CArG motifs, a feature seen with other, cellular genes (46).

Characterization of an ev 3′-proximal factor binding site. Ryden and Beemon (38) identified by DNase I footprinting one protein binding site in ev LTRs, located between positions −74 and −58 on the coding strand and positions −77 and −57 on the noncoding strand. The protein(s) that bound to this region was detectable in avian liver cell-derived nuclear extracts and was shown to be heat stable. It was further shown that the rat CCAAT/enhancer binding protein, c/EBP, bound the same sequence in an in vitro assay system. Although a binding site for this c/EBP-like protein was not seen in the exogenous virus LTR at the same position, two sites with similar characteristics were described at the 5′-most end of the RSV LTR, suggesting that ev and RSV LTRs may contain binding sites for the same or a similar cellular protein(s) at different locations. Since the region from −77 to −57 of the ev LTR was a potential component of the composite enhancer in the ev-RSV hybrid LTRs previously described (5), it was important to further characterize this site and determine its relationship to the 5′-most RSV enhancer domains. To investigate this question, we generated from the ev LTR a 34-bp fragment that extended from position −90 to an *Nde*I site at position −57 (Fig. 1) and analyzed it for protein binding activity as described above. As shown in Fig. 8A, the ev-derived fragment gave rise to one major DNA-protein complex (lanes 1 and 2) that was also seen when heat-treated extracts were used in EMSA (lanes 3 to 5). This treatment abolished binding of the EFI or FII and EFIII proteins to the RSV LTR (compare lane 7 with lanes 8 to 10).

To verify that this binding activity recognized sequences

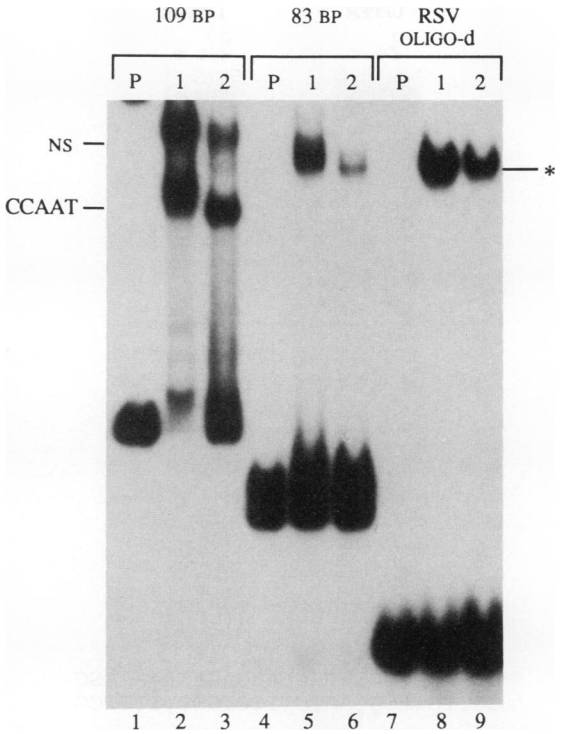

FIG. 6. Identification of a second CArG motif in the RSV LTR. The following U3 fragments from RSV were incubated with 15 μg of protein extracted from S13 nuclei, and DNA-protein complexes were analyzed by EMSA as described in the legend to Fig. 2: lanes 1 to 3, a 109-bp fragment that extended from the 5′ end of the RSV LTR through the EFI or FII binding site (primers RSV-a and RSV-b); lanes 4 to 6, an 83-bp fragment that extended from the 5′ end of the RSV LTR through position −147 and that lacked sequences required for EFI or FII binding (primers RSV-a and RSV-d); and lanes 7 to 9, the RSV-d double-stranded oligonucleotide that included the upstream CArG motif. The band marked with an asterisk is seen only after removal of the RSV EFI or FII binding site. Lanes 1, 4, and 7, labeled P, contained probe DNA not incubated with nuclear extracts. The numbers above each lane indicate the micrograms of salmon sperm DNA added to each sample as a nonspecific competitor. The band labeled NS was a nonspecific DNA-protein complex that was not seen when higher concentrations of salmon sperm competitor DNA, which do not affect binding to the inverted CCAAT motif, were added to the reaction (see the legend to Fig. 2).

within the previously identified c/EBP-like site, we conducted methylation interference assays with DNA labeled on the upper strand. A comparison of cleavage products generated from the free and bound fractions showed the depletion of two guanidine residues at positions −66 and −69 (Fig. 9) in the bound fraction. One adenine residue at position −71 that was a prominent cleavage product in the free fraction was depleted in the bound fraction. The identification by methylation interference assays of adenine residues as contact points for protein binding has been seen in many other cases and suggests that protein interacts with the minor groove of the DNA (44). The location of important contact points for protein binding to this region, as identified here by methylation interference and as previously defined (38) by DNase I footprinting, are shown in Fig. 9B. The region defined by these two techniques maps to the same location, although only a subset of residues within the DNase I-foot-

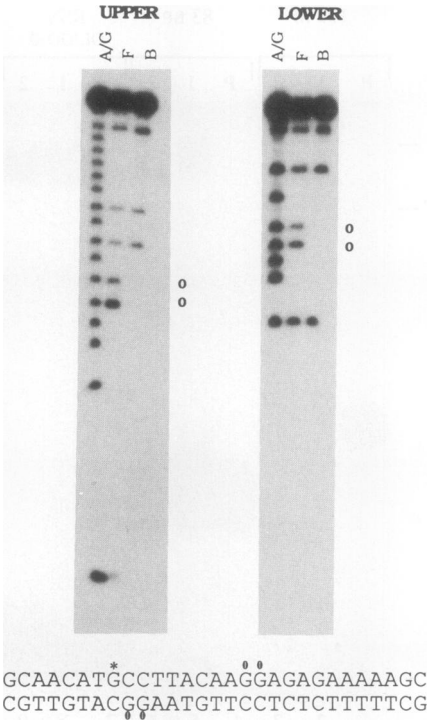

FIG. 7. Methylation interference footprint of the upstream RSV CArG motif. The upper and lower strands of the oligonucleotide RSV-d were labeled, and each was annealed with an excess of unlabeled complementary strand. After gel purification, oligonucleotides were methylated and incubated with S13 nuclear extracts, and the free and bound fractions were analyzed as described in the legend to Fig. 4, except that samples were run on an 18% urea-polyacrylamide gel. The sequence below the gel is the region from −175 to −147 in RSV (oligonucleotide RSV-d). The locations of guanidine residues whose presence was altered in bound fractions are marked with dots in the sequence below the gel. The guanidine residue marked with an asterisk varied in different experiments, and its involvement in protein binding to this region is therefore uncertain.

printed region were also identified by the methylation interference assay.

The finding of a heat-stable protein that bound within the sequence from −77 to −54 in ev LTRs with the extracts described here suggested that, in contrast to previous suggestions (38), this protein is different from a protein(s) that binds to the 5′-proximal enhancer domains of the RSV LTR, which were not detected with our crude extracts. This conclusion was further supported by the binding competition experiments shown in Fig. 8B. For this experiment, the ev U3 fragment from −90 to *Nde*I was end labeled and used in EMSA after the addition of various unlabeled competitor DNAs. The addition of a 100- to 400-fold molar excess of unlabeled DNA identical to the input probe efficiently blocked the formation of the ev DNA-protein complex (compare lane 2 with lanes 3 to 5), while an equivalent amount of nonspecific competitor DNA had no effect (data not shown). The addition of an RSV-derived DNA fragment (lanes 6 to 8) or a double-stranded oligonucleotide (lanes 9 to 11) that contained the 5′-proximal c/EBP-like site previously described (38) also had no effect on the formation of the ev DNA-protein complex. Together, these data indicate that the heat-stable ev DNA binding protein described here is distinct from that which recognizes the 5′ enhancer sequences in the RSV LTR and may therefore identify an ev-specific protein binding site involved in enhancer function.

DISCUSSION

The comparative analysis of factor binding sites in ev and RSV U3 regions described in this report led to the identification of two potential enhancer domains in the portion of the ev U3 region that is able to functionally replace the essential enhancer domains from the RSV LTR (5). A deficiency in ev LTR factor binding which may be of importance in the low-level transcription from intact ev LTRs was also found. Finally, a previously unidentified binding site was identified in the RSV LTR.

One of the protein binding sites identified as a candidate enhancer domain in the ev LTR is a CArG box located between positions −133 and −124. EMSA, methylation interference assays, and binding competition experiments indicated that this motif binds a factor that is the same as or similar to EFIII, which recognizes a CArG motif in the RSV LTR at an alternate location (3). It was also demonstrated that RSV contains a second CArG motif at the same relative location as the ev CArG motif and that this motif also binds a factor related to EFIII and the ev CArG binding factor, as judged by the size of the DNA-protein complex in EMSA and methylation interference footprints. This second CArG motif within the RSV LTR has also recently been detected by Sealy and colleagues (42a). Interestingly, this second, upstream CArG motif identified in the RSV LTR bound protein only when the adjacent CCAAT factor binding site was removed. It is unclear whether this finding is a result of the in vitro binding assay used here or reflects a potential regulatory component within the exogenous virus LTR. The finding of a second CArG motif in the RSV LTR is interesting, since other genes that include CArG motifs not flanked by palindromic sequences often contain multiple copies of this element (45). The presence of only one such sequence in the ev U3 region may therefore result in a deficiency that contributes to the low-level expression from intact ev LTRs. In the case of the ev-RSV hybrid LTRs previously described (5), the upstream CArG motif in the RSV LTR was included in that portion of the enhancer that was deleted during vector construction. The finding of a similar motif in the ev U3 fragment used to replace the upstream RSV enhancer sequences suggests that, in this case, there was a simple exchange of analogous elements to restore a functional enhancer.

Deletion studies have demonstrated that a region that includes the EFIII binding site is required for full activity of the exogenous virus enhancer (7), although the role of CArG motifs in viral transcriptional regulation is unclear. CArG boxes upstream of several skeletal and cardiac actin genes are essential for their cell type-specific expression (26, 29, 52), while the c-*fos* CArG-containing SRE is necessary and sufficient for the rapid induction of c-*fos* transcription following serum stimulation (12, 16, 28, 49, 50) and also for its subsequent repression (34). Although these elements mediate different patterns of expression of their associated genes in vivo, both protein binding and functional studies have demonstrated that these elements are, to some degree, interchangeable (47, 52). For example, CArG motifs without palindromic flanking sequences can confer serum inducibility on heterologous promoters (45, 47, 52), similar to results obtained with the c-*fos* SRE. These elements are not iden-

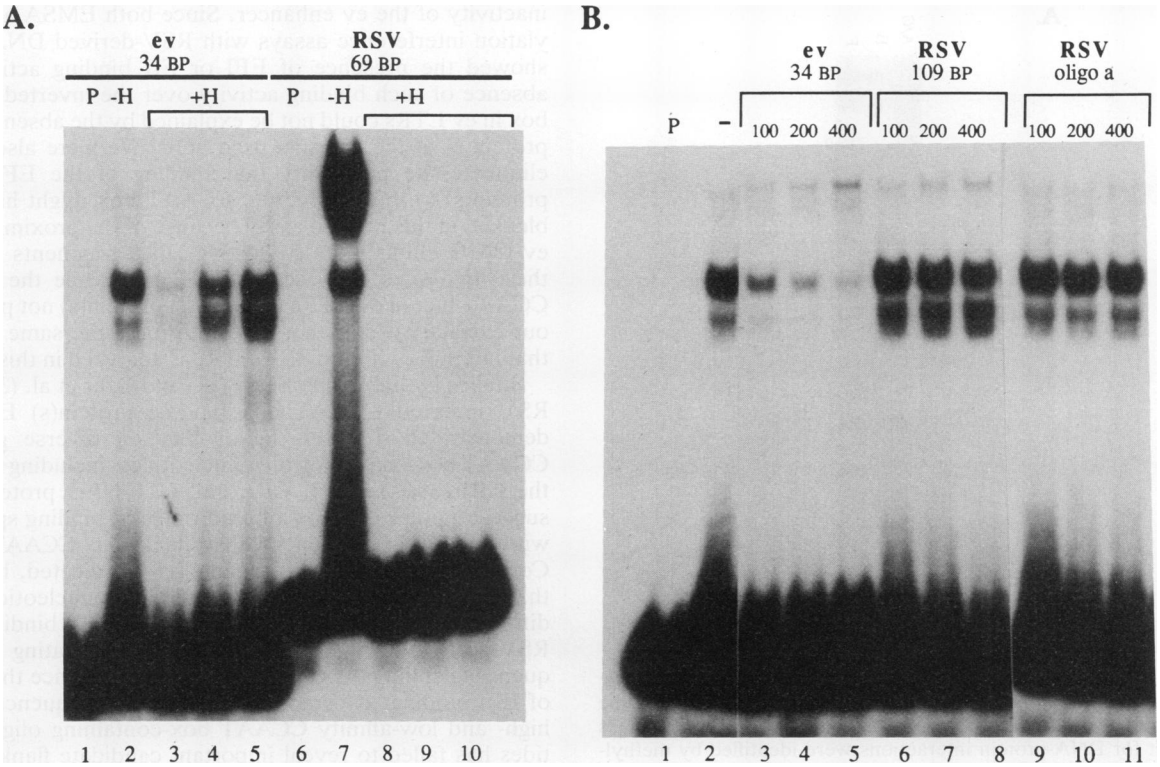

FIG. 8. Detection of a heat-stable binding activity in the 3′ portion of the ev U3 region. (A) Lanes 1 to 5 show the results obtained in EMSA when a 34-bp ev U3 fragment that was defined on its 5′ end by oligonucleotide ev-c and its 3′ end by an *Nde*I site at −57 was incubated with the following: lane 1, no extract; lane 2, 15 μg of S13 nuclear extract; and lanes 3 to 5, increasing amounts of S13 nuclear extract heated at 65°C for 10 min. All binding reaction mixtures contained 2 μg of sonicated salmon sperm DNA. The diffuse lower band was seen with this relatively low concentration of salmon sperm DNA but not with higher concentrations, which do not abolish the more prominent, slower-migrating complex (data not shown). Lanes 6 to 10 represent data obtained with an RSV 69-bp fragment (defined by primers RSV-b and RSV-c) that contained the EFI or FII and EFIII binding sites and that was incubated with the same extracts as those described for lanes 1 to 5. −H, no heat; +H, heat. (B) The 34-bp ev U3 fragment was incubated with 2 μg of sonicated salmon sperm DNA and 15 μg of nuclear protein extracted from S13 cells in the absence (lane 2) and presence (lanes 3 to 11) of a 100- to 400-fold molar excess of the following unlabeled competitor DNAs: ev 34 BP, unlabeled DNA identical to the input probe (lanes 3 to 5); RSV 109 BP, a fragment that was from the 5′ end of the RSV LTR and that contained the c/EBP-like binding site generated with the RSV-a and RSV-b primers; and RSV oligo a, double-stranded oligonucleotide RSV-a (Fig. 1). Lane P represents probe DNA not incubated with nuclear extracts.

tical, however. The c-*fos* SRE is flanked by palindromic sequences that are also involved in the binding of additional factors to the SRE (37, 43). In addition, although some CArG motifs efficiently cross-bind SRF both in vivo and in vitro (4, 27, 47, 52), other CArG motifs interact with distinct cellular proteins (18, 27, 48, 52). Sealy and colleagues have shown that the RSV EFIII binding site also confers a modest serum inducibility on linked promoters, a property shared by the upstream CArG motif (41a). The RSV EFIII binding site (3), the second RSV CArG motif, and the ev CArG motif all share important contact points involved in protein binding, as revealed by methylation interference assays. As originally noted (3), the pattern is the same as that seen with SRF binding to the c-*fos* SRE. Together, these data suggest that retroviral CArG motifs may play a role in the reported serum-inducible transcription of exogenous virus LTRs (10). Studies conducted with rat cells, however, identified the two inverted CCAAT motifs that flank the EFIII-CArG binding site as important determinants of serum-inducible transcription, at least for RSV LTRs with deletions of enhancer sequences upstream of the *Sph*I site at −141. Resolution of a role for CArG motifs in serum-responsive transcription from retroviral LTRs therefore awaits further experimentation.

The second protein binding site in ev LTRs studied here was localized near the extreme 3′ end of the ev U3 fragment that was included in the ev-RSV hybrid LTRs (5) and is therefore a second candidate domain of the defective ev enhancer. Contact points important for protein binding to this site were identified and found to lie within a larger region previously identified by DNase I footprinting to bind a protein(s) from avian liver nuclear extracts (38). Also, in agreement with previous results (38), it was found that heat treatment of nuclear extracts had no effect on protein binding to this site, suggesting that the protein binding activity seen in this study is the same as that previously described with nuclear extracts from avian liver (38).

As assayed in an in vitro binding assay, this region of the ev LTR also binds c/EBP purified from rat liver (38). Additional sites which showed similar binding properties with both purified c/EBP and avian liver nuclear extracts were also identified in other retroviral LTRs, including two sites which were located between positions −225 and −206 and positions −188 to −205 in the RSV U3 region. On the basis of these similarities, it was proposed that the RSV and ev LTRs contained binding sites for the same c/EBP-like protein but that these sites were present at different locations (38). A comparison with other c/EBP binding sites identified

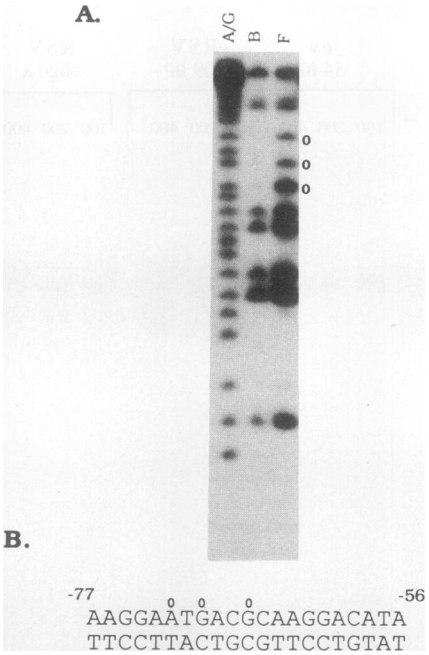

FIG. 9. Methylation interference footprint of the ev heat-stable binding activity. (A) The 34-bp ev U3 fragment described in the legend to Fig. 8 was end labeled on the upper strand, and residues important for DNA-protein interactions were identified by methylation interference assays as described in the legend to Fig. 4. Dots beside the sequence indicate residues specifically depleted in the bound (B) fraction. F, free. (B) Sequence of the region from −77 to −56 of the ev LTR that was involved in protein binding, as defined by DNase I footprinting (38) and methylation interference assays (dots above the sequence).

a consensus motif for this protein: 5′-T(T/G)NNG(C/T)AA (T/G).

In contrast to the studies cited above, we obtained no evidence that the factor that binds to the promoter-proximal site in the ev LTR described here is related to those which bind to the 5′-most enhancer domains of RSV. First, the unfractionated extracts used in these studies failed to reveal any binding activities over the 5′-most enhancer domains of RSV, whereas the protein(s) that bound to the ev 3′-proximal site was readily detectable. Second, DNA fragments that included sequences from the 5′ end of the RSV LTR were unable to compete for binding to the ev 3′-proximal site. One explanation for this discrepancy is that the protein(s) that we have identified here with avian B- and T-cell-derived extracts is different from that detected previously with avian liver cell-derived extracts. Alternatively, these data may indicate that, although the ev and RSV binding proteins show similarities (such as heat stability), they are nonetheless different proteins whose binding sites may each be able to bind c/EBP under in vitro conditions. Resolution of this point will require further study of the relevant protein(s) isolated from various cell types. In any case, the studies described in this report identify a protein factor that is a candidate transcriptional regulatory protein that is involved in ev LTR function and that to date appears to be unrelated to those previously described for exogenous virus LTR functions.

In addition to the identification of two candidate enhancer domains in the ev U3 region, a deficiency in EFI or FII binding to the inverted CCAAT box in ev LTRs was also found and may be responsible, at least in part, for the inactivity of the ev enhancer. Since both EMSA and methylation interference assays with RSV-derived DNA clearly showed the presence of EFI or FII binding activity, the absence of such binding activity over the inverted CCAAT box in ev LTRs could not be explained by the absence of this protein(s) in the extracts used here. We were also able to eliminate the possibility that binding of the EFI or FII protein(s) to the ev inverted CCAAT box might have been blocked in this in vitro assay because of the proximity of the ev CArG binding site by testing DNA fragments in which these two sites were separated. Thus, while the inverted CCAAT box in ev LTRs may bind a protein(s) not present in our extracts, it does not efficiently bind the same factor(s) that interacts with the RSV LTR, as analyzed in this system.

Studies by Faber and Sealy (11) and Ozer et al. (31) of the RSV inverted CCAAT box binding protein(s) EFI have demonstrated that this factor binds a diverse group of CCAAT box-containing oligonucleotides, including those for the CBF, NF-Y, CP1, CP2, and CTF/NF-1 proteins (11), suggesting that EFI has a broad range of binding specificity with regard to sequences that flank the CCAAT motif. Competition binding experiments have indicated, however, that different CCAAT box-containing oligonucleotides show differences in their ability to compete for EFI binding to the RSV-derived CCAAT box sequence, indicating that sequences flanking the CCAAT motif can influence the affinity of EFI binding. A comparison of flanking sequences of the high- and low-affinity CCAAT box-containing oligonucleotides has failed to reveal important candidate flanking residues, so the precise nucleotides required for high-affinity EFI binding have not been identified. It is possible, therefore, that the failure to detect EFI binding over the inverted CCAAT box in ev LTRs reflects differences in sequences flanking this motif. In this event, the maintenance of the conserved CCAAT motif at the same relative location in ev and RSV LTRs is surprising.

Exogenous virus LTRs contain two inverted CCAAT motifs within U3, one centered at position −131 and the other centered at position −67. Deletion and point mutation analyses have demonstrated that both of these motifs are required for high-level transcription from the RSV promoter. Studies by Dutta et al. (10) have shown that these elements may also be required for serum-induced transcription from the RSV LTR in rat cells. Sequence comparisons have shown that ev LTRs lack the promoter-proximal inverted CCAAT motif because of an insertion and base substitutions. The evidence reported here indicates that, although ev LTRs contain the upstream inverted CCAAT motif, this site does not bind the same protein factors as the upstream inverted CCAAT motif in exogenous virus LTRs and suggests that this site may therefore be nonfunctional. Experiments are ongoing to determine whether this apparent deficiency in ev LTRs is the primary defect in the ev enhancer.

The data presented in this report demonstrate that simple sequence comparisons are insufficient to accurately predict the presence of transcriptional control sequences in different enhancer elements. Such an analysis would have led to the conclusion that ev LTRs lack a site for EFIII-CArG binding activity and contain a functional inverted CCAAT box comparable to that found in RSV. It will therefore be interesting to extend these analyses to the 5′ portion of the ev LTRs, which exhibits only limited sequence similarity to the 5′-proximal RSV enhancer domains but has been shown to be functionally equivalent (5), and to compare the functional and structural properties of the protein(s) thus identified.

ACKNOWLEDGMENTS

We thank Kirstin Dohrer for excellent technical assistance, Brian Van Ness and members of his laboratory for helpful discussions, Keats Nelms for advice on EMSA and methylation interference assay, and L. Sealy et al. for communicating unpublished data. Plasmid DNAs were kindly provided by P. Tsichlis (ev-2–RAV-0) and C. Gorman (pRSVNeo).

This work was supported by Public Health Service grant R01-GM41571 from the National Institute of General Medical Sciences and by grants from the Minnesota Medical Foundation and the University of Minnesota Graduate School.

REFERENCES

1. **Akiyama, Y., and S. Kato.** 1974. Two cell lines from lymphomas of Mareks disease. Biken J. **17**:105.
2. **Atchison, M. L.** 1988. Enhancers: mechanisms of action and cell specificity. Annu. Rev. Cell Biol. **4**:127–153.
3. **Boulden, A., and L. Sealy.** 1990. Identification of a third protein factor which binds to the Rous sarcoma virus LTR enhancer: possible homology with the serum response factor. Virology **174**:204–216.
4. **Boxer, L. M., R. Prywes, R. G. Roeder, and L. Kedes.** 1989. The sarcomeric actin CArG-binding factor is indistinguishable from the c-*fos* serum response factor. Mol. Cell. Biol. **9**:515–522.
5. **Conklin, K. F.** 1991. Activation of an endogenous retrovirus enhancer by insertion into a heterologous context. J. Virol. **65**:2525–2532.
6. **Cullen, B. R., K. Raymond, and G. Ju.** 1985. Functional analysis of the transcription control region located within the avian retroviral long terminal repeat. Mol. Cell. Biol. **5**:438–447.
7. **Cullen, B. R., K. Raymond, and G. Ju.** 1985. Transcriptional activity of avian retroviral long terminal repeats directly correlates with enhancer activity. J. Virol. **53**:515–521.
8. **Cullen, B. R., A. M. Skalka, and G. Ju.** 1983. Endogenous avian retroviruses contain deficient promoter and leader sequences. Proc. Natl. Acad. Sci. USA **80**:2946–2950.
9. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
10. **Dutta, A., M. Y. Stoeckle, and H. Hanafusa.** 1990. Serum and v-src increase the level of a CCAAT-binding factor required for transcription from a retroviral long terminal repeat. Genes Dev. **4**:243–254.
11. **Faber, M., and L. Sealy.** 1990. Rous sarcoma virus enhancer factor I is a ubiquitous CCAAT transcription factor highly related to CBF and NF-Y. J. Biol. Chem. **265**:22243–22254.
12. **Gilman, M. Z., R. H. Wilson, and R. A. Weinberg.** 1986. Multiple protein-binding sites in the 5′-flanking region regulate c-*fos* expression. Mol. Cell. Biol. **6**:4305–4316.
13. **Gilmartin, G. M., and J. T. Parsons.** 1983. Identification of transcriptional elements within the long terminal repeat of Rous sarcoma virus. Mol. Cell. Biol. **3**:1834–1845.
14. **Goodwin, G. H.** 1988. Identification of three sequence-specific DNA-binding proteins which interact with the Rous sarcoma virus enhancer and upstream promoter elements. J. Virol. **62**:2186–2190.
15. **Gorman, C. M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. H. 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. USA **79**:6777–6781.
16. **Greenberg, M. E., Z. Siegfried, and E. B. Ziff.** 1987. Mutation of the c-*fos* gene dyad symmetry element inhibits serum inducibility of transcription in vivo and the nuclear regulatory factor binding in vitro. Mol. Cell. Biol. **7**:1217–1225.
17. **Greuel, B. T., L. Sealy, and J. E. Majors.** 1990. Transcriptional activity of the Rous sarcoma virus long terminal repeat correlates with binding of a factor to an upstream CCAAT box *in vitro*. Virology **177**:33–43.
18. **Gustafson, T. A., and L. Kedes.** 1989. Identification of multiple proteins that interact with functional regions of the human cardiac α-actin promoter. Mol. Cell. Biol. **9**:3269–3283.
19. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–480.
20. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of the acceptor site and termini of integrated avian endogenous provirus ev-1: integration creates a 6 bp repeat of host DNA. Cell **23**:155–164.
21. **Hughes, S. H.** 1982. Sequence of the long terminal repeat and adjacent segments of the endogenous avian virus Rous-associated virus 0. J. Virol. **43**:191–200.
22. **Laimins, L. A., P. Tsichlis, and G. Khoury.** 1984. Multiple enhancer domains in the 3′ terminus of the Prague strain of Rous sarcoma virus. Nucleic Acids Res. **12**:6427–6441.
23. **Linial, M., and M. Groudine.** 1985. Transcription of three c-myc exons is enhanced in chicken bursal lymphoma cell lines. Proc. Natl. Acad. Sci. USA **82**:53–57.
24. **Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Cappecchi.** 1983. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell **33**:705–716.
25. **Mitsialis, S. A., J. L. Manley, and R. V. Guntaka.** 1983. Localization of active promoters for eukaryotic RNA polymerase II in the long terminal repeat of avian sarcoma virus DNA. Mol. Cell. Biol. **3**:811–818.
26. **Miwa, T., and L. Kedes.** 1987. Duplicated CArG box domains have positive and mutually dependent regulatory roles in expression of the human α-cardiac actin gene. Mol. Cell. Biol. **7**:2803–2813.
27. **Mohun, T. J., A. E. Chambers, N. Towers, and M. V. Taylor.** 1991. Expression of genes encoding the transcription factor SRF during early development of *Xenopus laevis*: identification of a CArG box-binding activity as SRF. EMBO J. **10**:933–940.
28. **Mohun, T. J., N. Garrett, and R. H. Treisman.** 1987. Xenopus cytoskeletal actin and human c-fos gene promoters share a conserved protein-binding site. EMBO J. **6**:667–673.
29. **Mohun, T. J., M. V. Taylor, N. Garrett, and J. B. Gurdon.** 1989. The CArG promoter sequence is necessary for muscle-specific transcription of the cardiac actin gene in Xenopus embryos. EMBO J. **8**:1153–1161.
30. **Norton, P. A., and J. M. Coffin.** 1987. Characterization of Rous sarcoma virus sequences essential for viral gene expression. J. Virol. **61**:1171–1179.
31. **Ozer, J., M. Faber, R. Chalkley, and L. Sealy.** 1990. Isolation and characterization of a cDNA clone for the CCAAT transcription factor EFIA reveals a novel structural motif. J. Biol. Chem. **265**:22143–22152.
32. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene (c-myc) in bursal lymphomas. Nature (London) **295**:209–214.
33. **Prywes, R., and R. G. Roeder.** 1987. Purification of the c-*fos* enhancer-binding protein. Mol. Cell. Biol. **7**:3482–3489.
34. **Rivera, V. M., M. Sheng, and M. E. Greenberg.** 1990. The inner core of the serum response element mediates both the rapid induction and subsequent repression of c-fos transcription following serum stimulation. Genes Dev. **4**:255–268.
35. **Ruddell, A., M. Linial, W. Schubach, and M. Groudine.** 1988. Lability of leukosis virus enhancer-binding proteins in avian hematopoeitic cells. J. Virol. **62**:2728–2735.
36. **Ruddell, A., M. L. Linial, and M. Groudine.** 1989. Tissue-specific lability and expression of avian leukosis virus long terminal repeat enhancer-binding proteins. Mol. Cell. Biol. **9**:5660–5668.
37. **Ryan, W. A., B. R. Franza, and M. Z. Gilman.** 1989. Two distinct cellular phosphoproteins bind to the c-fos serum response element. EMBO J. **8**:1785–1792.
38. **Ryden, T. A., and K. Beemon.** 1989. Avian retroviral long terminal repeats bind CCAAT/enhancer-binding protein. Mol. Cell. Biol. **9**:1155–1164.
39. **Scholl, D. R., S. Kahn, R. Malavarca, S. Astrin, and A. M. Skalka.** 1983. Nucleotide sequence of the long terminal repeat and flanking cellular sequences of avian endogenous retrovirus *ev*-2: variation in Rous-associated virus-0 expression cannot be explained by differences in primary sequence. J. Virol. **45**:868–871.

40. **Schroter, H., P. E. Shaw, and A. Nordheim.** 1987. Purification of intercalator-released p67, a polypeptide that interacts specifically with the c-fos serum response element. Nucleic Acids Res. **15:**10145–10158.
41. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
42. **Sealy, L., and R. Chalkley.** 1987. At least two nuclear proteins bind specifically to the Rous sarcoma virus long terminal repeat enhancer. Mol. Cell. Biol. **7:**787–798.
42a. **Sealy, L., et al.** Personal communication.
43. **Shaw, P. E., H. Schroter, and A. Nordheim.** 1989. The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. Cell **56:**563–572.
44. **Siebenlist, U., and W. Gilbert.** 1980. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77:**122–126.
45. **Subramaniam, M., L. J. Schmidt, C. E. Crutchfield, and M. J. Getz.** 1989. Negative regulation of serum-responsive elements. Nature (London) **340:**64–66.
46. **Taylor, A., H. P. Erba, G. E. O. Muscat, and L. Kedes.** 1988. Nucleotide sequence and expression of the human skeletal α-actin gene: evolution of functional regulatory domains. Genomics **3:**323–336.
47. **Taylor, M., R. Treisman, N. Garrett, and T. Mohun.** 1989. Muscle-specific (CArG) and serum-responsive (SRE) promoter elements are functionally interchangeable in *Xenopus* embryos and mouse fibroblasts. Development **106:**67–78.
48. **Taylor, M. V.** 1991. A family of muscle gene promoter element (CArG) binding activities in *Xenopus* embryos: CArG/SRE discrimination and distribution during myogenesis. Nucleic Acids Res. **19:**2669–2675.
49. **Treisman, R.** 1985. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5′ element and c-fos 3′ sequences. Cell **42:**889–902.
50. **Treisman, R.** 1986. Identification of a protein binding site that mediates transcriptional response of the c-fos gene to serum factors. Cell **46:**567–574.
51. **Treisman, R.** 1987. Identification and purification of a polypeptide that binds to the c-fos serum response element. EMBO J. **6:**2711–2717.
52. **Walsh, K.** 1989. Cross-binding of factors to functionally different promoter elements in c-*fos* and skeletal actin genes. Mol. Cell. Biol. **9:**2191–2201.
53. **Weber, F., and W. Schaffner.** 1985. Enhancer activity correlates with the oncogenic potential of avian retroviruses. EMBO J. **4:**949–956.
54. **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.