MOLECULAR AND CELLULAR BIOLOGY, Mar. 1989, p. 1155–1164 Vol. 9, No. 3
0270-7306/89/031155-10$02.00/0
Copyright © 1989, American Society for Microbiology

Avian Retroviral Long Terminal Repeats Bind CCAAT/ Enhancer-Binding Protein

THOMAS A. RYDEN AND KAREN BEEMON*
Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218

Received 10 August 1988/Accepted 6 December 1988

DNA-protein interactions involving enhancer and promoter sequences within the U3 regions of several avian retroviral long terminal repeats (LTRs) were studied by DNase I footprinting. The rat CCAAT/enhancer-binding protein, C/EBP, bound to all four viral LTRs examined. The Rous sarcoma virus binding site corresponded closely to the 5′ limit of the LTR enhancer; nucleotides −225 to −188 were protected as a pair of adjacent binding domains. The Fujinami sarcoma virus LTR bound C/EBP at a single site at nucleotides −213 to −195. C/EBP also bound to the promoter region of the enhancerless Rous-associated virus-0 LTR at nucleotides −77 to −57. The avian myeloblastosis virus LTR bound C/EBP at three sites: nucleotides −262 to −246, −154 to −134, and −55 to −39. We have previously observed binding of C/EBP to an enhancer in the *gag* gene of avian retroviruses. A heat-treated nuclear extract from chicken liver bound to all of the same retroviral sequences as did C/EBP. Alignment of the avian retroviral binding sequences with the published binding sites for C/EBP in two CCAAT boxes and in the simian virus 40, polyoma, and murine sarcoma virus enhancers suggested $T^{T}_{G}NNG^{C}_{T}AA^{T}_{G}$ as a consensus sequence for binding of C/EBP. When two bases of this consensus sequence were altered by site-specific mutagenesis of the Rous sarcoma virus LTR, binding of the heat-stable chicken protein was eliminated.

Avian retroviruses have transcriptional regulatory elements within the U3 regions of their long terminal repeats (LTRs) which flank the provirus (39). An enhancer activity has also been detected internally within the *gag* coding sequences (1, 2, 18). A comparison of the nucleotide sequences of different avian retroviruses shows heterogeneity in the U3 LTR region (39). The relative transcriptional activities of several of these LTRs have been compared and found to correlate with the LTR enhancer activities (3). The Rous sarcoma virus (RSV) LTR stimulates a very high level of transcription, while the LTR of Rous-associated virus-0 (RAV-0) (an endogenous chicken virus which lacks an enhancer) is much less active (3). By comparison, the Fujinami sarcoma virus (FSV) LTR has an intermediate level of transcriptional activity (3).

The boundaries of the RSV LTR enhancer have been localized by deletion mutagenesis (4, 10, 21, 25, 28, 37). One essential enhancer domain lies between the 5′ end of the LTR and nucleotide −137 (numbering with the start of transcription at +1), while a second enhancer domain lies between nucleotides −137 and −54 (21). The FSV LTR enhancer has not been localized to this extent. These LTR enhancer sequences have been noted to contain several similarities (3, 21, 25) to the simian virus 40 (SV40) enhancer core sequence ($GTGG^{AAA}_{TTT}G$) (19, 38), the most critical motif in that enhancer (14, 40). The degree of similarity between these retroviral motifs and the SV40 core sequence correlates well with their relative enhancer activity (3).

Enhancer activity appears to require binding of *trans*-acting proteins to enhancer DNA (reviewed in reference 26). Several such proteins have been identified and isolated, and in some cases their genes have been cloned (reviewed in reference 16). One well-characterized SV40 enhancer core-binding protein whose gene has recently been sequenced is the CCAAT/enhancer-binding protein C/EBP (formerly called EBP20) (22). We have previously shown that C/EBP binds to the 5′ region of the enhancer within the *gag* gene of avian retroviruses (2).

To test whether C/EBP would bind to enhancer core sequences or to CCAAT boxes in the LTRs of avian retroviruses, we carried out DNase I footprinting experiments. We found in RSV that, rather than binding to any of the previously noted core motifs, two molecules of C/EBP bound to the 5′ end of the RSV LTR enhancer. The 5′ region of the FSV LTR also bound C/EBP, whereas the RAV-0 promoter bound C/EBP at a putative CCAAT box. To determine whether chickens have a factor analogous to C/EBP, we used partially purified (heat-treated) chicken liver nuclear extracts (htCLNE) in further footprinting experiments. These extracts gave footprints identical to those of C/EBP on the retroviral LTR sequences. Aligning the binding sites for C/EBP allowed us to generate a consensus DNA binding sequence: $T^{T}_{G}NNG^{C}_{T}AA^{T}_{G}$. To test the significance of this consensus sequence, we made a 2-base mutation within it in one of the RSV LTR binding sites. This mutation prevented htCLNE binding over the altered consensus site, but did not affect the adjacent normal binding sites.

MATERIALS AND METHODS

DNA fragments for footprinting. The source plasmid for the Schmidt-Ruppin D (SR-D) RSV LTR probes was pRSV cat, which contains the 3′ LTR of an integrated RSV provirus (9). 5′-end-labeled probes were generated by cutting at either the *Eco*RI or *Hin*dIII site (noncoding strand label) or *Oxa*NI site (coding strand label), treating with calf intestinal phosphatase, and 5′ end labeling with T4 polynucleotide kinase and [γ-^{32}P]ATP. Probes were isolated by restricting at either the *Nde*I (for the *Eco*RI or *Hin*dIII labeling) or the *Hin*dIII (for the *Oxa*NI labeling) site and size fractionating on a 1% agarose gel. Fragments were electrophoresed onto NA45-DEAE membranes (Schleicher & Schuell, Inc.) and eluted according to the manufacturer.

FSV footprinting probes were similarly obtained from

* Corresponding author.

plasmid pBRF16 (35). Probes were 5′ end labeled at the *Eco*RV (noncoding strand) or *Oxa*NI (coding strand) site and excised from the plasmid by a second cut with *Oxa*NI or *Bst*EII, respectively.

The RAV-0 probes were prepared from a plasmid of infectious RAV-0 (36), labeled at either the *Bst*EII or the *Oxa*NI site, and excised by a second cut with *Oxa*NI or *Bst*EII, respectively.

The probes for avian myeloblastosis virus (AMV) sequences were from a plasmid carrying the 3′ terminus of an integrated provirus, a gift from the laboratory of T. Papas (31). The plasmid was cut with either *Xho*I (noncoding label) or *Hin*dIII (coding strand label), labeled, and recut with *Hin*dIII or *Xho*I, respectively.

Nuclear extracts. Fresh livers from White Rock chickens were a gift from Dover Poultry, Baltimore, Md. The preparation of CLNE has been described previously (2). Protein precipitates were brought to a final concentration of 3 mg/ml. C/EBP was a gift from S. McKnight, P. Johnson, and B. Graves. The 130-fold-purified sample was described previously as an FPLC mono S fraction containing EBP20 (15). This fraction was at a concentration of 20 μg/ml.

DNase I footprinting. The procedure for DNase I footprinting has been described previously (2). Each footprint reaction contained 2 ng of labeled probe. DNase I was from Worthington Diagnostics. Reaction products were electrophoresed on 8% polyacrylamide sequencing gels.

Site-specific mutagenesis. Mutagenesis of the Prague C (Pr-C) RSV C/EBP binding site was by the method of Kunkel et al. (20). The template used for mutagenesis (previously described in reference 17) contains an RSV *Sac*I fragment (RSV nucleotides 6865 to 255) from plasmid pATV8K, inserted into the M13 vector mp18. The mutagenic oligonucleotide 5′ GGAGTA<u>G</u>A<u>G</u>CATAAGAC 3′ is complementary to Pr-C RSV nucleotides −212 to −228 (numbered relative to the start of transcription). The underlined sites designate the mutations. Mutants were sequenced according to Sanger et al. (32), using an oligonucleotide complementary to nucleotides −140 to −125.

Densitometry. The regions containing and flanking the footprint domains were scanned using an LKB 2202 UltroScan laser densitometer and 2220 recording integrator. Regions adjacent to the footprints (nonhypersensitive regions) were used to standardize the lanes. The total footprint region was treated as a unit, and the decrease in band intensity was measured between zero and the maximum protein concentration.

RESULTS

C/EBP binding to the RSV LTR enhancer. We were interested in identifying *trans*-acting factors which bind to enhancer and promoter elements of avian retroviruses. Our approach was to assay the binding of proteins by DNase I footprinting, using initially crude nuclear extracts and subsequently more defined protein fractions. Chicken liver was chosen as an abundant source of nuclear proteins. We first footprinted a fragment of SR-D RSV LTR which contained previously defined enhancer regions. When proteins from a crude CLNE were used, protection occurred over nearly the entire U3 region of the LTR, as well as some upstream sequences (data not shown).

Several sites within the RSV U3 region have been noted to contain sequence similarities to the SV40 enhancer core (3, 21, 25). Therefore, we next tested binding of C/EBP, a protein purified from rat liver on the basis of its binding to the SV40 enhancer core sequence and CCAAT box promoter elements (15). We obtained from Johnson et al. a 130-fold-purified protein fraction containing C/EBP (15). An SR-D RSV fragment, labeled at the *Hin*dIII site and incubated with increasing amounts of C/EBP, generated a footprint on the noncoding strand of SR-D RSV between nucleotides −225 and −188 (Fig. 1A). The DNA sequence of this region (data not shown) was identical to that of the SR-A strain of RSV (5). Deletion studies (4, 10, 21, 25, 28, 37) have shown that this protected region is a critical portion of the RSV LTR enhancer and lies within the central B enhancer domain of Laimins et al. (21).

The protected sequence, **TAGTCTTA**<u>**TGCAA**</u>T<u>AC</u>TCTT GT<u>**AGTCTTGCA**</u>ACATGGT, comprises a hyphenated direct repeat (bold), originally noticed by Tsichlis et al. (36), or alternatively a palindrome (underlined). Hypersensitivity is indicated by a ↓ on the coding strand an a ↑ on the noncoding strand. Since the protected region is about twice the size of the C/EBP binding sites on SV40, Moloney murine sarcoma virus (MSV), and polyoma enhancers (15) and is punctuated by a median hypersensitive site, we propose that it consists of a pair of tandem binding sites. The protected region did not correspond to any of the predicted enhancer core motifs and indeed showed no similarity to the SV40 enhancer core sequence. There are, however, a pair of GCAA sequences which we think are important for the binding of C/EBP (see below), since C/EBP has an affinity for CCAAT sequences and binds even more tightly to GCAAT (11).

Protein similar to C/EBP in chicken liver nuclei. The RSV LTR enhancer is active in many cell types (7, 9, 25, 29, 37); however, its highest activity is in avian cells, the natural host of this virus (9, 21, 37). Therefore, it was important to determine whether avian cells contained a protein related to C/EBP. Johnson et al. (15) have shown a 10- to 20-fold purification of C/EBP by a simple heat treatment step. We subjected CLNE to such a treatment. A crude CLNE sample was incubated at 68°C for 10 min and clarified by centrifugation to generate htCLNE. This resulted in a 10-fold reduction in the amount of soluble protein.

htCLNE generated a footprint on the noncoding strand of the RSV LTR (Fig. 1B) identical to that obtained with the purified rat C/EBP (Fig. 1A). Both the footprint limits (nucleotides −225 to −188) and the median hypersensitive site were the same. Therefore, htCLNE contained a protein indistinguishable from C/EBP in both its heat stability and its specificity of binding to the RSV LTR. The volume of htCLNE supernatant required to generate this footprint was identical to that of CLNE needed to protect this region; there was no apparent loss of activity upon heating, and crude extracts probably contain only a single protein binding to this region. htCLNE was also used to footprint the coding strand of this LTR (labeled at the *Oxa*NI site). The protected region on this strand extended from nucleotides −223 to −188, with a median hypersensitive site at nucleotide −203 (Fig. 1B).

C/EBP and the FSV LTR. While the sizes of the FSV and RSV LTRs are similar, their sequences are not closely related. The FSV LTR has a nucleotide sequence more similar to that of the endogenous avian virus RAV-0. FSV has an approximately 60-base-pair (bp) duplication and two unique regions of 7 and 8 nucleotides near the 5′ end of the LTR but is otherwise closely related to the RAV-0 LTR (35, 39). The enhancer activity of the FSV LTR is reduced relative to that of RSV (3).

To determine its protein binding pattern, an FSV DNA

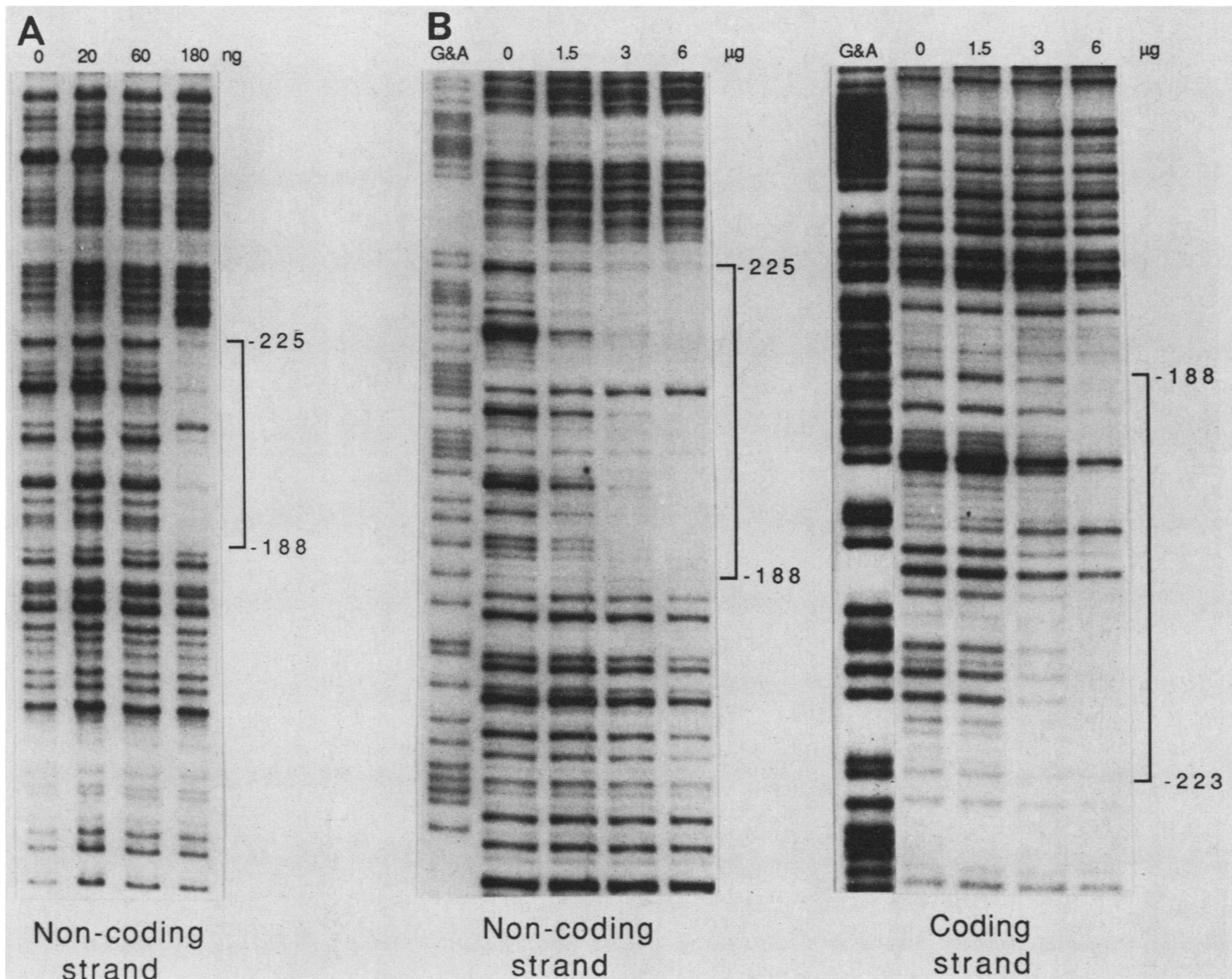

FIG. 1. Footprinting of the SR-D RSV LTR with C/EBP (A) and htCLNE (B). (A) The RSV LTR probe, labeled on the noncoding strand (*Hin*dIII site), was bound with C/EBP (0, 20, 60, and 180 ng) and digested with DNase I. The protected region is bracketed and the endpoints (nucleotides −225 and −188) are indicated. (B) The RSV LTR was labeled on the noncoding (*Eco*RI site) and coding (*Oxa*NI site) strands, bound with htCLNE (0, 1.5, 3, and 6 µg), and digested with DNase I. The protected region is bracketed and the endpoints of the footprints (nucleotides −225 and −188 and −223 and −188) are indicated. Purine cleavage reactions (Maxam and Gilbert [27]) are shown.

fragment was labeled at the *Oxa*NI site just upstream of the LTR, and binding of crude CLNE was assayed on the U3 region. CLNE protected sequences throughout U3, yielding a footprint as complex as in the analogous experiment with the RSV LTR (data not shown).

We again turned to footprinting with C/EBP fractions since a moderate similarity to the SV40 enhancer core sequence had been noted at nucleotides −124 to −114 (3). Rather than protecting this sequence, C/EBP was found to protect sequences between nucleotides −213 and −196 on the coding strand of the FSV LTR (Fig. 2A). This footprint was weaker than that with the RSV LTR (Fig. 1A). Densitometric analysis of the footprint showed that at the maximum amount of C/EBP used (180 ng), DNase I cutting intensity was reduced by threefold within the bracketed region (Fig. 2A). By comparison, the same amount of protein resulted in a greater than 10-fold protection of the RSV LTR (Fig. 1A). The FSV protected region did not contain any sequence similarity to the SV40 enhancer core sequence, but did contain a similarity to the RSV C/EBP binding site. The protected sequence on the noncoding strand, CTGTCTTAT GTAATGATGA, has significant similarity (underlined) to the more upstream half of the RSV footprint, TAGTCTTAT GCAATACTCTT. The length of the protected region on the FSV LTR is approximately equivalent to the length of the C/EBP protected region on SV40, polyoma, and MSV enhancers (15) and is about half that of the RSV LTR binding site. The position of the C/EBP footprint in the FSV LTR was near the 5′ border of the LTR, within one of the few FSV U3 regions which has diverged significantly from that of the RAV-0 LTR (35, 39).

To further test the relatedness of the heat-stable chicken binding protein and C/EBP, we bound htCLNE to the FSV LTR. The footprint obtained with the coding strand (Fig. 2B) appeared to be identical to that obtained with C/EBP (Fig. 2A). We further determined that the FSV noncoding strand was protected by htCLNE at nucleotides −214 to −197 (Fig. 2B).

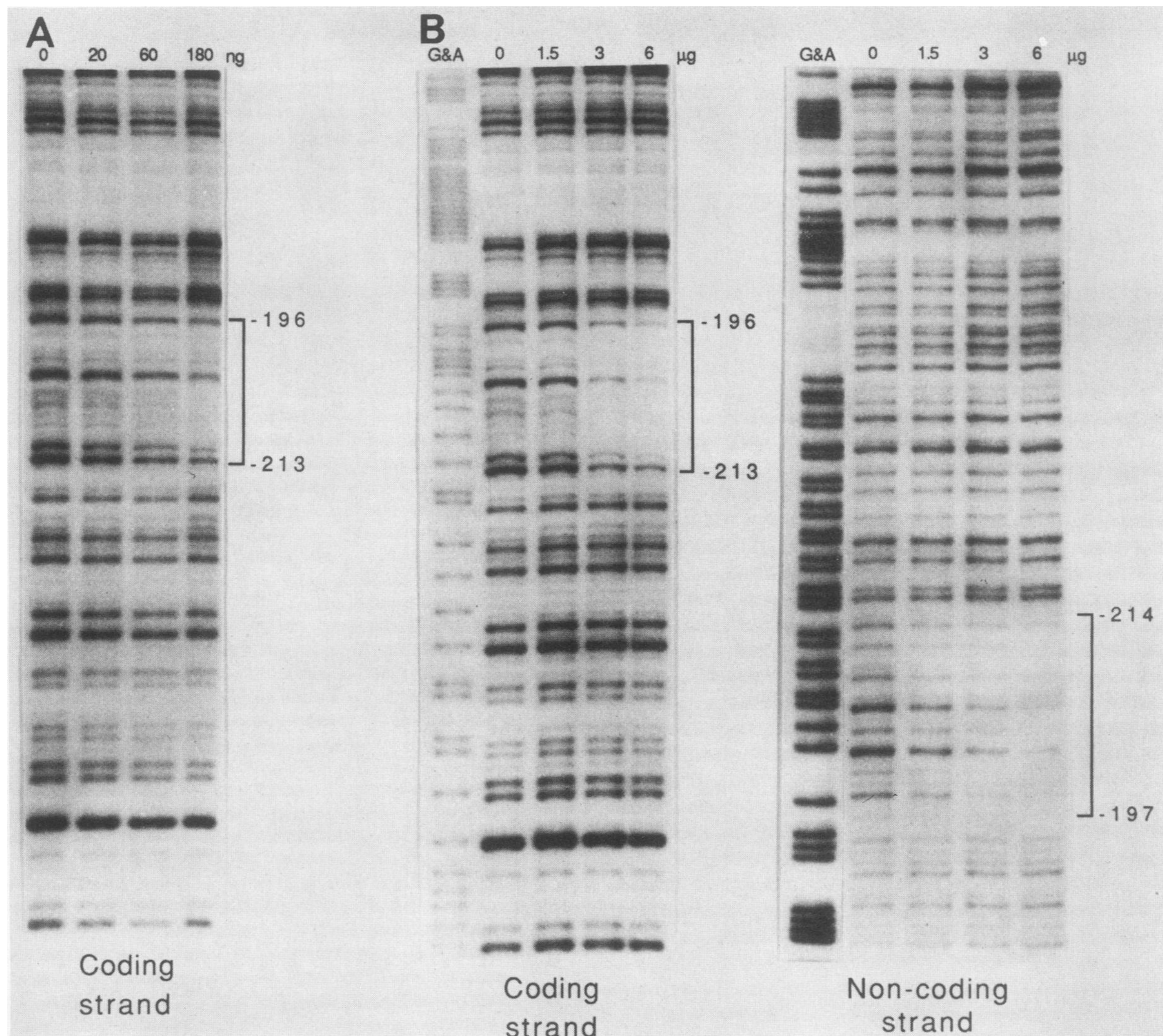

FIG. 2. Footprinting of the FSV LTR with C/EBP (A) and htCLNE (B). (A) The FSV LTR was labeled on the coding strand (*Oxa*NI site), bound with C/EBP (0, 20, 60, and 180 ng), and digested with DNase I. The footprint domain is indicated by the bracket, with endpoints at nucleotides −196 and −213. (B) The FSV coding and noncoding strands were labeled (the *Oxa*NI and *Eco*RV sites, respectively), bound with htCLNE (0, 1.5, 3, and 6 μg), and digested with DNase I. The protected regions (nucleotides −213 and −196 and −214 to −197) are indicated by brackets to the right. Purine cleavage sequencing reactions (27) were run in parallel. The plasmid from which these probes were obtained contains two adjacent LTRs flanked by the *Oxa*NI and *Bst*EII sites. The coding strand probe, limited by these two sites, contains two complete LTRs and can be compared directly with the RSV tandem C/EBP footprints. The noncoding strand probe, being labeled at a site internal to the LTR (*Eco*RV), bears only a single C/EBP binding site and therefore contains half the number of binding sites present in the FSV coding and RSV footprints.

RAV-0 promoter–C/EBP interaction. The RAV-0 LTR lacks enhancer activity (3). However, it has sequences (CGCAAGGACA, nucleotides −67 to −58) loosely conforming to the SV40 enhancer core (3). To determine whether this sequence bound C/EBP, the RAV-0 DNA was labeled at the *Oxa*NI site and assayed by DNase I footprinting. Figure 3A shows that C/EBP did protect a region of the RAV-0 coding strand from nucleotides −74 to −58, AAG GAATGACGCAAGGACATA. Thus, the protected region included the weak core similarity. Because of its position in the RAV-0 promoter and the lack of RAV-0 LTR enhancer activity (3), we propose that this binding is due to the CCAAT box binding activity of C/EBP (15).

When the htCLNE fractions were bound to RAV-0 (Fig. 3B), the pattern obtained on the coding strand was again identical to that with C/EBP (Fig. 3A). On the noncoding strand of the RAV-0 LTR, nucleotides −77 to −57 were protected by htCLNE.

AMV LTR and its C/EBP binding sites. The consistent appearance of C/EBP binding sites in these three avian retroviral LTRs prompted us to search other retroviruses for similar sequences. An 18-bp identity was found between the RAV-0 binding sequence and the sequence at nucleotides −154 to −134 in the U3 region of AMV. The U3 sequences of AMV have diverged significantly from those of other avian retroviruses (31). Its transcriptional activity has not been compared with the other viruses that we studied.

When the DNA was assayed by DNase I footprinting with C/EBP, we indeed observed a footprint over the sequence at nucleotides −154 to −134 (Fig. 4A). Additionally, two more binding sites were found. One was between nucleotides −262 and −246 and the other was between nucleotides −55

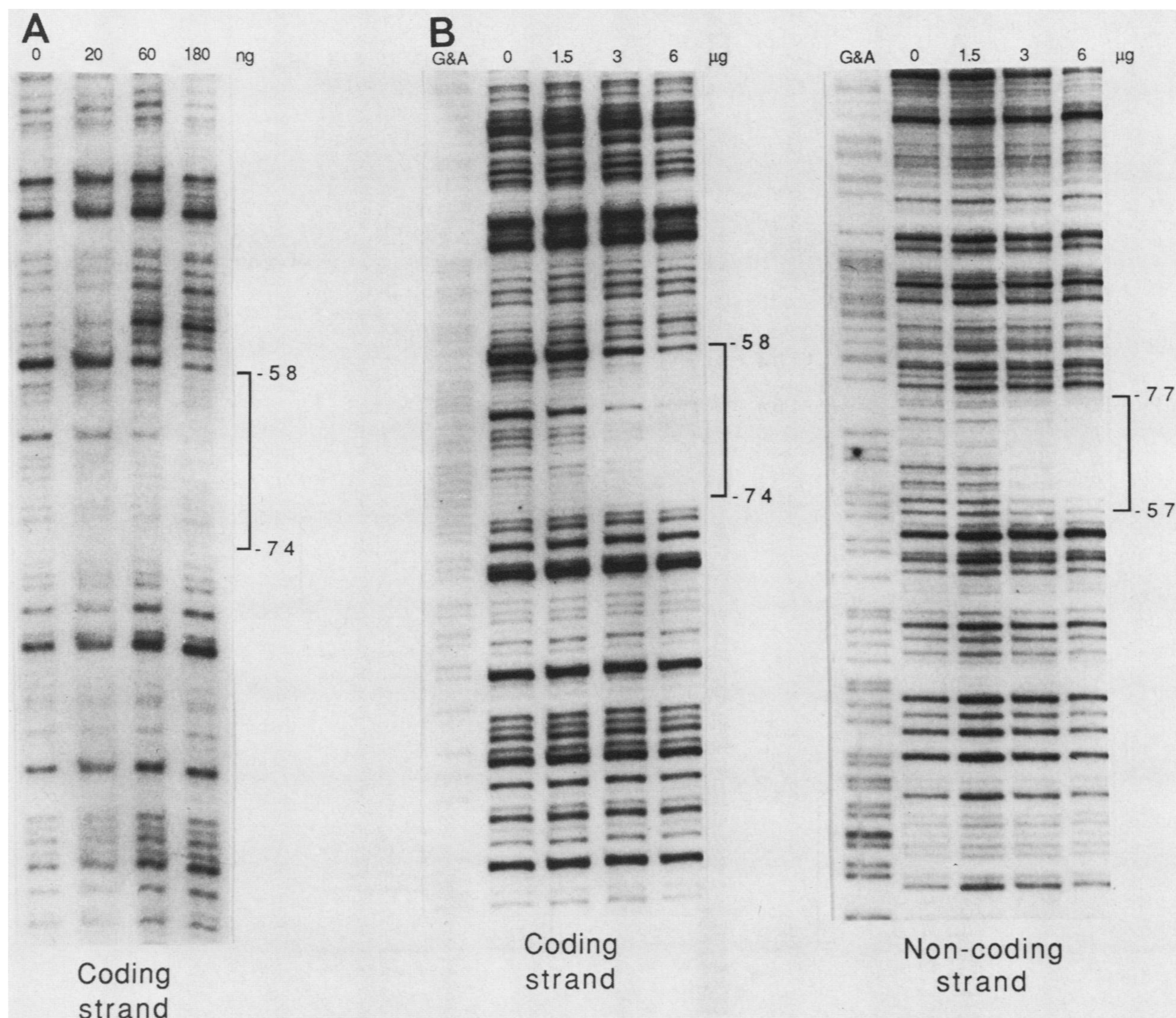

FIG. 3. RAV-0 footprints with C/EBP (A) and htCLNE (B). (A) C/EBP at 0-, 20-, 60-, and 180-ng amounts, bound to RAV-0 LTR (labeled on the coding strand at the *Oxa*NI site) and digested with DNase I. The protected region (nucleotides −74 to −58) is indicated by the bracket at the right. (B) RAV-0 sequences, labeled on the coding (*Oxa*NI site) and noncoding (*Bst*EII site) strands, were bound with 0, 1.5, 3, and 6 μg of htCLNE and digested with DNase I. Protected regions (nucleotides −74 to −58 and −77 to −57) are indicated at the right. Purine cleavages (27) were run in parallel. The plasmid from which these probes were obtained has two tandem LTRs flanked by the *Oxa*NI and *Bst*EII sites. These footprints can be compared directly with those of RSV (tandem C/EBP footprints) and the FSV coding strand.

and −39 (Fig. 4A). None of these sequences appeared to be related to the SV40 enhancer core sequence. We noted two interesting features of the more upstream binding site. First, the sequence is an imperfect palindrome: TGTATATTAC CAAATAAGGGAA. Second, as noted by Rushlow et al. (31), 11 bp of this sequence are directly repeated downstream (nucleotides −229 to −219; ACCAAATAAGG). This latter sequence, however, did not bind C/EBP. An imperfect palindrome and a direct repeat downstream were also found in the most upstream RSV C/EBP binding site. These AMV and RSV protected sequences share a match of 10 of 17 bp (see Fig. 6). The more downstream AMV site (nucleotides −55 to −39) was positioned between the putative CCAAT box and the TATA box, as identified by Rushlow et al. (31).

htCLNE was also tested for binding to the AMV LTR. The footprint obtained on the noncoding strand (Fig. 4B) contained the same three sites that were protected with C/EBP (Fig. 4A). On the coding strand, htCLNE protected similar sequences at nucleotides −260 and −241, −152 to −136, and −57 to −37 (Fig. 4B).

C/EBP binding sites in avian retroviral LTRs. Figure 5 presents a summary of binding sites for C/EBP and htCLNE in the U3 LTR region of four avian retroviruses. It is interesting that the three viruses having an active enhancer (RSV, FSV, and AMV) all bound C/EBP near the extreme 5′ end of their LTRs. In the case of RSV, this is known to be an essential enhancer region (4, 21, 28). C/EBP failed to bind the enhancerless LTR of RAV-0 at an upstream site, but did bind at a position appropriate for a promoter CCAAT box (6). There is a match of 11 of 16 bp between the sequence at the center of the RAV-0 binding site and the C/EBP binding site in the CCAAT box of the herpes simplex virus (HSV) thymidine kinase (*tk*) gene (11). The transcriptional activity of the RAV-0 LTR is comparable to that of the RSV LTR

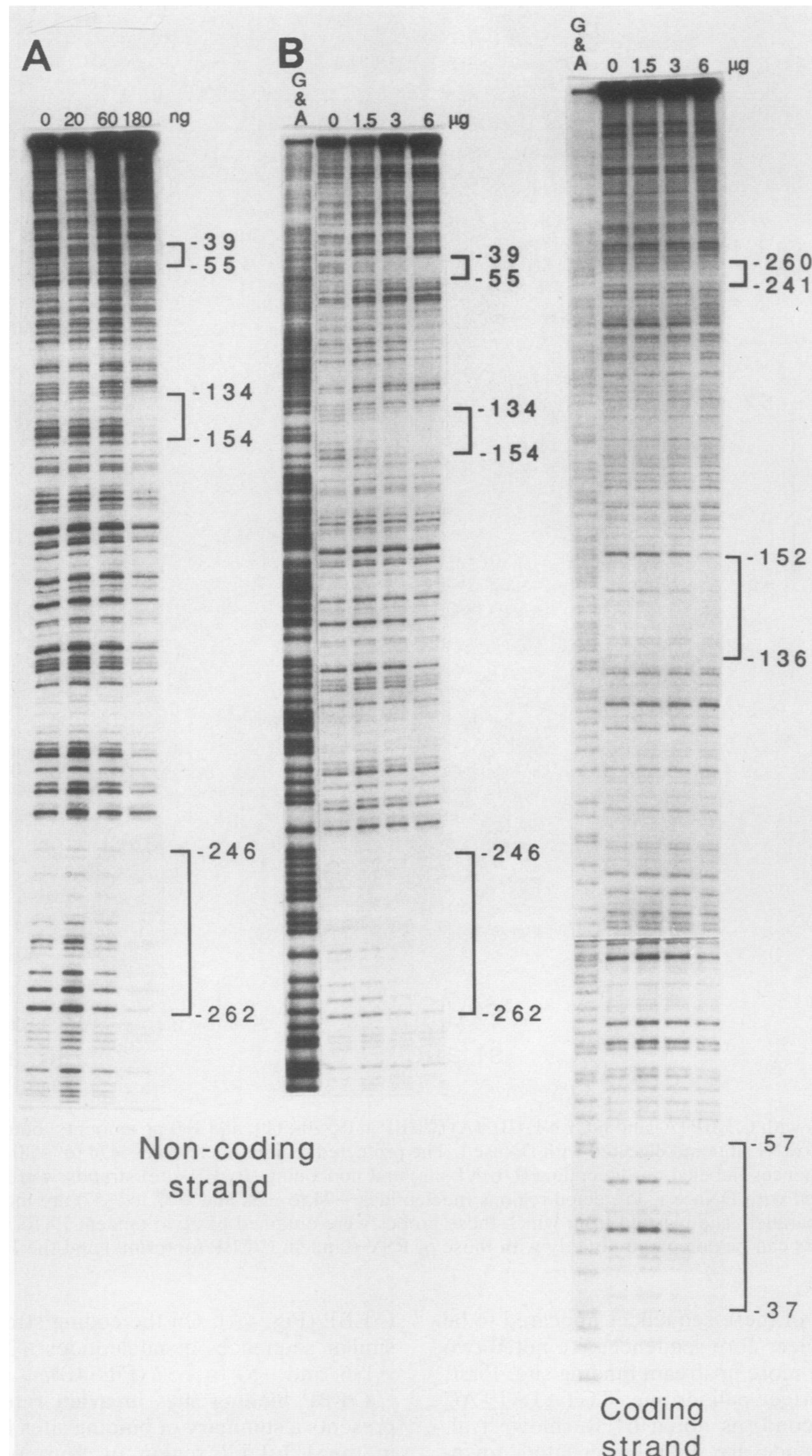

FIG. 4. AMV footprints with C/EBP (A) and htCLNE (B). (A) AMV LTR sequences, labeled on the noncoding strand (*Xho*I site), were allowed to bind C/EBP (0, 20, 60, and 180 ng) and were then cleaved with DNase I. The protected regions and their endpoints are shown on the right. (B) AMV LTR sequences were labeled on the coding (*Xho*I site) and noncoding (*Hin*dIII site) strands, bound with htCLNE at 0, 1.5, 3, and 6 µg, and digested with DNase I. The footprinted domains are indicated to the right. The G+A lanes are the Maxam and Gilbert (27) purine cleavage reactions.

without an enhancer (4); however, we have not observed binding of C/EBP to the RSV promoter. Thus, RSV and FSV contrast with MSV, which binds C/EBP both within its enhancer and at a CCAAT box sequence (11, 15). AMV may be more analogous in this respect to MSV. In addition to the binding site near the 5′ boundary of the LTR, the AMV LTR also bound C/EBP at two other sites. This pattern of three widely spaced binding sites is similar to the C/EBP binding

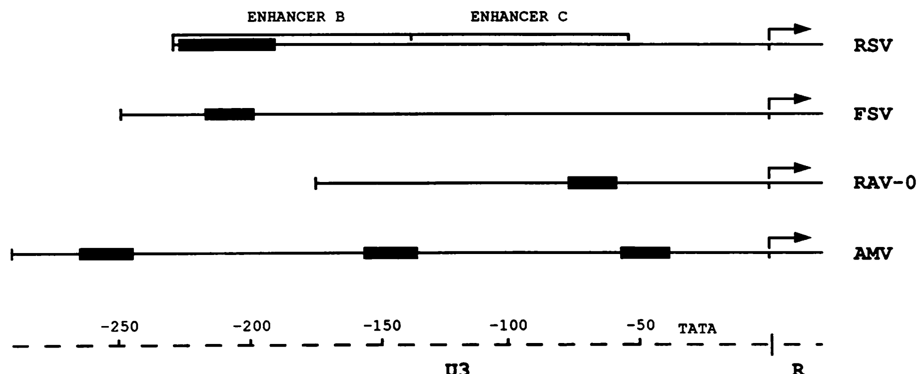

FIG. 5. Summary of the C/EBP binding sites on the avian retroviral LTRs. The U3 regions of the SR-D RSV, FSV, RAV-0, and AMV LTRs are diagrammed and their C/EBP binding domains are indicated with solid boxes. The start sites of transcription (arrows) serve as the points of alignment. For RSV the enhancer domains (B and C) identified by Laimins et al. (21) are indicated with brackets. The broken line at the bottom shows the position of the R region, the TATA boxes, and U3 nucleotides (numbered relative to the transcriptional start sites at +1).

pattern on the mouse albumin promoter, in which three sites are approximately equally spaced over 132 bp (24).

C/EBP and htCLNE consensus binding site: $T^{T}_{G}NNG^{C}_{T}AA^{T}_{G}$. To look for sequence specificity in binding of C/EBP and htCLNE, we aligned the sequences of the seven binding sites that we observed with the binding sequences for C/EBP that have been previously identified in the enhancers of SV40, polyoma, MSV, and the FSV *gag* gene and in the CCAAT boxes of MSV and of the HSV *tk* gene (Fig. 6). The consensus generated from the 14 sequences ($T^{T}_{G}NNG^{C}_{T}AA^{T}_{G}$) was completely conserved at the G and A sites and in 13 of the 14 sequences at the most upstream T. It is clear, however, that C/EBP will bind to variations of this sequence. Our alignment used the tight-binding GCAAT mutants of MSV and HSV *tk*, while C/EBP also binds with lower affinity to wild-type CCAAT sequences in these promoters (11).

In the four LTRs assayed, the consensus sequence appeared seven times, and C/EBP bound to each such sequence. Every footprint obtained with htCLNE was identical to footprints generated with C/EBP. No additional binding sites were detected with htCLNE. The simple heat treatment, therefore, yielded a chicken extract of considerable utility.

Requirement of the consensus sequence for protein binding. Site-specific mutagenesis was carried out to assess the effects of alterations in the consensus sequence on the binding of htCLNE. The Pr-C strain of RSV was chosen instead of the SR-D strain that we had been using previously since it is completely sequenced (33) and we wanted to eventually assay the effects of these mutations in the virus. The Pr-C LTR is very similar to the SR-D LTR except for a few nucleotide substitutions and a 5-bp insertion. These five nucleotides are inserted in the 3′ region of the sequences corresponding to the more downstream SR-D C/EBP binding site. This does not affect the consensus sequence of this binding site; however, it does introduce a third copy of this consensus motif (Fig. 7B). The htCLNE footprint on the Pr-C RSV LTR showed, in addition to the region protected in SR-D RSV, an extension downstream weakly covering

Site	Sequence
RSV (-225/-206)	TAGTC**TT**AT**GCAAT**ACTCTT
RSV (-188/-205)	ACCA**TG**TT**GCAAG**ACTAC
FSV (-196/-214)	CTGTC**TT**AT**GTAAT**GATGA
RAV-0 (-77/-57)	AAGGAA**TG**AC**GCAAG**GACATA
AMV (-241/-262)	TTCCCTTA**TT**TG**GTAAT**ATACA
AMV (-134/-154)	TGAGGAA**TG**AC**GCAAG**GACAG
AMV (-57/-37)	CTAAGGAG**TT**GT**GTAA**CCCAC
SV40 (255/236)	GTTAGGG**TG**TG**G**A**AAG**TCCC
Polyoma (5212/5232)	AGTGTGG**TT**TT**GCAAG**AGGAA
MSV (-236/-215)	AGGATATC**TG**TG**GTAAG**CAGTT
MSV CCAAT (-98/-71)	TACCTTATTTGAAC**T**AA**GCAAT**CAGTTC
HSV tk CCAAT (-67/-95)	GCGTGTTCGAA**TT**CG**GCAAT**GACAAGAC
FSV *gag* (848/819)	TAGCTG**TT**CC**GCAAT**GATAGCAGGATGTG
FSV *gag* (872/855)	AGGCTG**TG**GC**GCAAT**TAC
CONSENSUS	$\mathbf{T^{T}_{G}}NN\mathbf{G^{C}_{T}AA^{T}_{G}}$

FIG. 6. Summary and consensus of the C/EBP binding sequences. SR-D RSV, FSV, RAV-0, and AMV LTR sites are from the data in this report. The two RSV sequences indicate the two binding domains which were separated by the hypersensitive site on the noncoding strand. SV40, MSV, and polyoma sequences are from their enhancers, as reported by Johnson et al. (15). MSV and HSV *tk* CCAAT sequences are from their respective promoters, using the tight-binding mutants reported previously by Graves et al. (11). The FSV *gag* enhancer sequences are from a previous report by Carlberg et al. (2). Nucleotide numbers for these sequences are those reported in the respective references or, in the case of the retroviruses, relative to the transcriptional start sites at +1. To facilitate alignment, several sequences are inverted from their endogenous orientation. Nucleotides of the binding sites shown in boldface are those which are in common with the derived consensus.

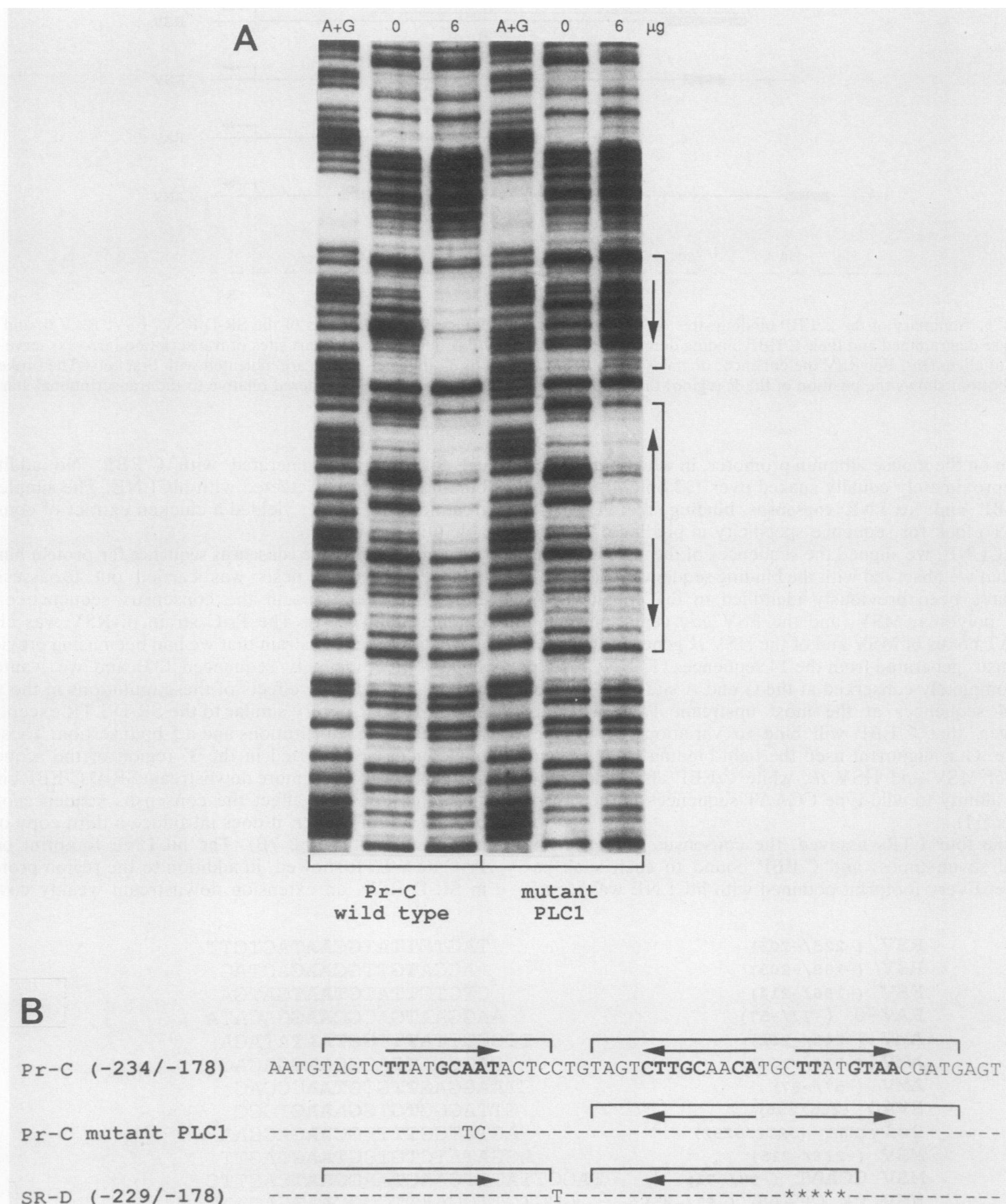

FIG. 7. Effect of mutation of the consensus sequence on the binding of htCLNE. (A) Mutagenesis of the Pr-C RSV binding site was performed as described in Materials and Methods. The AA nucleotides of the consensus at nucleotides −221 and −220 were mutated to TC to generate mutant PLC1. Wild-type and PLC1 DNAs were labeled at the *Eco*RI site and footprinted with 6 µg of htCLNE, as indicated at the top of the gel. Purine cleavage reactions were performed according to Maxam and Gilbert (27). Brackets to the right of the footprint indicate the protected domains on the wild-type sequence. The three consensus sequences are indicated with arrows. Mutant PLC1 lacks the upstream consensus sequence and did not footprint at this site. (B) Summary of nucleotide changes made in mutant PLC1 and comparison of the sequence differences between the Pr-C and SR-D strains of RSV. - - -, Nucleotides identical to the Pr-C RSV sequence shown (inserted nucleotides are those which are altered in the mutant or in SR-D RSV); *, nucleotides which are deleted in the SR-D strain. The htCLNE footprint domains (brackets) and the C/EBP binding site consensus sequences (arrows) are summarized above the sequences. Boldface letters indicate consensus nucleotides. Nucleotide numbers for Pr-C and SR-D RSV are given relative to the start site of transcription at +1.

this third motif (Fig. 7). The third sequence deviates at a single nucleotide from the consensus, which may explain its reduced affinity for htCLNE.

To assess the effects of mutating the consensus, the upstream binding site in the Pr-C RSV LTR was changed at nucleotides −221 and −220 from AA to TC (mutant PLC1) (Fig. 7B). Figure 7A shows footprints with htCLNE on the wild-type and mutated LTR sequences. The two-base substitution appeared to completely block binding of the heat-stable chicken protein at the upstream binding site. However, binding of htCLNE to the more downstream sites was not affected.

DISCUSSION

We have described binding of C/EBP to the U3 regions of the LTRs of four different avian retroviruses. C/EBP binding sites may be important elements in retroviral enhancers since they have been found near the 5′ LTR border in all three enhancer-containing viruses analyzed (RSV, FSV, and AMV) and were not found in the 5′ end of the enhancerless RAV-0 sequences. Rather, this LTR bound C/EBP at a potential CCAAT box in its promoter. We generated a consensus sequence for C/EBP binding: $T_{G}^{T}NNG_{T}^{C}AA_{G}^{T}$. In the four LTRs that we examined, the consensus sequence appeared seven times and binding of C/EBP to this sequence occurred in each case. Furthermore, we have mutated two of the nucleotides of this consensus in the upstream binding site in the Pr-C RSV LTR and found that this was sufficient to eliminate binding. Such mutants should be useful for studying the role of the C/EBP binding site in retroviral transcription.

C/EBP bound to the SR-D RSV LTR between nucleotides −225 and −188. The importance of this region has been suggested by deletion mutagenesis studies which place the C/EBP binding site at the 5′ limit of sequences required for the LTR enhancer activity (4, 21, 28). 5′ Deletions to nucleotide −190 reduce enhancer activity to 5% of wild-type levels (21). A similar 5′ deletion study showed that enhancer activity is reduced by 20- to 30-fold when sequences corresponding to nucleotides −203 to −196 (the downstream consensus) are removed (10). An enhancer trap assay selected only DNA fragments (nucleotides −223 to −64 and −221 to −88) containing both consensus binding sites (37). However, 3′ deletion studies indicate that the C/EBP binding region is not sufficient for the RSV LTR enhancer activity: activity is lost when sequences upstream of nucleotide −139 are deleted (21, 25). Footprints with crude nuclear extracts on the RSV LTR showed that other proteins bind in this region, and some of these have been studied by others (8, 12, 34). The full function of this enhancer probably requires multiple protein interactions, as has been seen previously for other enhancers (reviewed in reference 26).

While C/EBP was originally purified from rat liver (15), C/EBP RNA has been observed in a number of other tissues (S. McKnight, personal communication). However, the relevance of C/EBP for retroviral transcription is not yet clear. The expression of retroviral LTRs occurs in a wide range of tissues and cell types (9), but is highest in mesenchymal tissue (7, 29). Recently, several laboratories have reported binding of protein factors from other cell types to the RSV LTR. Sealey and Chalkley (34) observed binding with salt extraction fractions from a nuclear extract of quail fibroblasts. The 0.5 M NaCl extract (EFII) binds to the SR-A RSV LTR at nucleotides −229 to −203/−192. While the binding site of the EFII fraction is similar to that of C/EBP, there appears to be a difference in the other binding properties of these proteins. No competition was detected for EFII binding between the RSV LTR and SV40 and MSV enhancers (34). However, the affinity of C/EBP for the RSV enhancer that we observed was similar to that seen by Johnson et al. (15) for SV40 and MSV enhancers (P. Johnson, personal communication). In a similar study, Goodwin detected three factors binding to the RSV LTR enhancer in an avian erythroid cell line (8). Two of these chromatographically distinct factors bind within the C/EBP (EFII) protected region (8). Furthermore, the polyoma enhancer competes with the RSV LTR for binding of these factors (8). Ruddell et al., using a chicken bursal lymphoma cell line extract, detected several proteins which bind to the LTR enhancer region of avian leukosis virus, including one that is heat stable (30). The disparity between C/EBP and EFII may be due to tissue-specific differences in related proteins or to differences in the purity of the two protein preparations.

This laboratory has previously characterized an enhancer within the highly conserved *gag* coding region of several avian retroviruses, including RSV, FSV, and RAV-0 (1, 2). We further identified a C/EBP binding domain at nucleotides 813 to 872 within the FSV *gag* enhancer (2). The binding affinity observed at this site is similar to that reported here for binding between C/EBP and the RSV LTR enhancer. The *gag* enhancer displays a tandem C/EBP footprint pattern and a palindromic binding site like the one in the RSV LTR. Similarly, Karnitz et al. (18) have observed binding of a partially purified nuclear extract from BHK cells to the RSV *gag* enhancer. This binding can be inhibited by an oligonucleotide of the RSV EFII binding site in the LTR.

Binding of a common factor by two spatially separated retroviral enhancers suggests a possible interaction, perhaps by either a shuttling of *trans*-acting factors or a looping mechanism. Perhaps relevant to this latter mechanism, it has been observed that C/EBP forms stable dimers in solution, hypothesized to be a result of hydrophobic leucine interactions (23). We are tempted to speculate that the *gag* enhancer may be preferentially interacting with the more proximal of the two LTRs and thereby playing a role in the preference for initiation of retroviral transcription from the 5′ LTR (13).

ACKNOWLEDGMENTS

We gratefully acknowledge the provision of C/EBP and helpful discussions by Steve McKnight, Peter Johnson, and Barbara Graves.

This work was supported by Public Health Service grant CA 33199 from the National Cancer Institute and by biomedical research support grant SO7 RR07041 from the National Institutes of Health. T.A.R. was supported by Public Health Service predoctoral training grant ST32GM07231 from the National Institutes of Health. K.B. was a recipient of a faculty research award from the American Cancer Society.

LITERATURE CITED

1. **Arrigo, A., M. Yun, and K. Beemon.** 1987. *cis*-acting regulatory elements within *gag* genes of avian retroviruses. Mol. Cell. Biol. **7**:388–397.
2. **Carlberg, K., T. A. Ryden, and K. Beemon.** 1988. Localization and footprinting of an enhancer within the avian sarcoma virus *gag* gene. J. Virol. **62**:1617–1624.
3. **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.
4. **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.
5. **Czernilofsky, A. P., W. Delorbe, R. Swanstrom, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. The nucleotide sequence of an untranslated but conserved domain at the 3′ end of the avian sarcoma virus genome. Nucleic Acids Res. **8**:2967–2984.
6. **Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
7. **Flamant, F., D. LeGuellec, G. Verdier, and V. M. Nigon.** 1987. Tissue specificity of retrovirus expression in inoculated avian embryos revealed by in situ hybridization to whole-body section. Virology **160**:301–304.
8. **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.
9. **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.
10. **Gowda, S., A. S. Rao, Y. W. Kim, and R. V. Guntaka.** 1988. Identification of sequences in the long terminal repeat of avian sarcoma virus required for efficient transcription. Virology **162**:243–247.
11. **Graves, B. J., P. F. Johnson, and S. L. McKnight.** 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell **44**:565–576.
12. **Hatamochi, A., P. T. Golumbek, E. Van Schaftingen, and B. de Crombrugghe.** 1988. A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. J. Biol. Chem. **263**:5940–5947.
13. **Herman, S. A., and J. M. Coffin.** 1986. Differential transcription from the long terminal repeats of integrated avian leukosis virus DNA. J. Virol. **60**:497–505.
14. **Herr, W., and Y. Gluzman.** 1985. Duplications of a mutated simian virus 40 enhancer restore its activity. Nature (London) **313**:711–714.
15. **Johnson, P. F., W. H. Landschultz, B. J. Graves, and S. L. McKnight.** 1987. Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses. Genes Dev. **1**:133–146.
16. **Jones, N. C., P. W. J. Rigby, and E. B. Ziff.** 1988. Trans-acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. Genes Dev. **2**:267–281.
17. **Kane, S. E., and K. Beemon.** 1987. Inhibition of methylation at two internal N6-methyladenosine sites caused by GAC to GAU mutations. J. Biol. Chem. **262**:3422–3427.
18. **Karnitz, L., S. Faber, and R. Chalkley.** 1987. Specific nuclear proteins interact with the Rous sarcoma virus internal enhancer and share a common element with the enhancer located in the long terminal repeat of the virus. Nucleic Acids Res. **15**: 9841–9859.
19. **Khoury, G., and P. Gruss.** 1985. Enhancer elements. Cell **33**:313–314.
20. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
21. **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**:6427–6442.
22. **Landschulz, W. H., P. F. Johnson, E. Y. Adashi, B. J. Graves, and S. L. McKnight.** 1988. Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev. **2**:786–800.
23. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
24. **Lichtsteiner, S., J. Wuarin, and U. Schibler.** 1987. The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. Cell **51**:963–973.
25. **Luciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi.** 1983. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell **33**:705–716.
26. **Maniatis, T., S. Goodbourn, and J. A. Fischer.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236**: 1237–1245.
27. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**: 499–560.
28. **Norton, P. A., and J. M. Coffin.** 1987. Characterization of Rous sarcoma virus sequences essential for viral gene expression. Mol. Cell. Biol. **61**:1171–1179.
29. **Overbeek, P. A., S.-P. Lai, K. R. Van Quill, and H. Westphal.** 1986. Tissue-specific expression in transgenic mice of a fused gene containing RSV terminal sequences. Science **231**:1574–1577.
30. **Ruddell, A., M. Linial, W. Schubach, and M. Groudine.** 1988. Lability of leukosis virus enhancer-binding proteins in avian hematopoietic cells. J. Virol. **62**:2730–2735.
31. **Rushlow, K. E., J. A. Lautenberger, E. P. Reddy, L. M. Souza, M. A. Baluda, J. G. Chirikjian, and T. S. Papas.** 1982. Nucleotide sequence analysis of the long terminal repeat of avian myeloblastosis virus and adjacent host sequences. J. Virol. **42**:840–846.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
33. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
34. **Sealey, 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.
35. **Shibuya, M., L.-H. Wang, and H. Hanafusa.** 1982. Molecular cloning of the Fujinami sarcoma virus genome and its comparison with sequences of other related transforming viruses. J. Virol. **42**:1007–1016.
36. **Tsichlis, P. N., L. Donehower, G. Hager, N. Zeller, R. Malavarca, S. Astrin, and A. M. Skalka.** 1982. Sequence comparison in the crossover region of an oncogenic avian retrovirus recombinant and its nononcogenic parent: genetic regions that control growth rate and oncogenic potential. Mol. Cell. Biol. **2**:1331–1338.
37. **Weber, F., and W. Schaffner.** 1985. Enhancer activity correlates with the oncogenic potential of avian retroviruses. EMBO J. **4**:949–956.
38. **Weiher, H., M. Konig, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**: 626–631.
39. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1985. Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed., part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. **Zenke, M., T. Grundstrom, H. Matthes, M. Wintzerith, C. Schatz, A. Wildeman, and P. Chambon.** 1986. Multiple sequence motifs are involved in SV40 enhancer function. EMBO J. **5**:378–397.