Six Distinct Nuclear Factors Interact with the 75-Base-Pair Repeat of the Moloney Murine Leukemia Virus Enhancer

NANCY A. SPECK AND DAVID BALTIMORE*

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Binding sites for six distinct nuclear factors on the 75-base-pair repeat of the Moloney murine leukemia virus enhancer have been identified by an electrophoretic mobility shift assay combined with methylation interference. Three of these factors, found in WEHI 231 nuclear extracts, which we have named LVa, LVb, and LVc (for leukemia virus factors a, b, and c) have not been previously identified. Nuclear factors that bind to the conserved simian virus 40 corelike motif, the NF-1 motif, and the glucocorticoid response element were also detected. Testing of multiple cell lines showed that most factors appeared ubiquitous, except that the NF-1 binding factor was found neither in nuclear extracts from MEL cells nor in the embryonal carcinoma cell lines PCC4 and F9, and core-binding factor was relatively depleted from MEL and F9 nuclear extracts.

Eucaryotic transcriptional enhancers are *cis*-acting regions of DNA that can potentiate transcription from autologous and heterologous promoters, independent of orientation and location relative to the promoter (1, 2, 17, 19, 31, 45). There is substantial evidence that enhancers contain binding sites for proteins (33, 37, 42, 43), and it is generally assumed that the association of these proteins must mediate the transcriptional enhancement, although the mechanism for this process is unknown.

Dissection of enhancers by DNA mutational analysis (46) and, more recently, by protein-binding assays (3, 36, 43) has revealed a striking complexity in the composition and organization of enhancer sequences. There are generally multiple binding sites for nuclear factors within an enhancer, and they are often present in several copies. Viral enhancers, in particular, are composed of an impressive patchwork of conserved sequence motifs, found elsewhere in both cellular enhancers and other viral enhancers. The simian virus 40 (SV40) enhancer, for example, contains a core motif (TGTG-GTAA) that is also found in the immunoglobulin κ and μ enhancers and in the enhancers associated with the murine leukemia viruses, polyomavirus, and BK viruses (41). A sequence (GGGGACTTTCC) found in the immunoglobulin κ enhancer that binds a factor found only in B cells (36) and a conserved octanucleotide motif (ATTTGCAT) that is located 70 base pairs (bp) upstream of the transcription initiation site in immunoglobulin genes and also within the μ enhancer are also present in the SV40 enhancer (13, 30, 34, 38).

While the papovavirus enhancers have been well characterized, analysis of retroviral enhancers has not progressed as far. C-type retroviruses, as typified by Moloney murine leukemia virus (Mo-MuLV), contain enhancer elements in the U3 region of the viral genome (22). In Mo-MuLV, the enhancer is organized as a 75-bp repeated sequence located approximately 160 bp 5' to the viral cap site. With one notable exception, the Mo-MuLV enhancer is transcriptionally active in all cell types (18, 24). The disease specificity (thymic leukemia) of Mo-MuLV, however, is conferred by the U3 region of the genome (4, 9), indicating that subtle differences in transcriptional activity may influence the tissue specificity and general pathology of the disease.

To understand some of the interesting biological properties of retroviruses, it is necessary to study the transcriptional regulation conferred by the enhancer sequences. To this end, we have undertaken an analysis of the 75-bp repeat region of the Mo-MuLV enhancer. In this report, we identify six distinct binding sites for nuclear factors, as detected by an electrophoretic mobility shift assay. Methylation interference techniques have enabled us to define the sequence motifs for these nuclear factors and to demonstrate a considerable overlap of binding sites, indicating the complexity of organization of this viral enhancer.

MATERIALS AND METHODS

Extracts and cell lines. Nuclear extracts were made from tissue culture cell lines by the protocol of Dignam et al. (11) and usually contained between 3 and 8 mg of protein per ml. The following mouse cells were used: WEHI 231, a mouse B-cell line; EL4, a mouse T-cell line; MEL, a mouse erythroleukemia cell line; L cell, a mouse fibroblast; NIH 3T3, a mouse fibroblast; and PCC4 and F9, undifferentiated mouse embryonal carcinoma cells.

Column fractions of WEHI 231 nuclear extracts were kindly provided by L. Staudt. Briefly, the nuclear extract in buffer D (20 mM HEPES [N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid, pH 7.9], 0.04 M KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride, 20% [vol/vol] glycerol) was applied to either heparin Sepharose (CL-6B; Pharmacia, Inc.) or phosphocellulose (no. P11; Whatman, Inc.) equilibrated with the same buffer. Proteins were eluted from these columns in three steps of increasing KCl concentrations (0.25, 0.45, and 1.0 M) in buffer D. Fractions were dialyzed against buffer D and stored at −70°C.

Plasmids. The 450-bp $ClaI_{7676}$-$XbaI_{8111}$ fragment of the Mo-MuLV long terminal repeat, containing two 75-bp repeats, was digested with either HaeIII, HpaII, PvuII, or EcoRV, and the $PvuII_{7938}$-$PvuII_{8012}$, $PvuII_{8013}$-$XbaI_{8111}$, $EcoRV_{7960}$-$EcoRV_{8034}$, $EcoRV_{8035}$-$XbaI_{8111}$, $HpaII_{7984}$-$HpaII_{8058}$, $HpaII_{8059}$-$XbaI_{8111}$, $HaeIII_{7949}$-$HaeIII_{7993}$, and $HaeIII_{7994}$-$HaeIII_{8023}$ fragments were isolated and subcloned into the SmaI site of SP64 (see Fig. 1 for a map of the

---

* Corresponding author.

region). Approximately 500-bp competitor fragments containing the inserts were prepared by cutting in the vector sequences of SP64 (*Nae*I and *Bst*NI). The competitor fragments were isolated from 2% low-melting agarose gels by three extractions with phenol and one with chloroform followed by ethanol precipitation. The DNA was quantified by comparison with standard weights of pUC13 by being spotted on an agarose plate. Radioactive probe was prepared by cutting in the SP64 polylinker at either the *Eco*RI or *Bam*HI site, dephosphorylating the 5' end and end labeling with [γ-^{32}P]ATP in the presence of polynucleotide kinase (Boehringer Mannheim Biochemicals), and recutting at either *Bam*HI or *Eco*RI. Smaller fragments were prepared by kinasing or filling in 5' overhangs in the presence of α^{32}P-labeled ATCG by using Klenow enzyme (Boehringer Mannheim) and recutting at sites within the cloned fragment. End-labeled probes were purified and eluted from 8 to 12% native polyacrylamide gels. Typical specific activities ranged between 30,000 and 70,000 cpm/ng of DNA.

Gel-binding analysis. Binding reactions were carried out in 10 mM Tris (pH 7.5)–100 mM NaCl–1 mM EDTA–1 mM β-mercaptoethanol–4% glycerol for 20 to 30 min at room temperature. Poly(dIC) was added as a nonspecific carrier in the range of 0.5 to 8.0 μg per reaction, empirically determined for each probe. A typical reaction contained 50,000 cpm (Cerenkov) of end-labeled DNA with 8 to 12 μg of extract in a total volume of 20 μl. The binding mixture was fractionated by electrophoresis through a native 4% polyacrylamide gel (acrylamide/bisacrylamide ratio, 30:1) containing 6.7 mM Tris hydrochloride (pH 7.5), 3.3 mM sodium acetate, and 1 mM EDTA. The gel was dried and autoradiographed with a screen at −70°C. Competition experiments were performed under the conditions described above; in addition, 20 to 200 ng of unlabeled competitor fragments were included in the binding reaction prior to the addition of protein.

Methylation interference experiments. End-labeled DNA fragments were partially methylated at guanine residues as detailed by Maxam and Gilbert (26), with the following modification. The reaction was quenched with 1.5 M sodium acetate (pH 7.0)–1 M β-mercaptoethanol–0.0025% polyacrylamide carrier. Methylated DNA was ethanol precipitated twice, rinsed with 70% ethanol, and used in binding reactions. For a preparative binding, the usual conditions were scaled up 10-fold. Binding and electrophoresis were as described above. After electrophoresis, the gel was wrapped with Saran wrap and exposed wet for 4 to 6 h at room temperature. The complex and free fragments were excised, cast into 1% agarose gels, and electrophoresed onto DEAE-paper (12). DNA was eluted from the paper in 200 μl of 20 mM Tris (pH 8.0)–0.1 mM EDTA–1.0 M NaCl for 1 h at 68°C. The paper was rinsed with 200 μl of H$_2$O, and the solution was extracted with phenol (twice) and with chloroform (once), ethanol precipitated, and rinsed with 70% ethanol. The dried pellet was suspended in 1 M piperidine. Base-cleavage reactions were carried out for 45 min at 90°C, followed by lyophilization. The lyophilized fragments were analyzed on 8 to 12% polyacrylamide gels in the presence of 8 M urea and autoradiographed at −70°C with a screen.

RESULTS

Characterization of LVb and LVc. To analyze proteins that bind to individual sequence motifs, we subdivided the 75-bp repeat region of the Mo-MuLV enhancer by using several convenient internal restriction endonuclease sites. These fragments are identified by Roman numerals in Fig. 1. End-labeled fragments used for binding and unlabeled competitor fragments were isolated by cutting in the pUC13 polylinker; this adds nonspecific DNA sequences to the ends of the fragments, thus increasing their length and situating putative protein-binding sites in the middle of the fragments. Protein factors that bound to these probes were analyzed by an electrophoretic mobility shift assay (14, 16).

With this methodology, we found that WEHI 231 nuclear proteins bind to a number of sites within the Mo-MuLV enhancer. This became clear when fragment I was used because a multiplicity of protein-DNA complexes was observed (Fig. 2, lane 1). These bands could be partially resolved by chromatographically fractionating the WEHI 231 nuclear extract through heparin-agarose or phosphocellulose by using a three-step elution (0.25, 0.45, and 1.0 M KCl). The major band found with fragment I eluted in the 0.45 M KCl step from both heparin-agarose and phosphocellulose (lane 3), and its binding was specifically competed for by the addition of unlabeled fragment VII but not fragments X or XI (lanes 4 to 7), indicating that the factor bound over the *Eco*RV site. Direct-binding studies with labeled fragments showed that this complex was formed on fragment VII but not on fragments IX or XI, again indicating that the site was disrupted by cleavage at *Eco*RV (data not shown). This correlates well with data generated by Schulze et al. (34) in which a *Bgl*II linker inserted into the *Eco*RV site of the closely related Moloney murine sarcoma virus (Mo-MSV) enhancer significantly reduced the in vivo activity, as measured by CAT assays. We will refer to this factor as LVb (for leukemia virus factor b).

Two specific complexes were also formed with the 0.25 M KCl fraction (lane 2); both were competed for by unlabeled fragment I but not by a nonspecific competitor fragment (lanes 8 to 10).

Precise localization of the binding sites was obtained by using the technique of methylation interference (20). End-labeled probes were partially methylated on guanines and used in the binding reactions. The binding mixture was then fractionated by electrophoresis through polyacrylamide gels. Bands corresponding to both the complex and free DNA were eluted from the gel, subjected to piperidine cleavage, and analyzed by electrophoresis through polyacrylamide-urea sequencing gels. Methylated G residues that interfered with the binding of the factor were depleted from the protein-DNA complex band and appeared as gaps in the sequence ladder relative to free DNA.

With the 0.45 M KCl phosphocellulose fraction with fragment VII, the complex showed interference by methylation of two G residues on the coding strand and one G on the noncoding strand (Fig. 3, asterisks in lanes 1 to 4). This defined a site immediately 5' to the *Eco*RV cleavage site, correlating well with the in vivo 5' deletion data generated by Schulze et al., that localized the 5' boundary of an important regulatory sequence element of Mo-MSV to the 5'-CAGGAT-3' motif immediately 5' to the *Eco*RV site (see Fig. 6) (34).

The two protein-DNA complexes, found upon incubation of fragment I (or VII) with the 0.25 M KCl fraction, were also given methylation interference analysis. The interference pattern from the more rapidly migrating complex (open circle in Fig. 2, lane 8), hereafter referred to as the LVc-binding site, localized the DNA contacts to three G residues on the noncoding strand immediately 5' to the *Hpa*II site (data not shown). The more slowly migrating complex (diamond in Fig. 2, lane 8) showed interference patterns at both
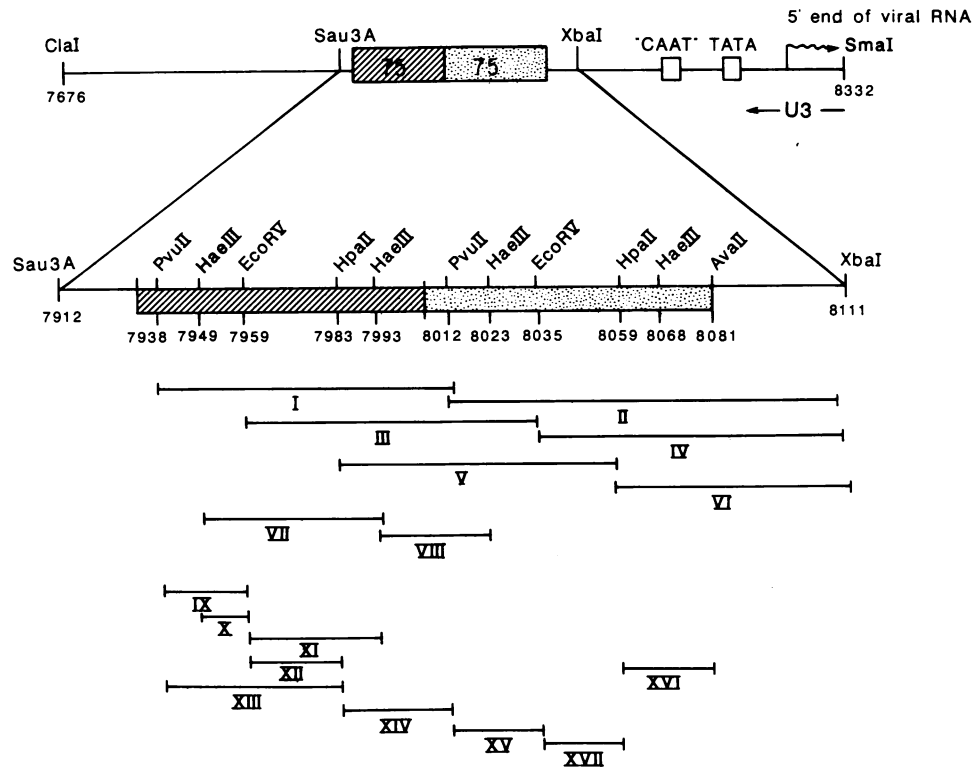
FIG. 1. Subclones of the Mo-MuLV 75-bp repeat. The top shows a schematic representation of the U3 region of Mo-MuLV contained on a $ClaI_{7676}$-$SmaI_{8332}$ fragment. Numbering starts with the 5' end of the viral genome. The 75-bp repeat region between $Sau3A_{7912}$ and $XbaI_{8111}$ is expanded below. The internal fragments $PvuII_{7938}$-$PvuII_{8012}$ (I) $HaeIII_{7949}$-$HaeIII_{7993}$ (VII), $EcoRV_{7960}$-$EcoRV_{8034}$ (III), $HpaII_{7984}$-$HpaII_{8059}$ (V), $HaeIII_{7994}$-$HaeIII_{8023}$ (VIII), $PvuII_{8013}$-$XbaI_{8111}$ (II), $EcoRV_{8035}$-$XbaI_{8111}$ (IV), and $HpaII_{8059}$-$XbaI_{8111}$ (VI) were subcloned into the $SmaI$ site of SP64, and the resultant plasmids were used to generate end-labeled probes. Fragments IX to XVII were prepared from plasmids containing fragments I to VIII. The probes and competitor fragments are referred to by Roman numerals (I to XVII) throughout the text.

the LVb site on the coding strand and the LVc site on the noncoding strand (Fig. 3, lanes 5 to 8). Although LVb was significantly depleted in the 0.25 M KCl fraction, the small amount remaining could apparently bind with LVc on the same DNA fragment. The association of both LVb and LVc on the same fragment to form a complex having a mobility distinct from those of LVb and LVc binding alone indicates that the distance between theLVb and LVc sites is sufficient to accommodate both proteins simultaneously. The arrow near lane 5 in Fig. 3b indicates a hypersensitive cleavage site in the LVb-LVc-DNA complex, between the LVb- and LVc-binding sites. We showed in an independent experiment that this did not depend on G-specific cleavage by isolating the protein-DNA complex and free DNA bands from a preparative gel and, without piperidine cleavage, analyzing the DNA in a denaturing polyacrylamide-urea gel, next to a sequence ladder. A single band migrating at the same position was seen in the protein-DNA band but not in free DNA (data not shown). This suggests that the association of LVb and LVc on the same fragment may render the DNA between these sites sensitive to cleavage by an endogenous endonuclease found in the nuclear extract (or possibly the LVb or LVc factors themselves).

Characterization of LVa. A third Mo-MuLV binding factor, which we call LVa, was found to bind to fragment VIII, a region covering the $PvuII_{8012}$ site (Fig. 4, lane 1). Nucleotides in the $PvuII$ site seemed important for this binding because the complex was not found with fragment XIV or XV at the same protein/poly(dIC) ratio [4 μg of protein to 1.0

μg of poly(dIC)] used for fragment VIII (data not shown). This complex could be specifically competed for by addition of an unlabeled competitor DNA fragment spanning the $PvuII$ site (fragment V) (lane 2) but not by a nonoverlapping fragment (fragment XI) (lane 3). The methylation interference assay on this complex from unfractionated WEHI 231 nuclear extracts (data not shown) localized important contact G residues on the coding strand immediately over the $PvuII$ site (see Fig. 6).

Figure 4, lanes 4 to 7, shows the methylation interference pattern generated on fragment VIII in a 0.45 M KCl heparin fraction. At a ratio of 12 μg of protein to 1.0 μg of poly(dIC), interference was apparent at G residues within both the LVa site (see asterisks, lane 4) and the adjacent glucocorticoid response element (GRE) site (see bracket, lane 4), for reasons that will be discussed in the following section.

Detection of factors binding to conserved sequence motifs: GRE. The Mo-MuLV 75-bp repeat contains, in addition to LVa, LVb, and LVc, several conserved sequence motifs found in other better characterized transcriptional enhancers. The purified glucocorticoid hormone receptor binds to three sites on the Mo-MSV long terminal repeat (8). The consensus sequence motifs (AGAACAGATG) for the GRE in the Mo-MuLV 75-bp repeat are located at nucleotides 7997 to 8006 and 8073 to 8081. Although fragment VIII contains a GRE motif, in addition to the site for LVa, methylation at G residues within the GRE did not interfere with binding in unfractionated WEHI 231 nuclear extracts at a protein/poly(dIC) ratio of 4:1 (data not shown). The
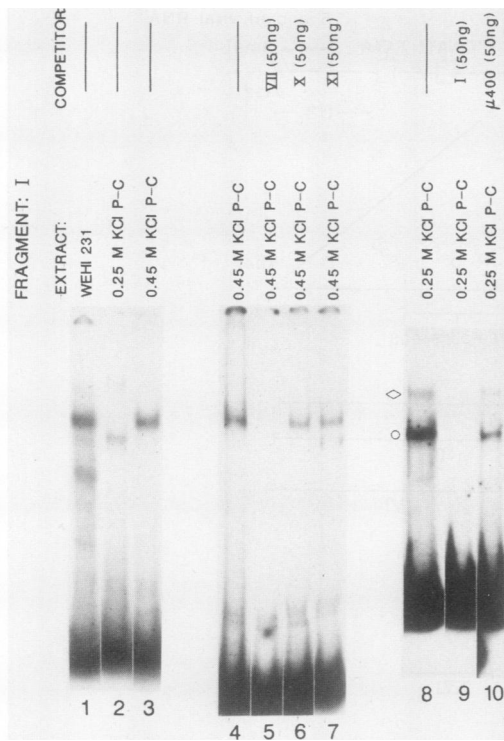
FIG. 2. Electrophoretic mobility shift assay of fragment I. End-labeled fragment I (50,000 cpm/2.5 ng) was incubated with unfractionated nuclear extract from WEHI 231 cells or column fractions from a stepwise KCl elution from a phosphocellulose column (for details, see Materials and Methods) in the presence of poly(dIC) and nonradioactive competitor DNA. Lanes: 1, unfractionated WEHI 231 nuclear extracts [12 μg of protein, 6 μg of poly(dIC)]; 2 and 3, 0.25 M KCl (lane 2) and 0.45 M KCl (lane 3) column fractions, 2 to 5 μg of protein, 1 μg of poly(dIC); 4 to 7, competition analysis of the complex formed in the 0.45 M KCl fraction (LVb); 50 ng of a 500-bp competitor fragment (containing approximately 400 bp of SP64 sequence in addition to the Mo-MuLV-derived insert) was included in the binding reaction (see Fig. 1 for nomenclature of fragments); 4, no competitor; 5, competition with fragment VII, which spans the EcoRV site; 6 and 7, competition with both halves of the same fragment (VII) when cut at EcoRV; 8 to 10, competition of the complexes formed in the 0.25 M KCl fraction (lane 8, no competitor; lane 9, binding in the presence of unlabeled fragment I; lane 10, competition with a nonspecific fragment). The nonspecific competitor, μ400, is an XbaI-PvuII fragment from the immunoglobulin μ enhancer, kindly provided by R. Sen. Symbols: ◇, ○, methylation interference patterns were obtained.

binding site for LVa and the GRE are in close proximity to each other; in fact, the furthest 3' nucleotide of the conserved GRE motif (nucleotide 7997 on the first 75-bp repeat) is immediately 5' to the first G residue showing interference in the binding site of LVa (nucleotide 7998).

To detect binding specific for the GRE, it was necessary to use a probe that truncated the binding site for LVa at the $PvuII_{8012}$ site. When fragment XIV was used in the gel mobility shift assay, specific complexes were most evident at a protein/poly(dIC) ratio of 24 μg of protein to 1.0 μg of poly(dIC) (Fig. 5, lane 1). The GRE complex was no longer detectable at the protein/poly(dIC) ratio [4 μg of protein to 1.0 μg of poly(dIC)] used to visualize the LVa complex on fragment VIII containing the LVa and GRE sites in unfractionated nuclear extracts (lanes 3 and 4). Complex formation was specifically competed for by addition of an unlabeled fragment overlapping the GRE site (fragment I) but not a

nonoverlapping fragment (fragment XI). Methylation interference was performed on the three bands indicated by the open symbols in Fig. 5, lane 1. The two upper bands (open triangle, open circle) showed interference at the expected G residues 7998, 8003, and 8006 on the coding strand and 8001 on the noncoding strand, while the lower band (open
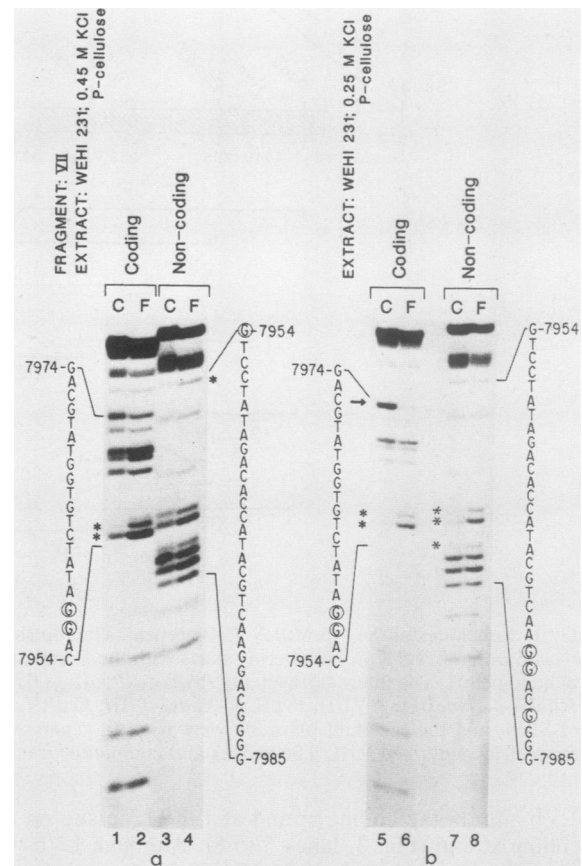

FIG. 3. Methylation interference experiments to define the binding sites of the proteins interacting with fragment VII. The complexes formed upon incubation of fragment I with the 0.25 and 0.45 M KCl column fractions were also seen with a smaller internal fragment, fragment VII. Fragment VII was end labeled at the EcoRI site in the SP64 polylinker to visualize the coding strand (lanes 1 and 2) and the BamHI site for the noncoding strand (lanes 3 and 4). A typical preparative reaction contained 500,000 to 1,000,000 cpm in 100 to 200 μl (final volume). (a) Nucleoprotein complex band isolated from a preparative low-ionic-strength polyacrylamide gel followed by piperidine treatment. Preparative binding reaction mixture contained 40 μg of protein and 5 μg of poly(dIC). Lanes 1 and 3 correspond to the protein-DNA complexes formed in the 0.45 M KCl fraction eluted from the phosphocellulose column. Lanes 2 and 4 are free DNA bands from the same binding reactions. Asterisks indicate the location of G residues whose methylation by dimethyl sulfate specifically inhibits the binding of the factor to its cognate sequence. At the protein/poly(dIC) ratio used in the binding reaction (8:1), partial interference at nucleotide 7967 in the core sequence was also apparent. (b) Methylation interference pattern from the low-mobility complex (see ◇, Fig. 2, lane 8) formed upon incubation of fragment VII with the 0.25 M KCl fraction from a phosphocellulose column [60 μg of protein, 5 μg of poly(dIC)]. Lanes: 5 and 6, bands corresponding to protein-DNA complex and free DNA, respectively, on the coding strand; 7, protein-DNA complex; 8, free DNA band visualized on the noncoding strand. Arrow (lane 5) indicates site of complex-specific enhanced cleavage on the coding strand, presumably by endogenous endonucleases in the nuclear extract.

square), which competed poorly, showed no methylation interference pattern (data not shown; see Fig. 6).

Both LVa and the GRE-binding factor elute from heparin in the 0.45 M KCl step, and the protein-DNA complexes comigrate in the electrophoretic mobility shift assay. At the relatively high protein/poly(dIC) ratios used to generate the methylation interference on fragment VIII (Fig. 4), both the GRE and LVa complexes are present in the isolated band and thus show interference at both sites.

Binding to the NF-1 motif. The 5' end of the conserved GRE motif in the Mo-MuLV enhancer overlaps the 3' end of a motif with a 7- to 8-bp homology to the consensus sequence for another well-characterized protein, nuclear factor 1 (NF-1) (28). NF-1 was first found to enhance in vitro replication of the adenovirus type 2 genome (10, 28, 32), with enhancement dependent upon NF-1 binding near the origin
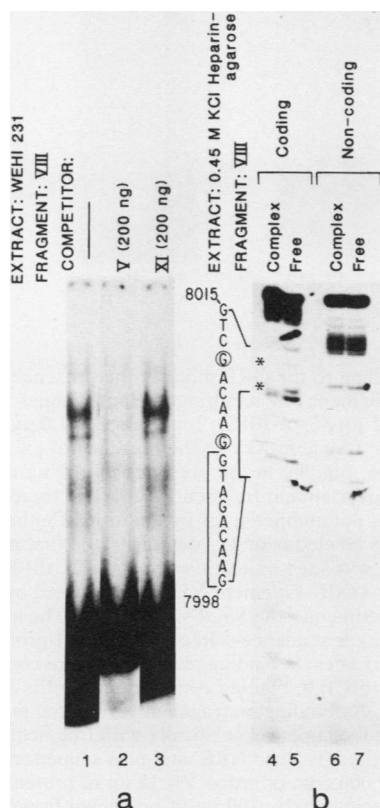

FIG. 4. Localization of the binding site for LVa. (a) Lanes: 1, the complex formed on fragment VIII in unfractionated WEHI 231 nuclear extracts [binding reaction mixtures contain 50,000 cpm of end-labeled fragment, 12 μg of protein and 3 μg of poly(dIC)]; 2, binding in the presence of 200 ng of unlabeled competitor DNA spanning the binding site (fragment V); 3, binding in the presence of 200 ng of unlabeled competitor DNA from adjacent, nonoverlapping sequences (fragment XI). (b) Methylation interference assay on protein-DNA complex formed on fragment VIII in a 0.45 M KCl fraction from heparin-agarose. Preparative binding reaction mixtures contained 5 μg of poly(dIC), 60 μg of protein, and 1,000,000 cpm of fragment VIII end labeled at either EcoRI (coding strand) or BamHI (noncoding strand) in the SP64 polylinker. Lanes: 4 and 5, protein-DNA complex and free DNA from probe labeled on coding strand; 6 and 7, complex and free DNA bands labeled on noncoding strand. Asterisks indicate methylated G residues that interfered with complex formation within the LVa site. Brackets indicate the adjacent GRE site.

(10, 32). Since then, NF-1 binding sites with the consensus sequence 5'- TGG(N)$_{6-7}$GCCA-3' have been identified in a variety of genes, including chicken lysozyme gene, BK enhancer, murine mammary tumor virus long terminal repeat, 5' to the c-myc gene, and near the switch regions of immunoglobulin M (23).

The fragment used to detect binding to the NF-1 site proximal to the GRE was prepared by cutting and 5' end labeling fragment VI in the polylinker of pUC13 and recutting at AvaII, which truncates the GRE at its 3' end (fragment XVI). A specific complex was formed on fragment XVI (solid circle in Fig. 5, lane 8), which was competed for by addition of unlabeled fragment I (lane 9) but not by fragment VIII (lane 10), which contains the GRE but not an intact NF-1 site. Methylation interference for the NF-1 binding site was obtained with a longer fragment (VI), which generated a more complex series of bands, by isolating the high-mobility complex (solid square in Fig. 5, lane 13) that was also competed specifically for by fragment I but not by fragment VIII (Fig. 5, lanes 11 to 13). A complex with similar mobility and an identical methylation interference pattern was found with a longer probe (fragment IV). The methylated residues that interfered with binding are shown in Fig. 6.

There is another sequence with a 7- to 8-bp match to the NF-1 consensus [TGA(N)$_6$GCCA] on the Mo-MuLV enhancer between the binding sites for LVa and LVb (nucleotides 7939 to 7951 and 8014 to 8026), making a total of four putative NF-1 sites on the two 75-bp repeats. Binding to this site was not tested directly, however, and only the homology is indicated on Fig. 6.

Detection of a core-binding factor. The Mo-MuLV 75-bp repeat contains two copies of an SV40 corelike motif (TGTGGTA$_A^T$) (41), located immediately 3' to the binding site for LVb. DNase I protection experiments on the SV40 enhancer in HeLa cell nuclear extracts yield protection over the core motif (43). Moreover, the SV40 core-binding factor has been partially purified (S. McKnight, personal communication) and found to also protect the Mo-MSV core motif from DNAse digestion. Because association of LVb could potentially interfere with binding to the Mo-MuLV core, a probe was prepared by 5' and 3' end labeling fragment III from a site in the pUC13 polylinker, and recutting at HpaII$_{7983}$ or HaeIII$_{7993}$ (fragments XII and XI), thus eliminating the binding site for LVb. A specific complex was formed on the XII or XI fragment (Fig. 5, lane 14 to 16). Interference analysis indicated that methylation at nucleotides 7964, 7966, and 7967 on the coding strand affected complex formation (Fig. 6), as well as partial interference at nucleotides 7962 and 7972 on the noncoding strand.

Tissue distribution of factors binding the Mo-MuLV enhancer. Nuclear extracts generated from cell lines representing a variety of tissue types were assayed for the presence of the Mo-MuLV enhancer-binding factors. Figure 7 shows a representative analysis. Lanes 1 to 4 and 7 show binding to fragment XIII with WEHI 231 (B cell), EL4 (mouse T thymoma), L cell (fibroblast), MEL (mouse erythroid leukemia cell), and NIH 3T3 (fibroblast) cells. Also shown are two undifferentiated embryonal carcinoma cell lines, PCC4 and F9 (lanes 5 and 6), in which the enhancer is transcriptionally inactive. All cell types contained the LVb-binding factor. The predominant complex on fragment XIII at the protein/poly(dIC) ratio of 2:1 is contributed by LVb (as seen for fragment I in Fig. 1). An attempt was made to normalize for LVb in this experiment and then to compare the relative levels of core-binding factor by using the same amounts of extract used in lanes 1 to 7 and a constant amount of
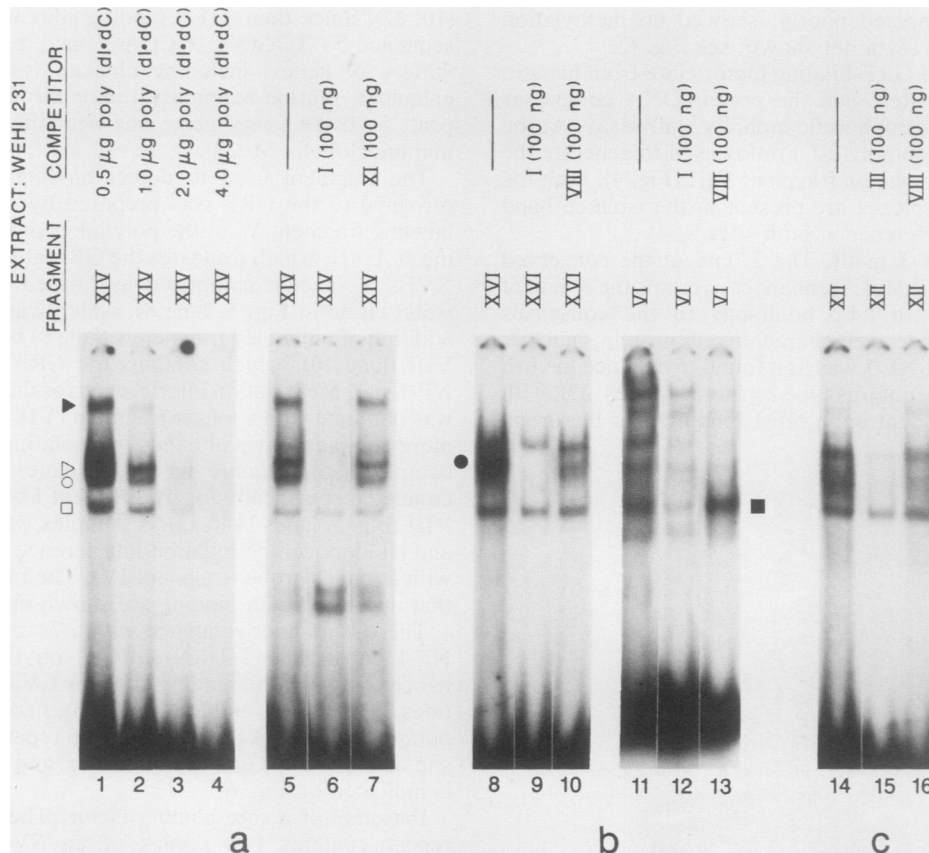
FIG. 5. Nuclear factors binding the GRE, NF-1, and core consensus motifs. (a) Binding to the GRE motif in the presence of increasing amounts of poly(dIC). The GRE probe (fragment XIV) was prepared from end-labeled fragment I by recutting at HpaII. Lanes: 1, incubation of 50,000 cpm of probe XIV with 12 μg of WEHI 231 nuclear extract in the presence of 0.5 μg of poly(dIC); 2, presence of 1.0 μg of poly(dIC); 3, presence of 2.0 μg of poly(dIC); 4, presence of 4.0 μg of poly(dIC); 5 to 7, binding on fragment XIV in the presence of excess unlabeled fragments; binding conditions are the same as in lane 2 (Lane 5, no competitor; lane 6, binding in the presence of 100 ng of fragment I, spanning the GRE site; lane 7, binding in the presence of 100 ng of an adjacent, nonoverlapping fragment [XI]). The three upper bands indicated in lane 1 by ▶, ▷, and ○ competed specifically, while the lower band (□) was not competed for by addition of either the specific or the nonspecific competitor fragment. The methylation interference assay on the bands labeled ▷ and ○ demonstrated that methylation of G residues within the GRE site perturbed binding. No methylation interference pattern was seen on the lower band (□). (b) Binding to the NF1-consensus. Probe XVI was prepared by end labeling fragment VI and recutting at AvaII. Fragment VI was subcloned by filling in the 5' overhang generated by HpaII with the Klenow fragment of DNA polymerase I and ligating into the SmaI site of SP64. The first nucleotide (C residue, position 8059) of the NF-1 site is therefore preserved and flanked by polylinker sequences. Recutting the VI probe with AvaII truncates the GRE site at its 3' end, probably resulting in destabilization of that complex. Lanes: 8, binding reaction mixture contained 50,000 cpm of probe XVI, 12 μg of protein from WEHI 231 nuclear extracts, and 1 μg of poly(dIC); 9, binding conditions were the same as those in lane 8, with the addition of 100 ng of unlabeled fragment I, which spans the NF-1 site; 10, binding on fragment XVI in the presence of 100 ng of unlabeled fragment VIII, containing the GRE site (the band indicated by ● in lane 8 competes specifically with fragment I, containing NF-1 site, but not with fragment VIII); 11, Binding to a longer fragment (VI) containing the NF-1 and GRE site, plus sequences between the 3' end of the direct repeat and the XbaI_{8111} site. Binding reaction mixtures contained 50,000 cpm of probe VI, 12 μg of protein, and 1 μg of poly(dIC). The protein-DNA complex indicated by ■ in lane 13 was competed for specifically by 100 ng of unlabeled fragment I but not fragment VIII (100 ng). The methylation interference pattern obtained for this protein-DNA complex demonstrated that the factor contacted the DNA at G residues within the NF-1 consensus sequence (see Fig. 6). (c) Binding to the core motif. Lane 14, Probe XII was prepared by cutting end-labeled fragment III at the HpaII site. Fragment III was cut and end labeled in the SP64 polylinker; thus, additional nonspecific sequences flank the EcoRV site, moving the core motif away from the end of the fragment. Binding reaction mixtures contained 50,000 cpm of probe XII, 15 μg of protein, and 1 μg of poly(dIC). Lane 15, binding to probe XII in the presence of 100 ng of unlabeled fragment III; lane 16, competition with fragment VIII.

poly(dIC). Fragment XI, used to detect the core-binding factor, also contains the LVc-binding site. Interference analysis on the predominant complex formed on fragment XI identified protein-DNA contacts at nucleotides within the core consensus sequence but not in the LVc site. It is evident that although all of the extracts contained the LVb-binding factor, some lacked or were greatly deficient in the core-binding factor (Fig. 7, lanes 8 to 14). Specifically, MEL and F9 cells were relatively depleted of the core-

binding factor, and the amount of factor in PCC4 cells was also low.

The results of an analysis of many extracts and fragments are summarized in Table 1. Several factors, notably LVa, LVb, and the GRE-binding factor, are present in every extract tested, although in varying amounts relative to each other. A complex was formed on the NF-1 probe in WEHI 231, EL4, L cell, and NIH 3T3 nuclear extracts but not in extracts generated from MEL, PCC4, or F9 cells. There
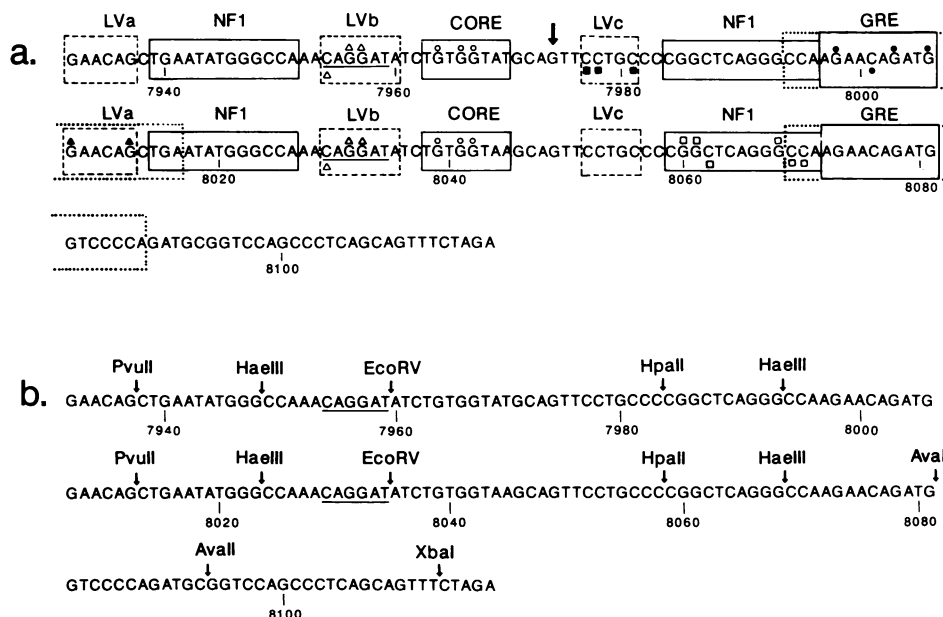
FIG. 6. Summary of the results from methylation interference experiments. (a) The binding sites for the various nuclear factors are indicated by boxes. Solid lines show the boundaries of previously characterized consensus motifs (NF1, core, GRE). Hatched lines indicate minimal boundaries for novel factors (LVa, LVb and LVc), as determined by methylation interference. The binding site for LVb is extended to nucleotide 7960, since truncation of the probe at EcoRV perturbed binding at the LVb site. The sequence CAGGAT, identified by Schulze et al. (34) to contain an important regulatory element for Mo-MSV, is underlined. Methylated G residues that interfered with complex formation are indicated above the sequence for the coding strand and below the sequence (on C residues) for the noncoding strand. Symbols: ○, core; △, LVb; ▲, LVa; ■, LVc; ●, GRE; □, NF1. Methylation interferences are shown only on the sites (i.e., the promoter proximal or distal 75-bp sequence) upon which they were obtained experimentally. The arrow at position 7974 shows the site of endonucleotic cleavage between the LVb and LVc sites (Fig. 3b). The dotted line extending from the box surrounding the GRE site indicates the boundaries of the DNAseI footprint generated by DeFranco and Yamamoto (8) using the purified glucorticoid receptor. For the purpose of orientation, the location of restriction sites in the enhancer sequence is shown in panel b.
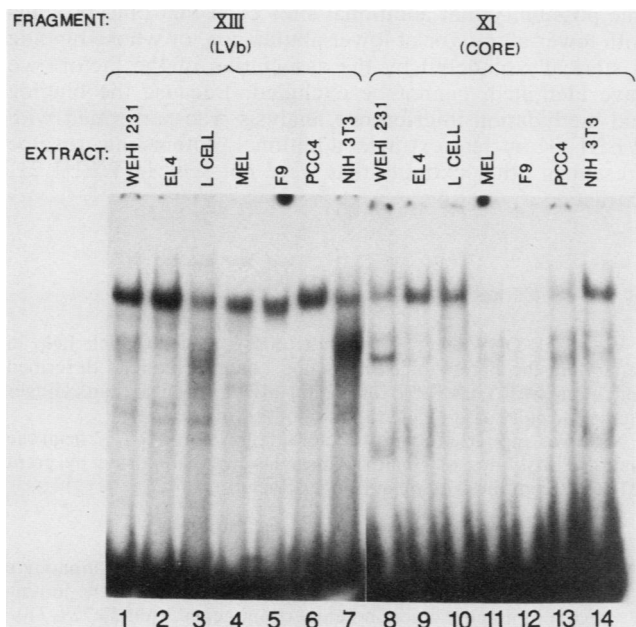

FIG. 7. Tissue distribution of LVb and core-binding factor. In lanes 1 to 7, end-labeled fragment XIII was incubated with 12 to 16 μg of protein from various cell extracts and 4.0 μg of poly(dIC) in a 20-μl binding reaction mixture, which was then analyzed by electrophoresis. Lanes 8 to 14 show binding to fragment XI containing the core motif. The same amount of nuclear extract for each cell line in lanes 1 to 7 was used in lanes 8 to 14, so that the relative

appears, therefore, to be some variation in the presence or relative amounts of the factors that bind the Mo-MuLV enhancer, which may determine its relative transcriptional efficiency in various tissues and ultimately influence the tropism and leukemogenicity of the virus.

DISCUSSION

We have detected six distinct factors that can bind to sequence motifs in the Mo-MuLV 75-bp repeat and have defined the important contact residues for binding. The binding sites for these factors are distributed along the length of the 75-bp repeat and are in close proximity to each other. The binding sites for three factors, LVa, LVb, and LVc, have not yet been identified, to our knowledge, on other eucaryotic enhancers.

By defining the important contact residues for protein-DNA interactions, the methylation interference assay gives a minimal size for the space on the DNA that is occupied by the protein. This minimal distance is indicated in Fig. 6 for LVa, LVb, and LVc by a hatched line surrounding the boundaries of important contact residues. For LVb, the first nucleotide residue 3' to the cleavage site for EcoRV is also

abundance of these two factors, LVb and core-binding factor, could be qualitatively compared. Fragment XI was prepared by cutting end-labeled fragment III at HaeIII; thus, SP64 polylinker sequences flank the 5' side of the EcoRV site, removing the binding site from the end of the fragment. Binding to fragment XI was done in the presence of 1.0 μg of poly(dIC). The cell lines used to derive the extracts are indicated above each line.

TABLE 1. Tissue distribution of factors binding the Mo-MuLV 75-bp repeat

Factor	Factors present in:						
	WEHI 231	EL4	MEL	L cells	NIH 3T3	PCC4	F9
LVa	+	+	+	+	+	+	+
LVb	+	+	+	+	+	+	+
LVc	+	ND[a]	ND	ND	ND	ND	ND
GRE	+	+	+	+	+	+	+
Core	+	+	±	+	+	+	±
NF-1	+	+	−	+	+	−	−

[a] ND, Not done.

included in the binding site, inferred from the fact that the complex is disrupted by cleavage at that site. Since the consensus motifs for NF-1 and core are separated from the LVb-binding site by only two nucleotides, it is highly probable that the association of LVb would sterically interfere with the binding of factors to the adjacent sites, or vice versa. The same situation is true for the NF-1/GRE, GRE/LVa, and LVa/NF-1 consensus motifs. The conserved GRE motif is shown boxed in by a solid line in Fig. 6. The dotted line that extends 3 bp 5' and 10 bp 3' to the conserved motif indicates the boundaries of the DNase I footprint generated by DeFranco and Yamamoto using the purified glucocorticoid hormone receptor (8). The 3' extension of the DNase I boundary, which gives some indication of the space that is being occupied by the protein, extends well over the site for LVa. It is obvious that the regulation of transcription must be complex, because the relative abundance or affinity of each of the various factors will presumably influence the binding of factors to adjacent sites.

The dissection of the Mo-MuLV enhancer-binding sites illustrates the power of combined gel retardation assays and methylation interference techniques. Many of the protein-DNA complexes detected on the Mo-MuLV enhancer were relatively weak, and often <5% of the probe was incorporated into the complex. By first electrophoretically resolving the protein-DNA complex from free DNA, it was possible to detect complex-specific deletions in the sequence ladder that would have been impossible to visualize in the presence of the vast excess of free DNA in the binding reactions. Once the binding site for a protein-DNA complex has been defined by methylation interference, the formation of that complex in a variety of extracts can be assayed. Thus, the presence of each of the Mo-MuLV enhancer-binding factors can be determined without chromatographically fractionating each extract.

Recently, several reports have been published describing the responsiveness of the closely related Mo-MSV enhancer to steroid hormones (8, 25, 37). It was observed that in addition to being stimulated through the binding of the glucocorticoid hormone, the Mo-MSV enhancer also contained high basal activity in the absence of stimulating hormones (8, 27). The data presented in this report suggest that this constitutive basal activity is conferred by the association of additional factors on the Mo-MSV enhancer. The observation by Miksicek et al. (27) that the basal activity of the Mo-MSV enhancer is destroyed by restriction with EcoRV correlates well with the binding data, which demonstrated that the predominant complex on the Mo-MuLV enhancer, that generated by LVb, is localized over the EcoRV site. Examination of enhancer sequences from a variety of murine leukemia viruses, including Mo-MuLV, Mo-MSV, spleen focus-forming virus, and Friend murine

leukemia virus (5), indicates strong conservation at the LVb site, compared with LVa and LVc, suggesting that it is an especially important regulatory element in murine leukemia viruses.

Although the Mo-MuLV enhancer contains three GRE consensus motifs (the third GRE is located 5' to the first 75-bp repeat and is not included in Fig. 6), in vivo data suggest that, unlike Mo-MSV, it is not positively regulated by steroid hormones (29). Comparison of the Mo-MuLV and Mo-MSV LTR sequences indicates only 6-bp differences between the two enhancers (6, 40), all occurring outside the consensus octanucleotide AGA^A_TCAG^A_T, that has been shown to be functional in GRE enhancement (7, 44). This would suggest that proteins other than the glucocorticoid hormone receptor may modulate the responsiveness to steroid hormones, perhaps by influencing the availability of the GRE for hormone binding.

A survey of extracts for five of the Mo-MuLV enhancer-binding proteins revealed that one factor, NF-1, was absent in the undifferentiated embryonal carcinoma cell lines F9 and PCC4. F9 cells also did not contain significant levels of the core-binding factor. It is interesting, in this respect, that mutations of the polyoma enhancer that allow replication in EC cells were different depending upon whether the virus was isolated from F9 or PCC4 cells (15, 21, 35). The most stringent restriction for growth was seen in F9 cells, in that mutants grown in F9 cells could also replicate in PCC4 cells, but the reciprocal was not true (39). Our data would suggest that F9 cells may be depleted of more transcription factors than PCC4, as seen for the Mo-MuLV enhancer. It may also be that the absence of sequence-specific transcription factors is involved in determining the inactive phenotype of Mo-MuLV in EC cells.

The six factors that were identified may not represent all of the factors that can bind to the Mo-MuLV 75-bp repeat. The possibility that additional sites exist that bind proteins with lower affinity or of lower abundance, or whose binding is sterically hindered by the association of the factors we have identified, cannot be excluded. Because the binding and methylation interference analysis was performed with WEHI 231 nuclear extracts, additional proteins may also be present in other extracts that were not seen in WEHI 231 extracts.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell 33:729–740.
2. Benoist, C., and P. Chambon. 1981. In vitro sequence requirements of the SV40 early promoter region. Nature (London) 290:304–310.
3. Boehnlein, E., and P. Gruss. 1986. Interaction of distinct nuclear proteins with sequences controlling the expression of polyomavirus early genes. Mol. Cell. Biol. 6:1401–1411.
4. Chatis, P. A., C. Holland, J. W. Hartley, W. P. Rowe, and N. Hopkins. 1983. Role of the 3'-end of the genome in determining

disease specificity of Friend and Moloney murine leukemia viruses. Proc. Natl. Acad. Sci. USA **80:**4408–4411.

5. **Clarke, S. P., R. Kaufhold, A. Chan, and T. W. Mak.** 1985. Comparison of the transcriptional properties of Friend and Moloney retrovirus long terminal repeats: importance of tandem duplications and the core enhancer sequence. Virology **144:**481–494.

6. **Dar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA **77:**3937–3941.

7. **DeFranco, D. O., Wrange, J. Merryweather, and K. R. Yamamoto.** 1985. Biological activity of a glucocorticoid regulated enhancer: DNA sequence requirements and interactions with other transcriptional enhancers. UCLA Symp. Mol. Cell. Biol. **20:**305–321.

8. **DeFranco, D., and K. Yamamoto.** 1986. Two different factors act separately or together to specify functionally distinct activities at a single transcriptional enhancer. Mol. Cell. Biol. **6:**993–1001.

9. **DesGroseillers, L., E. Rassart, and P. Jolicoeur.** 1983. Thymotropism of murine leukemia virus is conferred by its long terminal repeat. Proc. Natl. Acad. Sci. USA **80:**4203–4207.

10. **DeVries, E., W. van Driel, M. Tromp, J. H. van Boom, and P. C. van der Vliet.** 1985. Adenovirus DNA replication *in vitro*: site directed mutagenesis of the nuclear factor I binding site of the Ad2 origin. Nucleic Acids Res. **13:**4935–4952.

11. **Dignam, J. D., R. M. Lebowitz, 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.

12. **Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112:**295–298.

13. **Faulkner, F. G., and H. G. Zachau.** 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature (London) **310:**71–74.

14. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of Lac repressor operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9:**6505–6525.

15. **Fujimura, F. K., and E. Linney.** 1982. Polyoma mutants that productively infect F9 embryonal carcinoma cells do not rescue wild type polyoma in F9 cells. Proc. Natl. Acad. Sci. USA **79:**1479–1483.

16. **Garner, M. M., and A. Revzin.** 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the E. coli lactose operon regulatory system. Nucleic Acids Res. **9:**3047–3060.

17. **Gillies, S. D., S. L. Morrison, V. Oi, and S. Tonegawa.** 1983. A tissue specific enhancer is located in the major intron of a rearranged heavy chain gene. Cell **33:**717–728.

18. **Gorman, C. M., P. W. J. Rigby, and D. P. Lane.** 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell **42:**519–526.

19. **Gruss, P., R. Dhar, and G. Khoury.** 1981. Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Natl. Acad. Sci. USA **78:**943–947.

20. **Hendrickson, W., and R. F. Schlief.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the AraI DNA site. Proc. Natl. Acad. Sci. USA **82:**3129–3133.

21. **Katinka, M., M. Vasseur, N. Montreau, M. Yaniv, and P. Blangy.** 1981. Polyoma DNA sequences involved in control of viral gene expression in murine embryonal carcinoma cells. Nature (London) **290:**720–722.

22. **Laimins, L. A., P. Gruss, R. Pozzatti, and G. Khoury.** 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. **49:**183–189.

23. **Leegwater, P. A. J., P. C. van der Vliet, R. A. W. Rupp, J. Nowock, and A. Suppel.** 1986. Functional homology between the sequence-specific DNA-binding proteins nuclear factor I from HeLa cells and the TGGCA protein from chicken liver. EMBO J. **5:**381–386.

24. **Linney, E., B. Davis, J. Overhauser, E. Chao, and H. Fan.** 1984.

Non-function of a Moloney murine leukemia virus regulatory sequence in F9 embryonal carcinoma cells. Nature (London) **308:**470–472.

25. **Lowy, D. R., and E. M. Scolnick.** 1978. Glucocorticoids induce focus formation and increase sarcoma viral expression in a mink cell line that contains a murine sarcoma viral genome. J. Virol. **25:**157–163.

26. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:**560–564.

27. **Miksicek, R., A. Heber, W. Shmid, U. Danesch, G. Posseckert, M. Beato, and G. Schulze.** 1986. Glucocorticoid responsiveness of the transcriptional enhancer of Moloney murine sarcoma virus. Cell **46:**283–290.

28. **Nagata, K., R. A. Guggenheimer, T. Inomoto, J. H. Lichy, and J. Hurwitz.** 1982. Adenovirus replication *in vitro*: identification of a host factor that stimulates synthesis of the preterminal protein-dCMP complex. Proc. Natl. Acad. Sci. USA **79:**6438–6442.

29. **Overhauser, J., and H. Fan.** 1985. Generation of glucocorticoid reponsive Moloney murine leukemia virus by insertion of regulatory sequences from murine mammary tumor virus into the long terminal repeat. J. Virol. **54:**133–144.

30. **Parslow, T. G., D. L. Blair, W. J. Murphy, and D. K. Granner.** 1984. Structure of the 5'-ends of immunoglobulin genes; a novel conserved sequence. Proc. Natl. Acad. Sci. USA **81:**2650–2654.

31. **Queen, C., and D. Baltimore.** 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. Cell **33:**741–748.

32. **Rawlins, D. R., P. J. Rosenfeld, R. J. Wides, M. D. Challberg, and T. J. Kelly.** 1984. Structure and function of the adenovirus origin of replication. Cell **37:**309–319.

33. **Sassone-Corsi, P., J. Daugherty, B. Wasylyk, and P. Chambon.** 1984. Stimulation of *in vitro* transcription from heterologous promoters by the SV40 enhancer. Proc. Natl. Acad. Sci. USA **81:**308–312.

34. **Schulze, F., E. Boehnlein, and P. Gruss.** 1985. Mutational analysis of the Moloney murine sarcoma virus enhancer. DNA **4:**193–202.

35. **Sekikawa, K., and A. J. Levine.** 1981. Isolation and characterization of polyoma host-range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA **78:**1100–1104.

36. **Sen, R., and D. Baltimore.** 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell **46:**705–716.

37. **Sergeant, A., D. Bohmann, H. Zentgraff, H. Weiher, and W. Keller.** 1984. A transcriptional enhancer acts *in vitro* over distances of hundreds of base pairs on both circular and linear templates but not on chromatin-reconstituted DNA. J. Mol. Biol. **180:**577–600.

38. **Singh, H., R. Sen, D. Baltimore, and P. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319:**154–158.

39. **Tanaka, K., K. Chowdhury, L. T. Liang, K. S. S. Chang, M. Israel, and Y. Ito.** 1983. Relationships between growth restriction to polyoma virus and stages of differentiation in embryonic cell lines, p. 293–305. *In* L. M. Silver, G. R. Martin, and S. Strickland (ed.), Teratocarcinoma stem cells. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

40. **Van Beveran, C., E. Rands, S. K. Chattapadhyay, D. R. Lowy, and I. M. Verma.** 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions, and preintegration site. J. Virol. **41:**542–556.

41. **Weiher, H., M. Zonig, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219:**626–631.

42. **Wildeman, A., P. Sassone-Corsi, T. Grundstrom, M. Zenke, and P. Chambon.** 1984. Stimulation of in vitro transcription from the SV40 early promoter by the enhancer involves a specific transacting factor. EMBO J. **3:**2129–3133.

43. **Wildeman, A., M. Zenke, C. Schatz, M. Wintzerith, T. Grundstrom, H. Matthes, K. Takahashi, and P. Chambon.** 1986.

Specific protein binding to the simian virus 40 enhancer in vitro. Mol. Cell. Biol. **6:**2098–2105.

44. **Yamamoto, K. R.** 1985. Hormone-dependent transcriptional enhancement and its implications for mechanisms of multifactor gene regulation, p. 131–148. *In* L. Bograd and G. Adelman (ed.), Molecular developmental biology: expressing foreign genes. 43rd Symposium of the Society for Developmental Biology.

Alan Liss Inc., New York.

45. **Yaniv, M.** 1982. Enhancing elements for activation of eukaryotic promoters. Nature (London) **297:**17–18.

46. **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:** 387–397.