Involvement of Nuclear Factor I-Binding Sites in Control of Akv Virus Gene Expression

HENRIK STEEN OLSEN,[†] STEEN LOVMAND, JETTE LOVMAND, POUL JØRGENSEN,
NIELS OLE KJELDGAARD, AND FINN SKOU PEDERSEN*

*Department of Molecular Biology and Plant Physiology, University of
Aarhus, DK-8000 Aarhus C, Denmark*

Received 30 January 1990/Accepted 14 May 1990

The U3 region of Akv murine leukemia virus carries a 99-base-pair repeat that is associated with transcriptional enhancement in murine NIH 3T3 cells. Deletion analysis points to a critical function of a region within the repeat unit related to the recognition sequences for nuclear factor I proteins but distinct from the sites previously analyzed in related viruses. Nuclear proteins binding to the critical site were detected in NIH 3T3 cells and in mouse livers. A protein fraction binding to this site was purified from mouse livers by ion-exchange and DNA affinity chromatography and shown to have nuclear factor I properties. Mutations that caused a partial or complete reduction of the *in vitro* binding were introduced into an Akv long terminal repeat with one 99-base-pair repeat copy driving a reporter gene, and the expression activities of the mutants in NIH 3T3 cells were found to correspond to their *in vitro* binding activities. This correlation strongly supports the role of nuclear factor I proteins in Akv expression. Residual expression activity was, however, detected in mutants devoid of *in vitro* binding. This residual activity may relate to the presence of additional sequences with homology to nuclear factor I binding sites both within and outside the repeat region. The ability of these sites to bind crude and purified protein fractions with nuclear factor I activity was analyzed, and the role of the sites within and outside the repeat region for control of gene expression of Akv and related viruses is discussed.

The promoter-enhancer region in U3 of murine leukemia viruses (MuLVs) contains multiple sequence elements that interact with DNA-binding transcription factors of the host cell (33, 36, 48). Variations in the U3 structure among isolates of murine leukemia viruses can determine differences in their tissue specificity of transcription and oncogenesis (1, 3, 4, 17, 29, 31, 38, 46, 50, 53). In the cases investigated (15, 17, 25, 29, 32, 44), the structures within U3 that are critical for these differences have been located to tandem repeats about 200 to 400 base pairs upstream of the cap site at the U3-R border. In a previous study (32), we analyzed the overall functional organization of the enhancer region in a prototype virus, Akv MuLV, with a U3 region characterized by a 99-base-pair tandem repeat. In mouse fibroblast cells we found that sequences within the repeats were critical for enhancer function. Sequences outside the tandem repeats, however, were found to contribute cooperatively to enhancer function, a contribution that may be essential when one of the 99-base-pair tandem repeat units has been deleted (32).

In this study we analyzed a site in the 99-base-pair repeat of crucial significance for function of the Akv U3 in transcription. We present biochemical and genetic evidence that this site binds proteins of the nuclear factor I (NF-I) family and that the binding plays a role for expression in NIH 3T3 fibroblasts. The NF-I proteins represent a family of DNA-binding proteins (12, 39, 43), with roles in transcription and DNA replication that have been identified in birds and mammals (5, 20, 26, 30, 35, 41, 42). The consensus DNA recognition sequence is characterized by two symmetrical half-sites with a TGG motif separated by less critical spacer nucleotides (9, 16, 20, 26, 36, 37), but elements with only one

half-site have also been shown to be functional (2, 11, 12, 30, 39). Different sites may show functional differences, depending on the strength and exact specificity for binding of NF-I proteins (36) and on the possibility for interactions with other DNA-binding proteins (12).

The tandem repeat regions of different MuLVs differ with respect to the number of sites that may be implicated in NF-I binding and with respect to the exact structure and sequence environment of these sites (33, 36, 48), and sites corresponding to that studied here have not been previously analyzed. On the basis of results of biochemical and genetic studies, we discuss the possibility that expression is regulated by multiple NF-I-binding sites located both within and outside the repeat region.

MATERIALS AND METHODS

Expression vector plasmids. The transcription unit in the expression vector plasmids contains an Akv long terminal repeat with a stretch of viral 5' untranslated sequences linked to a chloramphenicol acetyltransferase reporter gene and simian virus 40 polyadenylation signals. The nucleotide sequence of the plasmids in the relevant portion of U3 is shown in Fig. 1. Plasmid p1-99 was derived from the standard Akv expression vector plasmid pAkv6-cat (8, 32, 38) by digestion with *Apa*I and self-ligation to generate an Akv U3 region carrying only one copy of the 99-base-pair tandem repeat. Plasmids pDD3, pD10, pD11, pDD4, pDD5, pD12, pD13, pDD7, pDD6, pD15, pD16, and pD17 are derivatives of p1-99 harboring previously described deletions in U3 (32), and pD19 is a parallel derivative described in Fig. 1. The plasmids pAkv8.1 (enhancer negative, promoter positive) (32) and pL6-cat (enhancer negative, promoter negative) (40) are pAkv6-cat derivatives with long terminal repeat deletions covering base pairs -441 to -89 and base pairs -441 to +27, respectively. Plasmids pM1, pM2, pM3, and pM4 were derived from p1-99 by site-specific

* Corresponding author.

[†] Present address: Department of Molecular Oncology and Virology, Roche Institute of Molecular Biology, Nutley, NJ 07110.

mutagenesis around base pairs -240 to -260 in U3. The mutagenesis was carried out by insertion of double-stranded DNA fragments harboring the mutations by using the unique *Pst*I and *Apa*I sites in the U3 region. The inserted fragments were generated by ligation of a chemically synthesized fragment for each mutation (*Msp*I-*Apa*I or *Rsa*I-*Apa*I) and a nonmutated, plasmid-derived fragment (*Pst*I-*Msp*I or *Pst*I-*Rsa*I). The structure of the plasmids in the relevant regions was confirmed by nucleotide sequence analysis.

Transfections and assays for transient gene expression. Transfections were performed by the calcium phosphate precipitation method (14) as previously described (32). The transient expression activity of the plasmids was determined as previously described (32). The procedure measures the activity of chloramphenicol acetyltransferase (13) under conditions that ensure substrate conversions within the time-linear range and corrects for variations in transfection efficiency by measurement of the activity of β -galactosidase expressed from a cotransfected plasmid (18). The results were corrected for background values from mock transfections, and for each transfection series the transient expression activities were calculated relative to the mean activity obtained for parallel pAkV6-cat plasmid transfections, arbitrarily set at 100.

DNA probes. The nucleotide sequences of the DNA probes are summarized in Fig. 1 and 6. DNA probes for electrophoretic mobility shift assays, DNase I footprinting, and methylation interference experiments were made in the following way. Chemically synthesized oligonucleotides (5 to 10 pmol) were radiolabeled with T4 polynucleotide kinase and [γ - 32 P]ATP (carrier free, crude; ICN Radiochemicals), heated to 90°C for 5 min, hybridized to the complementary oligonucleotide by slow cooling to room temperature, and purified by gel electrophoresis.

The DNA probes wt, M1, M2, M3, and M4 (Fig. 1) were made in several steps as follows. The oligonucleotides 5'-CCTCTAGACAGAGAGGCTGGAAAGTACCGGGAC TAGGGCCAAACAGGATATCTGTGG-3' (wt), 5'-CC TCTAGACAGAGAGGCTGGAAAGTACCGCCACTAGG GCCAAACAGGATATCTGTGG-3' (M1), 5'-CCTCTAGA CAGAGAGGCTGGAAAGTACCGGGACTAGGGGGAAA CAGGATATCTGTGG-3' (M2), 5'-CCTCTAGACAGAGA GGCTGGAAAGTACCGGGTTAGGGCCAAACAGGA TATCTGTGG-3' (M3), and 5'-CCTCTAGACAGAGAGGC TGGAAAGTACCGCCACTAGGGGGAAACAGGATATC TGTGG-3' (M4) (mutations underlined) were end labeled with T4 polynucleotide kinase, hybridized to the oligonucleotide 5'-CCTCTAGATTTCTGGGGACCATCTGTTCTTG GCCCTGGGCCGGGGCCCTAGTGCTTGACCACAGAT ATCCTGT-3' (the complementary nucleotides are underlined), incubated with reverse transcriptase and the four deoxynucleotide triphosphates for 1 h at 37°C to generate a completely double-stranded oligonucleotide, digested with *Eco*RV, and purified by gel electrophoresis. The structures of the resulting completely double-stranded probes wt, M1, M2, M3, and M4 are shown in Fig. 1.

Protein purification. A nuclear extract was prepared from approximately 25 fresh mouse livers (kindly donated by J. Tønnes Nielsen) as described previously (2), except that the following proteinase inhibitors were added: soybean trypsin inhibitor (50 μ g/ml), aprotinin (30 μ g/ml), benzamide (1 mM), leupeptin (20 μ g/ml), and phenylmethylsulfonyl fluoride (0.1 mM).

The crude nuclear extract (approximately 15 ml) was diluted to 100 mM KCl with buffer A (20 mM HEPES [N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid] pH

8.0), 2 mM MgCl₂, 10% glycerol, 0.2 mM EDTA, 1 mM dithiothreitol) and applied onto a heparin-Sepharose column (1.5 by 10 cm). The proteins were eluted from the heparin column with a linear gradient of KCl (100 to 500 mM) in buffer A, and the fractions containing DNA-binding activities were identified in this and in all subsequent steps by electrophoretic mobility shift assays with probe 82 (Fig. 1). The active fractions were pooled, diluted to 100 mM KCl with buffer A containing 0.1 mM phenylmethylsulfonyl fluoride, applied onto an FPLC Mono Q Column (Pharmacia Biotechnology, Uppsala, Sweden), and eluted with a linear gradient of KCl (100 to 1,000 mM) in buffer A.

The active fractions were again diluted to 100 mM KCl by using buffer A containing 0.1 mM phenylmethylsulfonyl fluoride, 0.1% Nonidet P-40, and 10 mM MgCl₂ and loaded onto an affinity column with matrix-bound double-stranded DNA (21). The column carried double-stranded DNA corresponding to probe 82 (Fig. 1), and the bound protein was eluted with buffer A containing 600 mM KCl, 0.1% Nonidet P-40, and 10 mM MgCl₂. The active fractions were pooled and diluted again for subsequent chromatography cycles on the same affinity column.

The final yield from 25 livers was in the range of 0.1 to 1 μ g of protein, and we estimate that the purification scheme results in a yield of about 20% and a purification of the specific binding activity of about 10⁴-fold from the crude nuclear extract.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis and silver staining analysis of purified protein fractions were performed by standard techniques (24, 49). Size exclusion chromatography on a Superose 12 column (Pharmacia) under nondenaturing conditions was carried out in buffer A with 0.1% Nonidet P-40 and 10 mM MgCl₂, and the protein fractions with binding activity were identified by the electrophoretic mobility shift assay with probe 82.

Crude nuclear protein extracts from NIH 3T3 cells grown in Dulbecco modified Eagle medium with 10% newborn calf serum were made as described by Dignam et al. (10), except that protease inhibitors were included as described above.

Electrophoretic mobility shift assay. The radiolabeled DNA probe (5,000 to 10,000 cpm) was mixed with poly(dI)-poly(dC), or with specific nonradioactive competitor DNA as specified, in 25 μ l of 50 mM KCl-20 mM HEPES (pH 7.5)-5% glycerol-10 mM MgCl₂-1 mM dithiothreitol-1 mM EDTA. A protein sample in 0.5 to 1 μ l was added, and the mixture was incubated for 10 to 20 min at room temperature. Then 5 μ l of 50% glycerol with 0.01% bromophenol blue was added, and the mixture was loaded onto a 5% polyacrylamide gel. The electrophoresis buffer for mobility shift assays was 0.19 M glycine-25 mM Tris-0.2 mM EDTA. After electrophoresis, the gel was dried and autoradiographed by using intensifying screens. The bands were cut out of the dried gel after exposure, and the amount of radioactivity was determined by Cerenkov counting.

DNase footprinting and methylation interference. DNase I footprinting was performed as a preparative mobility shift assay by using fivefold larger quantities of all components. Immediately before loading onto the gel, the mixture was digested with DNase I (0.05 U; Sigma Chemical Co., St. Louis, Mo.) for 1 min. After gel electrophoresis, the wet gel was exposed for 1 to 2 h, and the DNA in the retarded and nonretarded bands was extracted and run on a 12% polyacrylamide sequencing gel.

For the methylation interference experiment, a preparative mobility shift assay was performed with 100,000 cpm of a methylated probe (47), and DNA bands in the wet gel were

FIG. 1. Structure of plasmids and DNA probes. All plasmids are derivatives of the expression vector plasmid pAkv6-cat (32), in which a complete Akv long terminal repeat drives the expression of a chloramphenicol acetyltransferase (CAT) gene. The base pair numbers refer to the cap site of pAkv6-cat. Only the structures in the part of U3 where the plasmids differ from pAkv6-cat are shown. Plasmid p1-99 differs from pAkv6-cat by deletion of one copy of the 99-base-pair tandem repeat sequences. Plasmids pDD3, pD10, pD11, pDD4, pDD5, pD12, pD13, pDD7, and pDD6 all differ from p1-99 by further deletion of a single stretch of nucleotides shown by the broken lines. In these plasmids, the downstream deletion borders are located at various points in the 99-base-pair repeat sequence as indicated, and the deletions extend all through the upstream part of U3 into various points of the vector sequences (32). Plasmids pD15, pD16, and pD17 differ from p1-99 by internal deletions in the 99-base-pair repeat sequence as shown. Plasmid pD19 differs from p1-99 by deletion of sequences from the middle of the 99-base-pair repeat unit as shown and downstream through base pair -119. Plasmids pM1, pM2, pM3 differ from p1-99 by a single dinucleotide substitution around base pairs -250 to -240 as shown by the underlined bases. Plasmid pM4 carries the combined substitutions of pM1 and pM2. The chloramphenicol acetyltransferase activity measurements give the activities as means and standard deviations for at least six independent transfections into NIH 3T3 fibroblasts. The values are given relative to the mean activity for parallel transfections with the pAkv6-cat plasmid, arbitrarily set at 100. Plasmids pAkv8.1 (deletion from base pairs -441 to -89) and pL6-cat (deletion from base pairs -441 to +27) give chloramphenicol acetyltransferase activities of 0.2 ± 0.2 and 0.0 ± 0.0 , respectively. The perfectly double-stranded DNA probes are given as the sequence of one strand listed below the homologous Akv sequence. Eight base pairs at one end of the wt, M1, M2, M3, and M4 probes with no homology to Akv are shown as lowercase letters. The mutations in M1, M2, M3, and M4 and in 1102, 1103, 1104, and 1105, corresponding to pM1, pM2, pM3, and pM4, respectively, are underlined.

localized as described above. The DNA was electroeluted and treated with piperidine for 1 h at 95°C and electrophoresed on a 12% sequencing gel with size markers generated from the same probes by Maxam-Gilbert guanosine and adenosine cleavage reactions (34).

RESULTS

Deletion analysis of the 99-base-pair repeat unit. We wanted to identify *cis*-acting elements in the U3 region of Akv MuLV that are of critical significance for transcription in mouse fibroblasts. In our previous study of the overall

functional organization of this U3 region, the effect of deletions in U3 upon gene expression was analyzed (32). To identify important *cis*-acting elements, we inspected the nucleotide sequences of a number of critical deletion mutant plasmids, all affecting nucleotide sequences of the 99-base-pair repeat unit. Figure 1 shows the nucleotide sequences of the repeat unit and adjacent DNA, given in the upper line for plasmid p1-99, in which exactly one copy of the repeat sequence has been deleted, and below for 13 deletion mutants carrying less than one complete copy of the repeat unit. The transient expression activity of the linked chloramphenicol acetyltransferase gene is given relative to that of an

FIG. 2. Analysis of protein-DNA complexes formed with an NF-I site 1 probe. NIH 3T3 crude nuclear extract (lanes 1 and 3), liver crude nuclear extract (lanes 2 and 4), and affinity-purified extract from liver (lanes 5 and 6) were analyzed for binding to DNA probe 82 in the presence of 0.5 μ g of poly(dI)-poly(dC) by the electrophoretic mobility shift assay. The assays represented by lanes 1, 2, and 6 were performed as described in Materials and Methods. The protein fractions used for the assays represented by lanes 3, 4, and 5 were treated with trypsin (final concentration, 8 μ g/ml) for 10 min at room temperature; then trypsin activity was inhibited by addition of trypsinol (final concentration, 250 μ g/ml), and the protein samples were used in the electrophoretic mobility shift assay.

intact U3 with two copies of the tandem repeat unit, arbitrarily set to 100. By inspection of the series of 5' deletions, we note a sequential reduction in activity by deletion of sequences through base pairs -270 to -250, whereas further deletions through the 99-base-pair sequence have no significant effect upon expression activity. The critical sequences around base pair -250 are related to the recognition sequences for NF-I proteins (20). This site and a more downstream sequence that has also been proposed to represent a NF-I recognition site in related viruses (33, 36, 48) are marked in Fig. 1 as NF-I site 1 and NF-I site 2, respectively.

DNA-binding proteins with NF-I properties. To identify possible sequence-specific DNA-binding proteins of importance for the critical function of the nucleotides around base pair -250, double-stranded DNA probes covering this region (DNA probes 82, 794, and 933; Fig. 1) were used in mobility shift assays with nuclear extracts of NIH 3T3 cells. A prominent complex was formed for nucleotide probes containing sequences between base pairs -240 and -260, as shown in Fig. 2 for probe 82. To purify proteins able to form this complex in sufficient quantities, we started from a mouse liver nuclear extract. In most cases the DNA complexes formed with the liver proteins showed a somewhat faster electrophoretic mobility than did the NIH 3T3 complexes (Fig. 2). Different relative amounts of the different complexes were, however, formed with separate liver extracts. After mild proteolysis by incubation with trypsin, complexes with similar mobilities were observed in all cases (Fig. 2). The differences in complex mobilities may therefore relate to variable proteolytic degradation of the proteins in the crude nuclear extracts or possibly also to differences in accessory proteins associated with the complexes. We therefore believe that similar DNA-binding activities are present in the nuclear extracts from the two sources, an assumption that is further supported by analysis of binding activities toward different probes and by footprinting studies (data not shown). Also, our binding studies with mutated competitor DNAs (see below and Fig. 6) did not reveal any difference in binding specificity between the NIH 3T3 and liver activities.

FIG. 3. Analysis of the purified protein fraction. The DNA affinity-purified fraction of mouse nuclear proteins with binding activity toward NF-I site 1 was obtained as described in Materials and Methods. (A) Fractionation of the purified protein fraction on a Superose 12 column under non-denaturing conditions. The binding activity of the fractions was assayed in the electrophoretic mobility shift assay with probe 82. An autoradiogram of the part of the gel containing the complexes is shown, and the positions of proteins of known molecular weights on the same column are marked. (B) Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis of the purified protein fraction. The positions of marker proteins of known molecular masses (kilodaltons) in parallel lanes are marked by arrows.

A protein fraction that showed binding to DNA probes from this region (Fig. 2) was obtained from mouse liver nuclear extracts by using heparin-Sepharose, cation-exchange, and DNA affinity chromatography as described in Materials and Methods. The protein fraction subjected to the most extensive characterization (Fig. 3, 4, and 5) was derived from an extract with apparently little proteolytic degradation.

A rough estimate of the size of the protein with DNA-binding activity was obtained by size-exclusion chromatography on a Superose 12 column (Pharmacia) under non-denaturing conditions (Fig. 3A) in the presence of protein markers of known molecular weight. The specific DNA-binding activity was detected by the electrophoretic mobility shift assay with probe 82. We note binding activities in the size range between 44 and 67 kilodaltons. The protein species forming the two characteristic protein-DNA complex bands observed in the mobility shift assay can be separated on the size-exclusion column, indicating that proteins with different molecular weight interact with the binding site *in vitro*. The polypeptide composition of the purified protein fraction was analyzed by using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 3B). The purified fraction contained at least four major bands in the 50- to 70-kilodalton range, a molecular mass pattern somewhat similar to that reported for NF-I isolated from HeLa cells (20, 41).

To elucidate the precise binding specificity of the affinity-purified protein fraction, DNase I protection and methylation interference experiments were performed. For the DNase I protection assay, a DNA-protein complex was formed and digested briefly with DNase I before separation of the complexed and noncomplexed DNA by electrophoresis on non-denaturing gels. The DNA bands were isolated and rerun on a denaturing gel for identification of DNase I cleavage sites (Fig. 4B). With probe 136 spanning one full 99-base-pair repeat segment, two protected areas can be found, each containing a match of 7 of 8 base pairs to the consensus sequence for binding of NF-I proteins 5' TGG^A/

FIG. 4. DNase I footprinting and methylation interference analysis. The analysis was carried out with samples of the affinity-purified protein fraction analyzed in Fig. 3. (A) DNase I footprint analysis with probe 136 (Fig. 1), spanning a full 99-base-pair repeat sequence. Lanes: 1, free DNA probe; 2, probe-protein complex. Only the coding strand was radioactively labeled. The nucleotides in the protected regions are indicated. The upper site corresponds to NF-I site 2 of Fig. 1, and the lower site corresponds to NF-I site 1. The extent of the protection at the upstream border of NF-I site 1 is not clearly defined and may not include the first four nucleotides indicated at the bottom. (B and C) Methylation interference analysis. The analysis was conducted as described in Materials and Methods with probe 94. In experiment B the coding strand was radioactively labeled, and in experiment C the noncoding strand was radioactively labeled. Lanes: 1 and 4, noncomplexed probe; 2, probe from the complex with the faster migration; 3, probe from the complex with slower migration. Guanosine residues that show methylation interference are marked by filled circles (●) and one residue that shows weaker interference is marked by an open circle (○).

C_N GCCAA 3' (16). Likewise, for the methylation interference experiment, a preparative mobility shift experiment was set up, this time with methylated probe 94. The nonretarded band and the two retarded bands were isolated after nondenaturing gel electrophoresis, and the DNA was analyzed on a denaturing gel after treatment with piperidine (Fig. 4C). On the coding DNA strand, the interfering methylations were observed at all three G residues that matched the consensus sequence (base pairs -243, -250, and -251 in Fig. 1). On the noncoding strand the GG sequence (base pairs -241 and 242 in Fig. 1) interacted with protein binding, and a weaker interaction was found in some experiments for the G at base pair -248, representing a nonhomology with the consensus sequence. Identical interference patterns were observed for both the faster- and slower-moving complexes. The patterns of methylation interference and DNase I protection are in accordance with results obtained for NF-I proteins analyzed in other circumstances (6, 9, 20, 41).

Effect of mutations in NF-I-binding site. We wanted to introduce mutations in the sequence of the NF-I recognition

FIG. 5. Electrophoretic mobility shift assays with wild-type and mutated probes. Approximately 50 pg of radioactively labeled DNA probes (M1, M2, M3, M4, and wt; Fig. 1) were mixed with various amounts of nonradioactive probe 136. The mixtures were analyzed by the electrophoretic mobility shift assay with samples of the affinity-purified protein fraction analyzed in Fig. 3. (A) Autoradiograms of the part of the gel containing the protein-DNA complexes. Lanes 1 through 6 represent assays with 0, 50, 100, 200, 500, and 1,000 pg of nonradioactive probe 136, respectively. (B) Graphic representation of the percentage of the radioactivity in the probe retained in the complexes. DNA probes: wt (□), M1 (●), M2 (○), M3 (△), M4 (▲).

sequence around base pair -250 to test the effect of such mutations both on protein binding in vitro and on gene expression in cultured cells. The four mutations chosen for this study represent three dinucleotide substitutions (pM1, pM2, and pM3) and one substitution of two dinucleotides (pM4) (Fig. 1; see Materials and Methods). The substitutions of pM1, pM2, and pM4 involve nucleotides that represent a perfect match with the consensus sequence, whereas the pM3 change was introduced at a position suspected to be less critical.

Five different synthetic DNA probes (Fig. 1) carrying either the wild-type sequence (wt probe) or the four mutations (probes M1, M2, M3, and M4) were used for analysis of binding of the purified protein fraction. To obtain further information on the quantitative aspects of DNA binding and to confirm that proteins bound to the mutated probes would also bind to the wild-type sequence, assays were performed with various amounts of nonradioactive competitor DNA of the wild-type sequence (probe 136). The complexes were autoradiographed (Fig. 5A), and the probe retardation efficiency was determined (Fig. 5B). The mutation in the downstream half-site of the consensus NF-I binding sequence, as found in pM2 and pM4, completely abolished protein binding, indicating that this half-site has a critical function. The mutation in the upstream half-site (as in pM1) led to some reduction in binding, whereas the mutation in pM3 had only a minor effect. The addition of excess DNA of the wild-type sequence led to competition in complex for-

FIG. 6. Effect of mutations in DNA-binding competition assays. Approximately 0.06 pmol of radioactively labeled probe of the wild-type sequence (probe 933) was mixed with various amounts of the five nonradioactive probes, 1102 (M1), 1103 (M2), 1104 (M3), 1105 (M4), and 933 (wt). The mixtures were analyzed by the electrophoretic mobility shift assay as described in Materials and Methods with NIH 3T3 crude nuclear extract or affinity-purified extract from mouse livers. Each assay with the NIH 3T3 extract contained 0.5 μ g of poly(dI)-poly(dC). Lanes: 1, assays without specific competitor DNA; 2 to 6, assays with a 10-, 50-, 100-, 500-, and 1,000-fold molar excess of specific competitor DNA, respectively.

mation for the M1 and M3 probes as well as for the wild-type probe. For the two mutated probes pM1 and pM3, however, we noted an increase in complexed radioactive DNA upon the addition of smaller amounts of competitor DNA, a reproducible phenomenon for which we have no explanation (Fig. 5).

To study the effect of the same mutations on the binding activity under different experimental conditions, we performed electrophoretic mobility shift assays with a probe of wild-type sequence (probe 933, Fig. 1) in the presence of various concentrations of competitor DNAs of mutated or wild-type sequence (probes 1102, 1103, 1104, 1105, and 933; Fig. 1). This type of experiment was performed both with the affinity-purified liver fraction and with a crude nuclear extract from NIH 3T3 cells.

The mutations had the same effect in the assays with the purified liver fraction and in the assays with the crude nuclear extract from NIH 3T3 cells (Fig. 6). The competitor DNAs with the M2 or M4 mutations did not affect protein binding to the wild-type probe, even in a 1,000-fold molar excess, whereas strong competition was observed with a competitor of the wild-type sequence (Fig. 6). The probes of M1 or M3 sequence showed intermediate competitor activity. These conclusions, derived from inspection of the autoradiograms, are supported by quantitative analysis of the amount of radioactivity in the bands representing complexed

probe and free probe in two independent experiments (data not shown).

The results of the competition experiments therefore indicate that the different relative effects of the mutations observed in Fig. 5 are valid for NIH 3T3 cells as well as for liver cells. The difference between the M1 and M3 sequences was not resolved by this competition assay.

The effect of the mutations in the plasmids pM1, pM2, pM3, and pM4 were studied by comparing their activity with that of p99-1 in a transient expression assay. All mutations caused a reduction in expression of chloramphenicol acetyltransferase activity (Fig. 1). The activity of the pM3 plasmid was about 30% of the p1-99 value, whereas the activities of the remaining three plasmids ranged from 20 to 10% of the p1-99 activity. A statistical analysis with Student's *t* test showed that all five values (for p1-99, pM3, pM1, pM2, and pM4) were significantly different ($P < 0.01$). The relative effect of the mutations on the expression level corresponded to their effect on the *in vitro* binding of the purified protein fraction with NF-I-like properties, with the exception that the difference between the M2 dinucleotide substitution mutant and the M4 double dinucleotide substitution mutant could not be resolved in the protein binding assay. These results therefore strongly support the notion that binding of NF-I proteins to this site is important for expression in the cells.

The values observed for the pM1, pM2, and pM4 plasmids were significantly higher than the values for the deletion mutants that lacked the complete NF-I site 1 (plasmids pDD5, pD12, pD13, pDD7, pDD6, pD15, pD16, and pD17; Fig. 1). This may indicate that sequences immediately downstream of NF-I site 1 are important or that exact positioning of sequences upstream of the NF-I site plays a role.

The second NF-I site in the 99-base-pair repeat unit. The nucleotide sequence termed NF-I site 2 in Fig. 1 has previously been proposed to represent an NF-I-binding site in murine leukemia viruses (33, 36, 48). Even though this site was intact in pM1, pM2, pM3, and pM4, these plasmids showed a severe reduction in expression activity. More than one NF-I-binding site may therefore be required, or a functional NF-I site in the natural sequence environment of NF-I site 1 may be essential. These possibilities are supported by the low activity of plasmids pD15 and pD16 (Fig. 1). The deletions of these two plasmids bring NF-I site 2 into the upstream sequence environment of NF-I site 1, except for the additional removal of one (pD16) or two (pD15) base pairs. That NF-I site 2 was less important than NF-I site 1 for expression in our assay was supported by results obtained with plasmid pD19 (Fig. 1), which showed an activity similar to that of p1-99. In pD19, NF-I site 2 and adjacent sequences in the 99-base-pair unit were deleted together with 55 base pairs downstream of the repeat unit. Although the activity of this plasmid may be affected by alterations of promoter-enhancer distances, the result clearly demonstrates that NF-I site 2 is not absolutely required for expression.

Potential NF-I-binding sites outside the repeat unit. In a previous study (32) we found that the function of one 99-base-pair repeat sequence alone is dependent upon its normal sequence environment, whereas the function of two 99-base-pair sequences in tandem is independent of sequence environment. Since we have found a critical function of binding sites for NF-I proteins, and since two repeats contain four NF-I sites and one contains only two, we speculated that the sequences in the Akv U3 outside the tandem repeats may contain NF-I sites that are of impor-

FIG. 7. Probes for sequences in Akv U3 with partial homology to NF-I-binding sites. The map corresponds to the U3 sequence of Akv. Only one copy of the 99-base-pair tandem repeat is shown (arrow). The nucleotide numbers refer to the cap site as in Fig. 1. The five sites were identified by partial homology to the NF-I-binding site consensus sequence TGG^A/C_N₃GCCAA (16). The dots above the sequences indicate the consensus matches. Sites 1 and 2 are also marked in Fig. 1. The double-stranded sequences given correspond to the DNA probes 790, 792, 794, 796, and 798.

tance when only one 99-base-pair sequence is present. We therefore searched for possible NF-I-binding sites in U3 upstream or downstream of the repeats. The NF-I-binding sites are typically characterized by an inverted repeat, but sequences containing only one half-site of a repeat have also been implicated in binding (2, 11, 12, 30, 39). The published NF-I consensus recognition sequences include TGG^A/C_N₃GCCAA (16), PyTGG^A/C_N₃^A/TGCCA (37), TGG^A/C_N₃^A/T_N₃^A/TGCCAA (26), TTGGCTN₃AGCCAA (20), and TTGGAN₃GCCAAT (9) and the ideal recognition sequence PyPy-PyTGGCACAGTGCCAPuPuPu (36). We found two sequences upstream and one sequence downstream of the 99-base-pair repeat that exhibited good homology to the NF-I consensus (Fig. 7).

Synthetic double-stranded probes (probes 790, 792, 794, 796, and 798) corresponding to these sequences (Fig. 7) were used to assess protein binding. The probes were made short (18 base pairs) to minimize the possibility of complexes due to interactions with non-NF-I proteins. A series of parallel band shift assays was first carried out with the affinity-purified fraction. A weak binding activity was detected for probe 792, and no or little binding was observed for the two remaining probes (probes 790 and 798) (data not shown). As expected, the parallel experiments demonstrated clear binding to the sequences of NF-I sites 1 and 2 within the repeats (probes 794 and 796, respectively).

To further investigate the large differences in the protein-binding abilities of the five probes with NF-I binding site homology, we performed binding studies with crude nuclear extracts or with semipurified NF-I preparations. The protein fractions used were liver crude nuclear extract (Fig. 8A) and pooled top fractions from a heparin column chromatography (Fig. 8B). The binding capacity was assessed without or with two different concentrations of poly(dI)-poly(dC). Protein-DNA complexes could be detected with all probes in the absence of poly(dI)-poly(dC). Complexes of the same mobilities occurred with the probes within the repeats (probes 794 and 796), although less complex was formed with probe 796. The complexes formed with probe 792 were less efficiently competed with by poly(dI)-poly(dC) than were the complexes formed with the other two probes (790 and 798). Furthermore, the positions of shifted bands for probe 792 were more like the bands occurring for probes 794 and 796, representing repeat recognition sites, although the same degree of complexity did not show up.

The complexes observed for probes 790 and 798 are most likely due to unspecific protein binding, since the bands disappeared after competition with low concentrations of poly(dI)-poly(dC) (0.1 μ g for probe 798 and 0.5 μ g for probe

790). Further studies will be required to determine whether the poly(dI)-poly(dC)-resistant complexes formed with probe 792 represent specific protein binding and whether the proteins belong to the NF-I family.

DISCUSSION

In the present study we analyzed a sequence in the tandem repeat of the U3 region of Akv murine leukemia with the properties of a NF-I-binding site and provided evidence that binding of NF-I-like proteins to this site plays a role for transcriptional activity of this U3 in mouse fibroblasts.

Homology to NF-I recognition sites can be found in the repeat modules of a number of other MuLVs. The viruses within this group differ with respect to the organization and

FIG. 8. Analysis of protein binding to probes representing possible NF-I-binding sites in the U3 region of Akv. A crude nuclear protein extract from mouse livers (A) and pooled top fractions from a subsequent heparin column fractionation step (B) were tested in the electrophoretic mobility-shift assay with the five double-stranded DNA probes shown in Fig. 6. The assays were performed as described in Materials and Methods with the following additions (lanes): 1, 1 μ g of poly(dI)-poly(dC); 2, 0.5 μ g of poly(dI)-poly(dC); 3, 0.1 μ g poly(dI)-poly(dC); 4, no addition.

the primary nucleotide sequences of repeat modules, and their repeat regions show differences in the number of sites with homologies to the NF-I consensus recognition sequence and in the nucleotide sequences of these sites. Different sites in one virus or sites in different viruses are therefore not necessarily functionally equivalent.

The Akv repeat region contains four elements with homology to the NF-I consensus recognition sequence, two in each tandem repeat unit represented by the site 1 and site 2 sequences in Fig. 1. In this study we focused on the function of site 1. The site 1 sequence is also found in the repeat regions of the Akv-related viruses SL3-2 (8), Gross passage A (52), OA-1 (27), FBJ MuLV (51), and Soule MuLV (7), in all cases in a homologous position and in only one copy. This site has not previously been subjected to functional studies in any of the viruses. The site 2 sequence is found in the repeat region of all viruses that are closely related to Akv. In SL3-3 virus the ability of this sequence and mutants thereof to bind proteins in nuclear extracts suggests that it represents a NF-I-binding site (36). Three copies of this sequence are found in SL3-3 (28) and in FBJ MuLV (51), two copies are found in OA-1 (27) and SL3-2 (8), and one copy is found in Gross passage A virus (52). The Moloney (45) and Friend (23) MuLVs both contain an element nearly identical to site 2 in their tandem repeat sequences (23, 33, 45, 48). The function of these sequences in the Moloney and Friend MuLVs as NF-I-binding sites has been subject to biochemical and genetic studies (33, 48). In addition, the Moloney MuLV tandem repeat contains another potential NF-I site (5'-TGAATATGGGCCAA-3') that is more different from the sites in Akv (45). This site is located in the promoter-distal part of the repeat unit in a position corresponding to the NF-I site 1 of Akv. It has been found (44) that this region can be deleted from a Moloney MuLV U3 with one tandem repeat copy without significant effect on its ability to drive transient expression in fibroblasts. Potential NF-I sites of different structures are also found in the repeat region of MCF viruses like MCF 247 (5'-TGGGACAGGGGCCAA-3') (22) and of a radiation leukemia virus isolate (5'-TAGACAT AGGCCAA-3') (19).

This comparison stresses the complex differences among the individual viruses with respect to the number and organization of sites that may be implicated in NF-I binding. Interestingly, all the sites discussed above contain a downstream half-site with the sequence 5'-GGCCAA-3' but differ substantially in their upstream half-sites. We also note the presence of two site 1 elements in Akv, whereas the other viruses contain one or no copies of this sequence.

Our first indication for a role of NF-I site 1 in transcription directed by the U3 of Akv came from deletion analysis, which pointed to a critical function of sequences around base pair -250 within the 99-base-pair repeat sequence. The nucleotide sequence in this region matches the NF-I consensus recognition sequence 5'-TGG^A/C_NGGCCAA-3' (16), except for a difference in the first base pair. From mouse liver nuclei we purified a protein fraction with specific binding activity toward the Akv sequence in this region. The purification procedure included DNA affinity chromatography on matrix-bound double-stranded DNA of the Akv sequence in this region. NF-I proteins purified from HeLa cell fractions were shown to consist of a family of polypeptides with molecular weight range from 52,000 to 66,000 (20, 41). The purified fraction of mouse nuclear proteins used in the present study shows a similar molecular weight pattern, and our DNase I footprint and methylation interference results are in accordance with published results with purified NF-I

preparations (6, 9, 20, 41). We therefore believe that our purified proteins belong to the NF-I family.

To test whether the critical function of the sequences around base pair -250 is in fact due to their ability to bind NF-I proteins, we used site-specific mutagenesis. Three dinucleotide substitution mutants (pM1, pM2, and pM3) and one double dinucleotide substitution mutant (pM4) with the combined mutations of pM1 and pM2 were first analyzed for their ability to bind NF-I proteins in vitro in the electrophoretic mobility shift assay and then for their transcriptional function in mouse fibroblasts. The pM1 mutation represents a change in the base pairs of the two methylation-sensitive G residues of the upstream half-site from 5'-CGGGAC-3' to 5'-CGCCAC-3' and the pM2 mutation a change in the base pairs of the two methylation sensitive G residues of the downstream half-site from 5'-GGCCAA-3' to 5'-GGG GAA-3'. The pM3 mutation represents the change of presumably less critical base pairs between the half-sites. The pM2 mutation in the downstream half-site was found to abolish NF-I binding in vitro, whereas the pM1 mutation in the upstream half-site allowed some binding activity. The downstream half-site, which represents a perfect match with the consensus sequence and is conserved in all of the NF-I sites discussed above, therefore seems critical for binding. The pM3 mutation led to a binding with an activity between those of the wild-type sequence and the pM1 sequence.

The mutated plasmids pM1, pM2, pM3, and pM4 all showed reduced activity in the transient expression assay, when compared with that of their parent plasmid, p1-99. The relative order of expression activities was p1-99 > pM3 > pM1 > pM2 > pM4, in agreement with the in vitro binding activities. We take the covariation between in vitro binding and in vivo activity as a strong indication of a functional role of the NF-I binding in this region. However, complete impairment of NF-I site 1, as in our pM2 and pM4 mutations, still allows transcriptional activity above background levels.

This residual activity in the absence of a functional NF-I site 1 may depend upon the binding of NF-I to other sequences in the U3 region of Akv and/or upon the binding of other transcription factors. As expected from the previous studies of other viruses, our purified fraction of NF-I proteins also bound to the sequences in Akv termed NF-I site 2 (Fig. 1), and the observed DNase I protection (Fig. 4) and methylation interference patterns (data not shown) fit those expected for NF-I interactions. NF-I site 2 alone, however, is not sufficient for the residual expression activity of the plasmids with an impaired NF-I site 1, since deletions affecting site 1 all led to abolishment of activity. Specifically, we note the low expression activity of mutants pD15 and pD16, with deletions that bring NF-I site 2 into the site 1 position, and of mutant pD17, with a deletion that eliminates the site 1 sequences. Most likely additional protein-DNA interactions in this region are required, and the interpretation may be quite complex. The inability of site 2 to replace site 1 function in these deleted constructs may relate to the different structures of the two sites, or it may reflect a role of interactions with elements immediately downstream of site 1 that are deleted in pD15 and pD16 or with elements upstream of site 1 that may not interact properly with site 2 in pD17. Apparently, NF-I site 2 is not as important as NF-I site 1 for expression in NIH 3T3 cells, as suggested by the results obtained with the pD19 deletion mutant. This plasmid shows high-level expression, although it lacks NF-I site 2 and adjacent sequences.

Obviously, our deletion analysis is too crude to clearly identify specific contributions from individual binding sites

outside of NF-I site 1. Site-specific mutagenesis of NF-I site 2 and of other potential binding sites may be required in further analyses. We note, however, that three different levels of disturbance of the functional organization may be reflected in the levels of activity of pM2 and pM4 (6.8 and 4.8 relative units, respectively), in which NF-I site 1 is impaired by substitution, of pD15 and pD16 (2.6 and 3.3 units), in which NF-I site 2 with downstream sequences is transposed to the position of NF-I site 1, and of pD17 (1.7 units), in which NF-I site 1 has been deleted.

Our previous results (32) have shown that the transcriptional activity of one 99-base-pair repeat unit is dependent upon its sequence environment, indicating a requirement for cooperativity between one repeat unit and its flanking sequences in U3. To investigate the possible role of NF-I interactions for this cooperativity, we searched the Akv U3 sequences outside the repeats for homologies to the recognition sequence for NF-I. Two sites (sites 3 and 4) were found upstream of base pair -385, and one site (site 5) was found downstream of the repeats. These three sites each contain one TGG sequence and at least two matches with the TGG sequences of the other half-site. The formation of complexes with DNA probes corresponding to these sites was analyzed with affinity-purified, semipurified, and crude nuclear protein fractions. Whereas sites 4 and 5 gave rise to low-specificity complexes, site 3 produced complexes of apparently higher specificity. Furthermore, site 3 gave rise to complex mobilities almost similar to those obtained with site 1 and site 2 probes, although the binding activity of the site 3 probes was somewhat lower. Interestingly, the only differences outside the repeat region between the U3s of Akv and the related viruses SL3-2 (8), SL3-3 (28), and Gross passage A virus (52) are located in the potential NF-I-binding site 3. The Akv and Gross passage sequences both have a 6-base-pair spacer between the TGG and CCA half-sites but differ in one position of the spacer. The SL3-2 and SL3-3 sequences carry an additional base pair at this position and hence contain 7 base pairs between two half-sites. The Moloney MuLV sequence (45) also differs from the Akv sequence in this region. Additional experiments will be required to elucidate the possible regulatory function of the site 3 sequence in Akv and the variant viruses.

In summary, we have identified two sequence elements in the tandem repeats of Akv MuLV involved in NF-I binding and potential NF-I-binding sites located in U3 outside the tandem repeats and pointed to identical and divergent sites in the U3 regions of other MuLVs. A number of features may influence the exact function of each of these sites. The existence of different subclasses of NF-I proteins present in various amounts and possibly with slightly different binding specificities may affect the pattern of binding to multiple NF-I sites. Molecular cloning of the genes encoding NF-I proteins from humans (43), rats (39), and hamsters (12) confirms the existence of different subclasses of NF-I proteins that all recognize the 5'-TGG-3' sequence. The analyses predict proteins with nearly identical amino acid sequences in their NH₂-terminal half, which contains the DNA-binding domain and with divergence in the COOH-terminal regions. Proteins binding to DNA flanking the NF-I sites may also contribute to their functional diversity through protein-protein interactions affecting the DNA binding of NF-I and other proteins. Also, an NF-I functional monomer could bind to half the recognition site, 5'-NNTG-GNN-3' in a complex with another protein with affinity for a neighboring sequence, as proposed previously (12). Finally, the exact location of an NF-I site in U3 may affect its

possibility to participate in formation of a functional transcription complex.

Studies employing permutations of NF-I sites within one virus or between viruses, together with specific mutagenesis of individual sites, may provide insight into the role of these different regulatory principles. These viruses may thereby provide interesting material for studies of the role of NF-I-binding sites in regulation of gene expression.

ACKNOWLEDGMENTS

We thank Peter V. Mathiasen, Lone Højgaard, and Bente Christensen for technical assistance and J. Tønnes Nielsen for mouse livers.

This work was supported by grants from the Carlsberg Foundation, the Danish Cancer Society, Euratom (B16-086-DK), and the Danish Biotechnology Programme.

LITERATURE CITED

1. Boral, A. L., S. A. Okenquist, and J. Lenz. 1989. Identification of the SL3-3 virus enhancer core as a T-lymphoma cell-specific element. *J. Virol.* **63**:76-84.
2. Borgmeyer, U., J. Nowock, and A. E. Sippel. 1984. The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. *Nucleic Acids Res.* **12**:4295-4311.
3. Böse, Z., H.-J. Thiesen, and P. Charnay. 1986. A transcriptional enhancer with specificity for erythroid cells is located in the long terminal repeat of the Friend murine leukemia virus. *EMBO J.* **7**:1615-1623.
4. Celander, D., and W. A. Haseltine. 1984. Tissue-specific transcription preference as a determinant of cell tropism and leukemogenic potential of murine retroviruses. *Nature (London)* **312**:159-162.
5. Cereghini, S., M. Raymondjean, A. G. Carranca, P. Herbolmel, and M. Yaniv. 1987. Factors involved in control of tissue-specific expression of albumin gene. *Cell* **50**:627-638.
6. Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp. 1988. Human CCAAT-binding proteins have heterologous subunits. *Cell* **53**:11-24.
7. Corcoran, L. M., J. M. Adams, A. R. Dunn, and S. Cory. 1984. Murine T lymphomas in which the cellular *myc* oncogene has been activated by retroviral insertion. *Cell* **37**:113-122.
8. Dai, H. Y., M. Etzerodt, A. J. Bækgaard, S. Lovmand, P. Jørgensen, N. O. Kjeldgaard, and F. S. Pedersen. 1990. Multiple sequence elements in the U3 region of the leukemogenic murine retrovirus SL3-2 contribute to cell-dependent gene expression. *Virology* **175**:581-585.
9. De Vries, E., W. van Driel, S. J. L. van den Heuvel, and P. C. van der Vliet. 1987. Contact point analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. *EMBO J.* **6**:161-168.
10. Dignam, J. D., R. M. Lebovitz, and R. G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**:1475-1489.
11. Gil, G., T. F. Osborne, J. L. Goldstein, and M. S. Brown. 1988. Purification of a protein doublet that binds to six TGG-containing sequences in the promoter for hamster 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *J. Biol. Chem.* **263**:19009-19019.
12. Gil, G., J. R. Smith, J. L. Goldstein, C. A. Slaughter, K. Orth, M. S. Brown, and T. F. Osborne. 1988. Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Proc. Natl. Acad. Sci. USA* **85**:8963-8967.
13. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
14. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**:456-467.

15. Graves, B. J., R. N. Eisenman, and S. L. McKnight. 1985. Delineation of transcriptional control signals within the Moloney murine sarcoma virus long terminal repeat. *Mol. Cell. Biol.* 5:1948-1958.
16. Gronostajski, R. M., S. Adhya, K. Nagata, R. A. Guggenheimer, and J. Hurwitz. 1985. Site-specific DNA binding of nuclear factor I: analysis of cellular binding sites. *Mol. Cell. Biol.* 5:964-971.
17. Hallberg, B., and T. Grundström. 1988. Tissue specific sequence motifs in the enhancer of the leukaemogenic mouse retrovirus SL3-3. *Nucleic Acids Res.* 16:5927-5944.
18. Herbolme, P., B. Bourachot, and M. Yaniv. 1984. Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. *Cell* 39:653-662.
19. Janowski, M., J. Merregaert, J. Bonvier, and J. R. Maisin. 1985. Proviral genome of radiation leukemia virus: molecular cloning of biologically active proviral DNA and nucleotide sequence of its long terminal repeat. *J. Virol.* 55:251-255.
20. Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48:79-89.
21. Kadonaga, J. T., and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* 83:5889-5893.
22. Kelly, M., C. A. Holland, M. L. Lung, S. K. Chattopadhyay, D. R. Lowy, and N. H. Hopkins. 1983. Nucleotide sequence of the 3' end of MCF 247 murine leukemia virus. *J. Virol.* 45:291-298.
23. Koch, W., W. Zimmerman, A. Oliff, and R. Friedrich. 1984. Molecular analysis of the envelope gene and long terminal repeat of friend mink cell focus-inducing virus: implications for the functions of these sequences. *J. Virol.* 49:828-840.
24. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680-685.
25. 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.
26. Leegwater, P. A. J., P. C. van der Vliet, R. A. W. Rupp, J. Nowock, and A. E. Sippel. 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.
27. Leib-Mösch, C., J. Schmidt, M. Etzerodt, F. S. Pedersen, R. Hehlmann, and V. Erfle. 1986. Oncogenic retrovirus from spontaneous murine osteomas. II. Molecular cloning and genomic characterization. *Virology* 150:96-105.
28. Lenz, J., D. Celandier, R. L. Crowther, R. Patarca, D. W. Perkins, and W. A. Haseltine. 1984. Determination of the leukaemogenicity of a murine retrovirus by sequences within the long terminal repeat. *Nature (London)* 308:467-470.
29. Li, Y., E. Golemis, J. W. Hartley, and N. Hopkins. 1987. Disease specificity of nondefective Friend and Moloney murine leukemia viruses is controlled by a small number of nucleotides. *J. Virol.* 61:693-700.
30. 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.
31. LoSardo, J. E., L. A. Cupelli, M. K. Short, J. W. Berman, and J. Lenz. 1989. Differences in activities of murine retroviral long terminal repeats in cytotoxic T lymphocytes and T-lymphoma cells. *J. Virol.* 63:1087-1094.
32. Lovmand, S., N. O. Kjeldgaard, P. Jørgensen, and F. S. Pedersen. 1990. Enhancer functions in U3 of Akv virus: a role for cooperativity of a tandem repeat unit and its flanking DNA sequences. *J. Virol.* 64:3185-3191.
33. Manley, N. R., M. A. O'Connell, P. A. Sharp, and N. Hopkins. 1989. Nuclear factors that bind to the enhancer region of nondefective Friend murine leukemia virus. *J. Virol.* 63:4210-4223.
34. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
35. Nagata, K., R. A. Guggenheimer, T. Enomoto, J. H. Lichy, and J. Hurwitz. 1982. Adenovirus DNA 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.
36. Nilsson, P., B. Hallberg, A. Thornell, and T. Grundström. 1989. Mutant analysis of protein interactions with a nuclear factor I binding site in the SL3-3 virus enhancer. *Nucleic Acids Res.* 17:4061-4075.
37. Nowock, J., U. Borgmeyer, A. W. Püschel, R. A. W. Rupp, and A. E. Sippel. 1985. The TGGCA protein binds to the MMTV-LTR, the adenovirus origin of replication, and the BK virus enhancer. *Nucleic Acids Res.* 13:2045-2061.
38. Paludan, K., H. Y. Dai, M. Duch, P. Jørgensen, N. O. Kjeldgaard, and F. S. Pedersen. 1989. Different relative expression from two murine leukemia virus long terminal repeats in unintegrated transfected DNA and in integrated retroviral vector proviruses. *J. Virol.* 63:5201-5207.
39. Paonessa, G., F. Gounari, R. Frank, and R. Cortese. 1988. Purification of a NF1-like DNA-binding protein from rat liver and cloning of the corresponding cDNA. *EMBO J.* 7:3115-3123.
40. Pedersen, F. S., M. Etzerodt, S. Lovmand, H. Y. Dai, A. J. Bækgaard, J. Sørensen, P. Jørgensen, N. O. Kjeldgaard, J. Schmidt, C. Leib-Mösch, A. Luz, and V. Erfle. 1987. Transcriptional control and oncogenicity of murine leukemia viruses, p. 17-35. In N. O. Kjeldgaard and J. Forchhammer (ed.), *Viral carcinogenesis, Alfred Benzon Symposium 24*. Munksgaard, Copenhagen.
41. Rosenfeld, P. J., and T. J. Kelly. 1986. Purification of nuclear factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* 261:1398-1408.
42. Rossi, P., G. Karsenty, A. B. Roberts, N. S. Roche, M. B. Sporn, and B. de Crombrughe. 1988. A Nuclear Factor I binding site mediates the transcriptional activation of a type I collagen promoter by transforming growth factor- β . *Cell* 52:405-414.
43. Santoro, C., N. Mermod, P. C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. *Nature (London)* 334:218-224.
44. Schulze, F., E. Boehnlein, and P. Gruss. 1985. Mutational analyses of the Moloney murine sarcoma virus enhancer. *DNA* 4:193-202.
45. Shinnick, T., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* 293:543-548.
46. Short, M. K., S. A. Okenquist, and J. Lenz. 1987. Correlation of leukemogenic potential of murine retroviruses with transcriptional tissue preference of the viral long terminal repeats. *J. Virol.* 61:1067-1072.
47. Siebenlist, U., and W. Gilbert. 1980. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. *Proc. Natl. Acad. Sci. USA* 77:122-126.
48. Speck, N. A., and D. Baltimore. 1987. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol. Cell. Biol.* 7:1101-1110.
49. Switzer, R. C., III, C. R. Merrill, and S. Shifrin. 1979. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal. Biochem.* 98:231-237.
50. Thiesen, H.-J., Z. Bösze, L. Henry, and P. Charnay. 1988. A DNA element responsible for the different tissue specificities of Friend and Moloney retroviral enhancers. *J. Virol.* 62:614-618.
51. Van Beveren, C., F. van Straaten, T. Curran, R. Müller, and I. M. Verma. 1983. Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini. *Cell* 32:1241-1255.
52. Villemur, R., E. Rassart, L. DesGroseillers, and P. Jolicœur. 1983. Molecular cloning of viral DNA from leukemogenic Gross passage A murine leukemia virus and nucleotide sequence of its long terminal repeat. *J. Virol.* 45:539-546.
53. Yoshimura, F. K., B. Davison, and K. Chaffin. 1985. Murine leukemia virus long terminal repeat sequences can enhance gene activity in a cell-type-specific manner. *Mol. Cell. Biol.* 5:2832-2835.