4-1993

Two Factors that Bind to Highly Conserved Sequences in Mammalian Type C Retroviral Enhancers.

Nancy R. Manley
Massachusetts Insitute of Technology

Mary M. O'Connell
Massachusetts Institute of Technology

Wanwen Sun
Dartmouth College

Nancy A. Speck
Dartmouth College

Nancy Hopkins
Massachusetts Institute of Technology

JOURNAL OF VIROLOGY, Apr. 1993, p. 1967–1975
0022-538X/93/041967-09$02.00/0

Vol. 67, No. 4

Two Factors That Bind to Highly Conserved Sequences in Mammalian Type C Retroviral Enhancers

NANCY R. MANLEY,[1]† MARY O'CONNELL,[1]‡ WANWEN SUN,[2] NANCY A. SPECK,[2] AND NANCY HOPKINS[1]*

Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,[1] and Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755[2]

Received 3 September 1992/Accepted 29 December 1992

The transcriptional enhancers of the Moloney and Friend murine leukemia viruses (MLV) are important determinants of viral pathogenicity. We used electrophoretic mobility shift and methylation interference assays to study nuclear factors which bind to a region of these enhancers whose sequence is identical between Moloney and Friend viruses and particularly highly conserved among 35 mammalian type C retroviruses whose enhancer sequences have been aligned (E. Golemis, N. A. Speck, and N. Hopkins, J. Virol. 64:534–542, 1990). Previous studies identified sites for the leukemia virus factor b (LVb) and core proteins in this region (N. A. Speck and D. Baltimore, Mol. Cell. Biol. 7:1101–1110, 1987) as well as a site, overlapping those for LVb and core, for a third factor (N. R. Manley, M. A. O'Connell, P. A. Sharp, and N. Hopkins, J. Virol. 63:4210–4223, 1989). Surprisingly, the latter factor appeared to also bind two sites identified in the Friend MLV enhancer, Friend virus factor a and b1 (FVa and FVb1) sites, although the sequence basis for the ability of the protein to bind these diverse sites was not apparent. Here we describe the further characterization of this binding activity, termed MCREF-1 (for mammalian type C retrovirus enhancer factor 1), and the identification of a consensus sequence for its binding, GGN_8GG. We also identify a factor, abundant in mouse T-cell lines and designated LVt, which binds to two sites in the Moloney MLV enhancer, overlapping the previously identified LVb and LVc binding sites. These sites contain the consensus binding site for the Ets family of proteins. We speculate on how distinct arrays of these factors may influence the disease-inducing phenotype.

The transcriptional enhancers of nondefective mouse type C retroviruses are important determinants of viral pathogenesis whose precise nucleotide sequences influence leukemogenicity, disease specificity, and latent period of disease induction (7, 8, 11–14, 25, 32, 34). We and others are interested in identifying nuclear factors that interact with retroviral enhancers which may be candidates to mediate enhancer function in specific cell types and hence presumably to influence viral pathogenesis. Our efforts have focused on the nondefective Friend and Moloney murine leukemia virus (MLV) enhancers. Genetic and biochemical analyses have identified regions important for the regulation of disease specificity as well as binding sites for multiple cellular factors throughout the enhancers of these two viruses (19, 35, 43, 44).

Alignment of the nucleotide sequences of the enhancers of 35 mammalian type C retroviruses revealed that some regions contain highly conserved sequence motifs (20). Particularly striking is a central region almost invariably included in direct repeat sequences of mammalian type C retrovirus enhancers. In the Moloney and Friend MLVs, this region contains adjacent binding sites for the leukemia virus factor b (LVb) (43), for the core binding factor (CBF) (51) (otherwise known as the SL3-3 enhancer factor 1 [SEF1] [50], the SL3 and AKV CBF [S/A-CBF] [4], and polyomavirus enhancer core binding proteins 2 and 3 [PEBP2/3] [41]), and for a third factor whose binding site was previously ill defined but which appeared to include sequences in both the LVb and core sites (35). This third factor was also able to bind to sequences in the Friend virus enhancer that are less well conserved in mammalian type C virus enhancers.

Mutations introduced into the highly conserved LVb and core binding sites in the Moloney MLV enhancer have significant effects on enhancer function and on viral pathogenesis (44, 45). For this reason and because of its striking sequence conservation, we chose to further investigate factors that bind to this region. This study focuses on two factors. One is the previously ill-defined factor, one of whose binding sites overlaps those for LVb and core (35). Here we designate this factor mammalian type C retrovirus enhancer factor 1 (MCREF-1), define a consensus sequence for its binding, explain the basis for its ability to bind to the Friend virus factor a (FVa) and FVb1 sites, and define an additional binding site for the factor in the Moloney virus enhancer region. The other factor we describe, LVt, binds to sequences that overlap the highly conserved LVb site as well as to sequences overlapping the less well-conserved site for a previously defined LVc on the Moloney MLV enhancer (43). LVt appears to be enriched in T cells and may be a member of the Ets family of nuclear factors. We also summarize our current understanding of the array of factors that may bind the Friend and Moloney MLV enhancers and thereby contribute to the different pathogenicities encoded by these two elements.

MATERIALS AND METHODS

Extracts and cell lines. Nuclear extracts were made from the mouse EL4 T-cell line and the mouse erytholeukemia MEL cell line by the method of Dignam et al. (15). Rat brain nuclear extracts were prepared as described previously (35). Bovine MCREF-1 was partially purified from calf thymus as described in the accompanying report (46). Protein concen-

* Corresponding author.
† Present address: University of Utah, Salt Lake City, UT 84112.
‡ Present address: Biozentrum der Universitat Basel, CH-4056 Basel, Switzerland.

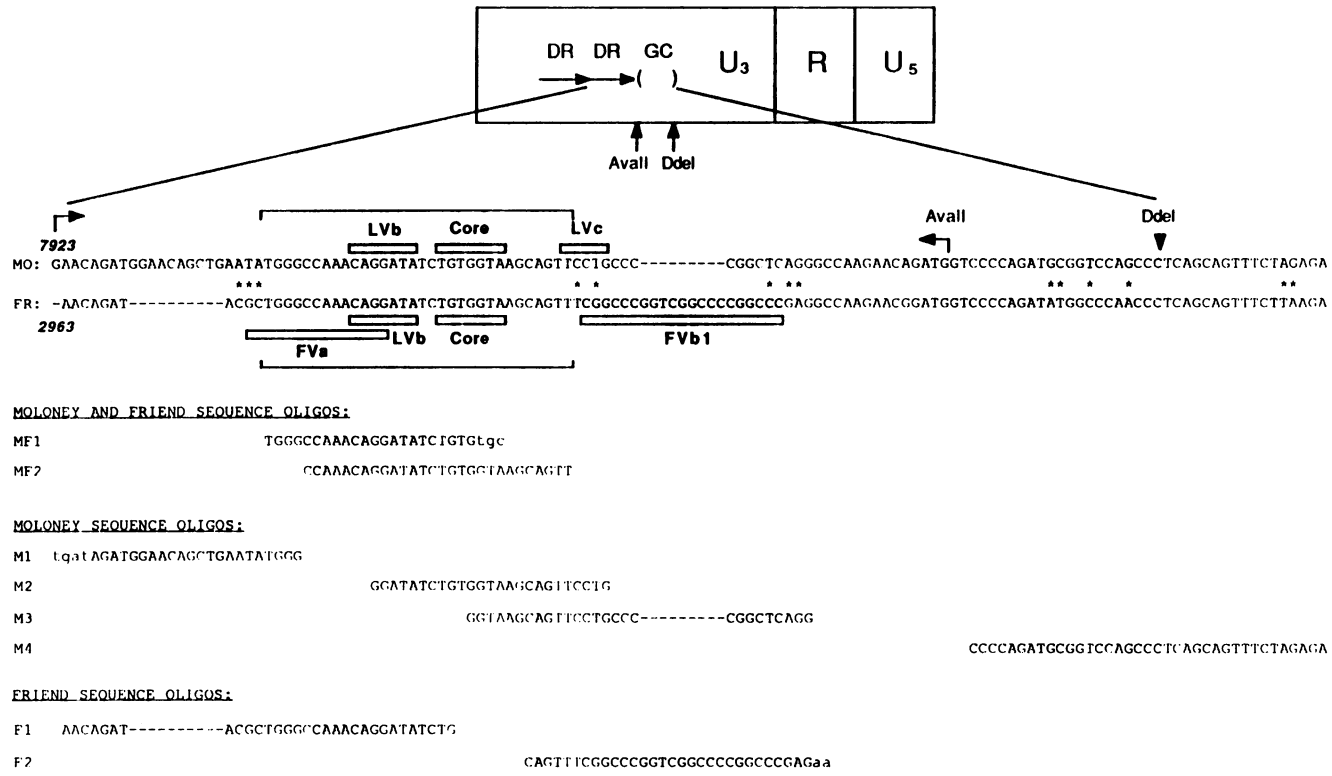

FIG. 1. Origins of wild-type and mutant oligonucleotides from the Moloney and Friend MLV enhancers. A schematic representation of the Moloney and Friend MLV long terminal repeat is shown on the top. The sequences underneath are from the second copy of the direct repeat of the Moloney (MO) and Friend (FR) enhancers and the adjacent GC-rich region. Numbering of the enhancer sequences is from that of Weiss et al. (54). Asterisks between the Moloney and Friend MLV enhancer sequences indicate positions of sequence divergence. Dashes within the sequence indicate gaps in the alignment between the two enhancers. Previously defined binding sites in the Moloney and Friend enhancers relevant to the discussion are indicated by rectangles above and beneath the sequences (35, 43). The conserved sequence which contains the LVb and core sites is bracketed. The sequences of oligonucleotide probes and competitors are listed below the enhancer sequences. Oligonucleotides from sequences conserved between Moloney and Friend enhancers are called MF1 and MF2. Oligonucleotides from areas of sequence divergence between the Moloney and Friend enhancers are called M1 to -4 and F1 to -2, respectively. Mutations in sequences are indicated by lowercase letters.

trations were determined by the method of Bradford (5), with reagents purchased from Bio-Rad.

Substrates for protein binding in electrophoretic mobility shift assays. The origins of the oligonucleotides from the Moloney and Friend MLV enhancers that were used in biochemical assays are shown in Fig. 1. Complementary oligonucleotides were synthesized with an Autogen 6500 DNA synthesizer at the Center for Cancer Research, Massachusetts Institute of Technology. All oligonucleotides were purified by electrophoresis through 20% polyacrylamide–7 M urea gels. Radioactive probes were made by labeling one oligonucleotide, either the top or bottom strand of the binding site, with [γ-^{32}P]ATP (7,000 Ci/mmol; ICN) and T4 polynucleotide kinase (New England Biolabs) and then annealing the labeled oligonucleotide with its complementary strand (26). The double-stranded probes were purified by electrophoresis through 20% native polyacrylamide gels.

Competitor oligonucleotides were prepared by annealing equal molar amounts of unlabeled complementary oligonucleotides in a buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1.0 mM EDTA, 1 mM 2-mercaptoethanol, and 4% glycerol at a final double-stranded DNA concentration of 100 ng/μl. Annealing and concentration were verified by agarose gel electrophoresis against a DNA standard (ϕX174 *Hae*III digest; New England Biolabs). The annealed oligonucleotides were used directly as competitors in binding reactions.

Protein-DNA binding analysis. Binding activity was detected by the electrophoretic mobility shift assay (16, 17, 42). Electrophoretic mobility shift assays were performed according to three protocols, which will be referred to throughout the text by the different gel buffer systems used: Tris-borate, Tris-acetate, and Tris-glycine. Binding reaction mixtures for EL4, MEL cell, and rat brain nuclear extracts that were assayed by Tris-borate or Tris-acetate protocols were identical and contained 50,000 cpm of ^{32}P-labeled oligonucleotide probe, 10 to 15 μg of nuclear extract, and 0.5 to 1.0 μg of poly(dI-dC)-poly(dI-dC) (Pharmacia) in binding buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, and 4% glycerol (total volume, 20 μl). Partially purified bovine MCREF-1 was analyzed by the Tris-borate protocol. Binding reaction mixtures for partially purified bovine MCREF-1 contained 10,000 cpm of ^{32}P-end-labeled probe, 0.2 μg of poly(dI-dC)-poly(dI-dC), and 3 μl of partially purified bovine MCREF-1 (heparin-Sepharose fraction [46]) in a total volume of 15 μl.

Binding reaction mixtures for the Tris-glycine protocol contained 10,000 cpm of ^{32}P-labeled probe, 3 to 5 μg of nuclear extract, 50 to 100 ng of poly(dI-dC)-poly(dI-dC), and

10 to 20 ng of a 25-base single-stranded oligonucleotide of heterologous sequence (5′-GATCAAGGTCAGGTACACG AAAA-3′) in buffer containing 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) (pH 7.5), 1.0 mM EDTA (pH 8.0), 5 mM dithiothreitol, and 10% glycerol (total volume, 10 μl).

All binding reaction mixtures were incubated at room temperature for 15 minutes and then loaded directly onto 4 or 6% native polyacrylamide gels. Tris-borate gels contained 22 mM Tris, 22 mM boric acid, 0.5 mM EDTA, and 3% glycerol and were electrophoresed in the same buffer minus glycerol. Both the Tris-acetate gel and the electrophoresis buffer contained 6.7 mM Tris (pH 7.5), 3.3 mM sodium acetate, and 0.1 mM EDTA. Tris-glycine gels contained 8.5 mM Tris (pH 8.5), 65 mM glycine, and 0.35 mM EDTA and were electrophoresed in 0.6× buffer (5.1 mM Tris [pH 8.5], 39 mM glycine, 0.21 mM EDTA). Tris-borate and Tris-acetate gels were electrophoresed at 10 V/cm, and Tris-glycine gels were electrophoresed at 12 V/cm. Radioactivity was detected by autoradiography of dried gels.

In summary, the principle differences among the three buffer systems used are the buffer and ion concentrations, pH, and the presence or absence of glycerol in the gel. Any or all of these components may contribute to the differences in protein binding seen in this study.

Methylation interference analyses. Binding reaction mixtures for methylation interference analysis of LVt contained 250,000 cpm of end-labeled, methylated (36) probe, 24 μg of EL4 nuclear extract, 100 ng of single-stranded competitor, and 350 ng of poly(dI-dC)-poly(dI-dC), with the Tris-glycine protocol. Binding reaction mixtures for murine MCREF-1 contained 500,000 cpm of end-labeled methylated probe, 40 μg of MEL cell nuclear extract, and 3 μg of poly(dI-dC)-poly(dI-dC) and were analyzed with the Tris-borate protocol. Following electrophoresis and overnight exposure of the gels, bands corresponding to the protein-DNA complex and free DNA were excised from the gel, and the DNA was purified as described previously (35), subjected to 1 M piperidine cleavage, and analyzed by electrophoresis through 15% polyacrylamide–7 M urea sequencing gels.

RESULTS

The transcriptional enhancer regions of the Moloney and Friend MLVs consist of an approximately 200-bp sequence containing a direct repeat and a 3′ GC-rich region and are located in the U3 segment of the long terminal repeat (29). Figure 1 shows a sequence comparison of the two viral enhancers in the second copy of the direct repeat and the GC-rich region. There is a 32-bp sequence in the center of the repeat which is identical between the two viruses (shown in brackets). This sequence is also highly conserved among many other mammalian type C retroviral enhancers (20). We synthesized double-stranded oligonucleotides containing sequences from both conserved and unique regions of the Moloney and Friend MLV enhancers to analyze binding of nuclear factors. The sequences of these oligonucleotides and their positions within the enhancers are shown in Fig. 1.

Identification of LVt. In order to look for proteins that bind specifically to the conserved LVb-core region and which were not identified in previous studies of the Moloney, Friend, and SL3-3 MLV enhancers (4, 43, 49), we compared binding of unfractionated EL4 T-cell nuclear extracts to an oligonucleotide probe containing the LVb and core sequences (oligonucleotide MF2 in Fig. 1) with different, but commonly employed, gel buffer systems in the electrophoretic mobility shift assay. In Tris-borate or Tris-acetate buffer systems the major protein-DNA complex detected is primarily due to binding over the core site (Fig. 2A and B). Formation of this complex is significantly inhibited by the addition of unlabeled competitor oligonucleotides which contain the core binding site sequence TGTGGTA (lanes 2 and 3) but is not inhibited by competitor oligonucleotides which contain only an LVb site (lane 7) or tandem LVb sites (lane 4; see Figure 5A for sequence of this LVb-LVb oligonucleotide). This binding activity is also not inhibited by sequences from either the Friend or Moloney MLV direct repeats which flank the conserved LVb-core region (lanes 5, 6, 8, and 9). Methylation interference analysis of this complex is consistent with the pattern described for previously characterized core binding proteins, the bovine CBF (51), SEF1 (49), and PEBP2/3 (41), which may be identical (data not shown).

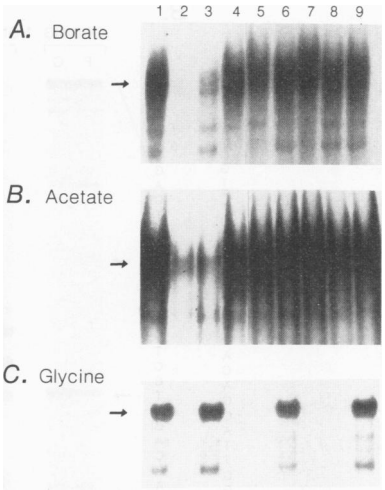

FIG. 2. Comparison of the protein-DNA complex obtained with the LVb-core probe in Tris-borate, Tris-acetate, and Tris-glycine gel systems. (A) Binding of EL4 nuclear extract to the MF2 probe, electrophoresed through a Tris-borate gel. Binding reaction mixtures contained EL4 nuclear extract and MF2 probe in the presence or absence of 100 ng of unlabeled double-stranded competitor oligonucleotides. The regions of the gels containing the sequence-specific protein-DNA complexes are shown. Lanes: 1, binding in the absence of competitor oligonucleotides; 2, binding in the presence of unlabeled MF2 competitor DNA; 3, 100 ng of M2 competitor; 4, LVb-LVb competitor (see Fig. 5A for sequence of this oligonucleotide); 5, F1; 6, F2; 7, MF1; 8, M3; 9, M1. (B) Electrophoretic mobility shift assay with a Tris-acetate gel system. Lanes correspond to those in panel A. (C) Electrophoretic mobility shift assay with a Tris-glycine gel system.

When the binding reactions just described are performed with a Tris-glycine gel buffer system, a dramatically different pattern is seen (Fig. 2C). Comparison of the sequences of oligonucleotides which are capable of interfering with the major specific protein-DNA complex detected in this gel system shows that all contain the sequence 5′-CAGGA-3′ (Fig. 2C, lanes 2, 4, 5, 7, and 8 [oligonucleotide competitors MF2, LVb-LVb, F1, MF1, and M3]). Oligonucleotides whose sequences terminate within the CAGGA sequence (Fig. 2C, lane 3) or do not contain it (Fig. 2C, lanes 6 and 9) do not inhibit formation of the protein-DNA complex seen with Tris-glycine gels. The M3 oligonucleotide, which contains sequences from the Moloney MLV direct repeat 3′ of the conserved region, contains the sequence CAGGA on the

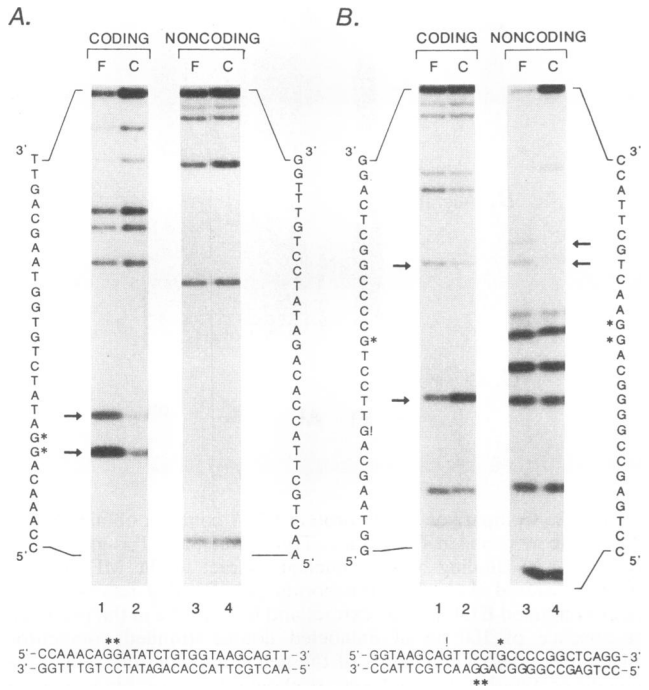

FIG. 3. Methylation interference analysis of LVt. (A) Methylation interference pattern generated by LVt from EL4 T-cell nuclear extracts on the MF2 probe. Lanes 1 and 2, top (coding) strand of the MF2 probe; lanes 3 and 4, bottom (noncoding) strand. Lanes F and C correspond to the free DNA bands and the protein-DNA complex bands, respectively. Arrows indicate the locations of guanines whose methylation specifically inhibits the binding of LVt to the MF2 probe. The sequence of the MF2 probe is shown on each side of the panel and on the bottom. Asterisks next to the sequence indicate the guanine contacts. (B) Methylation interference pattern generated by LVt on the M3 probe. Exclamation mark indicates a position of enhanced cleavage.

bottom or noncoding strand. The CAGGA sequence on the MF2 and M3 probes overlaps the previously defined LVb and LVc sites on the Moloney MLV enhancer (43). When the M3 oligonucleotide is used as a probe in mobility shift assays in the Tris-glycine gel system, the resultant protein-DNA complex has a similar mobility and shows the same competition profile as that seen with the MF2 oligonucleotide probe (data not shown). Thus, we have identified a factor that binds to both the MF2 and the M3 probes. We call this factor LVt. This name was selected because a comparison of nuclear extracts prepared from a variety of cell lines showed that the protein-DNA complex seen with the Tris-glycine gel system is present at relatively high levels in T-cell lines and in primary spleen cells enriched for T cells by culturing in the presence of interleukin-2 (data not shown), while lower levels of binding, or no binding, were observed in non-T-cell extracts.

Methylation interference analysis of the LVt complexes formed over its sites on the MF2 and M3 probes confirms that the protein-DNA complex detected in Tris-glycine gels is specific for the CAGGA sequence in both probes (Fig. 3). In both cases, methylation of the two adjacent G residues in the middle of this sequence interferes with formation of the protein-DNA complex. No interference was seen on the noncoding strand on the MF2 probe (Fig. 3A). However, a guanine on the coding strand of the M3 probe does appear to be involved in binding (Fig. 3B). The more extensive interference pattern seen at the LVt site in the M3 probe may be due to the slight difference in sequence between the two sites in the MF2 and M3 probes.

Identification of a consensus binding site for MCREF-1. Previously we reported the identification of a protein, which we now call MCREF-1, that appeared to bind specifically to at least three sites within the Friend MLV direct repeat sequence (35). Two of these sites, FVa and FVb1, flank the conserved core site, while the third is located within the LVb-core region and is shared with Moloney MLV. It was also observed that an additional binding site for MCREF-1 might be present in the Moloney virus enhancer in sequences corresponding to the Friend virus FVa site, although the existence of this site in the Moloney MLV enhancer was difficult to confirm with crude nuclear extracts because of the presence of an overlapping binding site for nuclear factor 1 (35) (see reference 46 for further characterization of this site). Because the sequences of the MCREF-1 binding sites were very different and the methylation interference patterns of MCREF-1 in the FVa, FVb1, and LVb-core region varied considerably, we were unable at that time to deduce a consensus binding site for the protein (35). Binding studies with probes and competitor oligonucleotides that have variations of the LVb-core region sequence have now allowed us to deduce a consensus sequence for MCREF-1 among the three previously identified binding sites. In addition, we have identified an MCREF-1 site in the GC-rich region 3′ to the direct repeat in the Moloney virus.

To continue our analysis of the factor, now called MCREF-1, which binds over the LVb-core region, we first used crude nuclear extracts prepared from rat brain. These extracts are deficient in core binding factor, which obscures MCREF-1 binding in T-cell nuclear extracts and thus facilitated studies of MCREF-1. Formation of the MCREF-1 protein-DNA complex on the MF2 probe is specifically inhibited by sequences containing the FVa and FVb1 sites, as described previously (Fig. 4A, lanes 5 and 6 [35]), and also by an oligonucleotide containing two tandem LVb sites (lane 4) but not by an oligonucleotide truncated in the core site (lane 7) or truncated in the LVb site (lane 3). When the sequences of those oligonucleotides that can inhibit binding of MCREF-1 to the MF2 oligonucleotide were aligned along with the cross-inhibiting FVa and FVb1 sites identified previously and the guanine contacts for MCREF-1 in these sites were noted, a consensus binding site sequence emerged (Fig. 5A). This sequence is GGN_8GG, or, as described below, $CNGGN_6CNGG$.

At this point in our studies, in order to obtain larger amounts of material, attempts were made to partially purify MCREF-1 from calf thymus (46). With a partially purified bovine preparation (heparin-Sepharose fraction [46]), competition assays were repeated to confirm that bovine MCREF-1 was equivalent to MCREF-1 in rat brain nuclear extracts (Fig. 4B). Formation of the bovine MCREF-1 protein-DNA complex is inhibited by the LVb-core A oligonucleotide (Fig. 4B, lane 2 [see Figure 5B for the sequence of this oligonucleotide]), the LVb-LVb oligonucleotide (lane 4), and the FVa and FVb1 binding sites (F1 and F2; lanes 5 and 6), but not by oligonucleotide MF1 (lane 7) or M2 (lane 3) or the M3 or M1 oligonucleotide (lanes 8 and 9), results consistent with the specificity of MCREF-1 from rat brain nuclear extracts.

In order to further test the hypothesis that the sequence GGN_8GG is responsible for the cross-competition seen between FVa, FVb1, and the LVb-core sequences, we

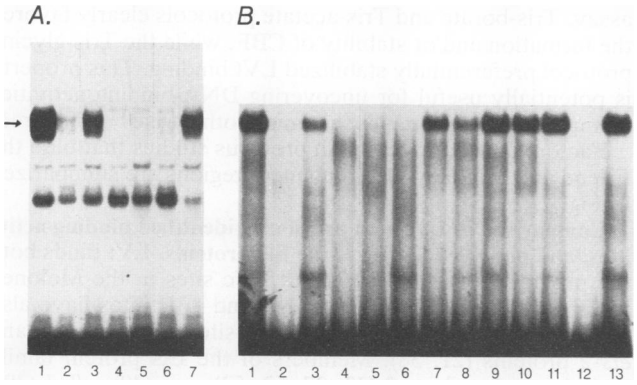

FIG. 4. Binding specificity of MCREF-1 for sequences in the conserved LVb-core region. (A) Binding of MCREF-1 from rat brain nuclear extracts to the MF2 probe, in the presence or absence of 100 ng of unlabeled double-stranded competitor oligonucleotides. Lanes: 1, binding in the absence of competitor oligonucleotides; 2, binding in the presence of unlabeled MF2 competitor DNA; 3, M2 competitor; 4, LVb-LVb; 5, F1; 6, F2; 7, MF1. (B) Binding of partially purified bovine MCREF-1 to the LVb-core A probe. Lanes: 1, binding to the LVb-core A probe in the absence of competitor oligonucleotides; 2, binding in the presence of 50 ng of LVb-core A competitor; 3, 50 ng of M2; 4, LVb-LVb; 5, F1; 6, F2; 7, MF1; 8, M3; 9, M1; 10, LVb-core B; 11, LVb-core C; 12; LVb-core E; 13, LVb-core D. Sequences of the LVb-core A to E oligonucleotides are listed in Fig. 5B.

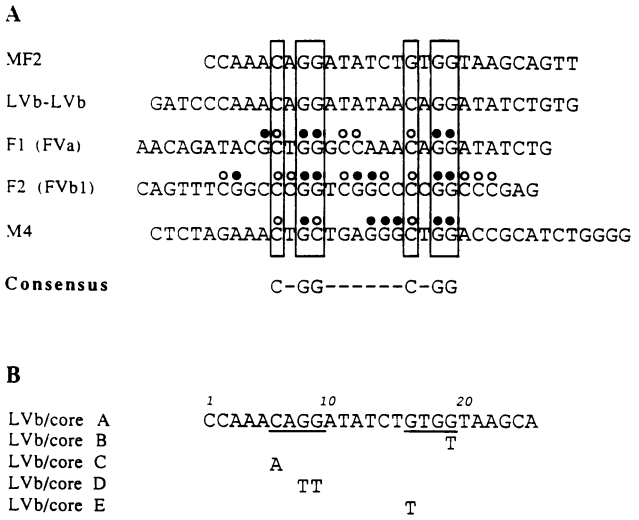

FIG. 5. Determination of a consensus binding site for MCREF-1. (A) Listed are the identified MCREF-1 binding sites from the Moloney and Friend MLV enhancers, plus the tandem LVb site oligonucleotide (LVb-LVb). Closed circles indicate guanine contacts made by MCREF-1 on the F1, F2, and M4 oligonucleotide probes (35; this paper), and open circles indicate the guanine contacts made on the bottom (noncoding strand) of the probe. The consensus CNGG sequences at each end of the binding site are boxed. (B) Sequences of oligonucleotides with mutations in the derived consensus binding site for MCREF-1. Shown is the conserved LVb-core region from mammalian type C retroviral enhancers. Each of the oligonucleotides LVb/core B to E contains mutations at one or two positions, indicated by the substituted nucleotide. Positions in the oligonucleotide are numbered above the sequence.

prepared a series of oligonucleotides containing specific mutations in the putative MCREF-1 binding site in the conserved LVb-core region sequence (oligonucleotides LVb core B to E in Fig. 5B) and used these as competitors of MCREF-1 binding to the LVb-core A probe. For this analysis, we used partially purified bovine MCREF-1. For the purpose of discussion, we have numbered the positions in the LVb-core site in Figure 5B. Mutation of G residues at either end of the putative GGN_8GG consensus sequence (positions 8 and 9, or position 19) results in inability to inhibit MCREF-1 protein-DNA complex formation (Fig. 4B, LVb-core D and B, lanes 13 and 10). Mutation of the C residue at position 6 also eliminates the ability to inhibit MCREF-1 binding (LVb-core C, lane 11). Mutation of a G residue at position 16 did not affect MCREF-1 binding (LVb-core E, lane 12).

The Moloney and Friend virus enhancer GC-rich segments have been defined in functional studies of the enhancer (29) and in genetic studies of disease specificity (19) as a 22-bp sequence extending from the end of the direct repeat to the downstream *Dde*I site (Fig. 1). Within this segment, there are only four base pair differences between the two viral sequences. To investigate the possible binding of cellular factors to the Moloney virus GC-rich segment, we performed mobility shift and methylation interference assays with the M4 oligonucleotide probe, which contains the GC-rich region and an additional 16 bp 3′ to the *Dde*I site.

Incubation of the M4 oligonucleotide probe in a MEL cell nuclear extract generated a single protein-DNA complex that appeared to have the same sequence specificity as MCREF-1 from rat brain and calf thymus in competition assays (data not shown). Formation of this protein-DNA complex was also inhibited by the M4 oligonucleotide but not by an oligonucleotide containing the corresponding sequences from the Friend MLV enhancer (data not shown). These results indicate that the Moloney GC-rich region contains yet another binding site for MCREF-1. Methylation interference analysis of MCREF-1 on the M4 probe identifies a 14-bp sequence, 5′-CCAGCCCTCAGCAG-3′, which is involved in specific binding (Fig. 6). This sequence includes only a single base difference in the corresponding Friend C region sequence (5′-CCAACCCTCAGCAG-3′), indicating that a single base difference at this position is sufficient to eliminate MCREF-1 binding at this site. Note that as discussed above, this same base change did not eliminate binding to the MCREF-1 site present in the MF2 oligonucleotide (i.e., MCREF-1 binds to LVb/core E in Fig. 5B). The two MCREF-1 sites, that overlapping the LVb-core sequence and the one in the Moloney virus GC-rich region, have a number of sequence differences which may explain this discrepancy. In particular, the MCREF-1 site in the Moloney GC-rich region has the sequence 5′-GCN_8GG rather than the more common GGN_8GG, and this might explain why this site cannot tolerate a change at position 16 (Fig. 5B) while an MCREF-1 site with the more common sequence can (see reference 46 for further discussion of this point).

Together the results support the GGN_8GG consensus sequence for the MCREF-1 binding site. The frequency at which a C appears at positions 6 and 16 in Fig. 5B suggests that $CNGGN_6CNGG$ is the most complete representation of MCREF-1 sequence specificity. As noted, there appears to be variation in the importance of any one of these residues at a particular MCREF-1 binding site. Overall sequence context at each of these divergent sites may influence what variation from the consensus can be tolerated and still allow specific binding.

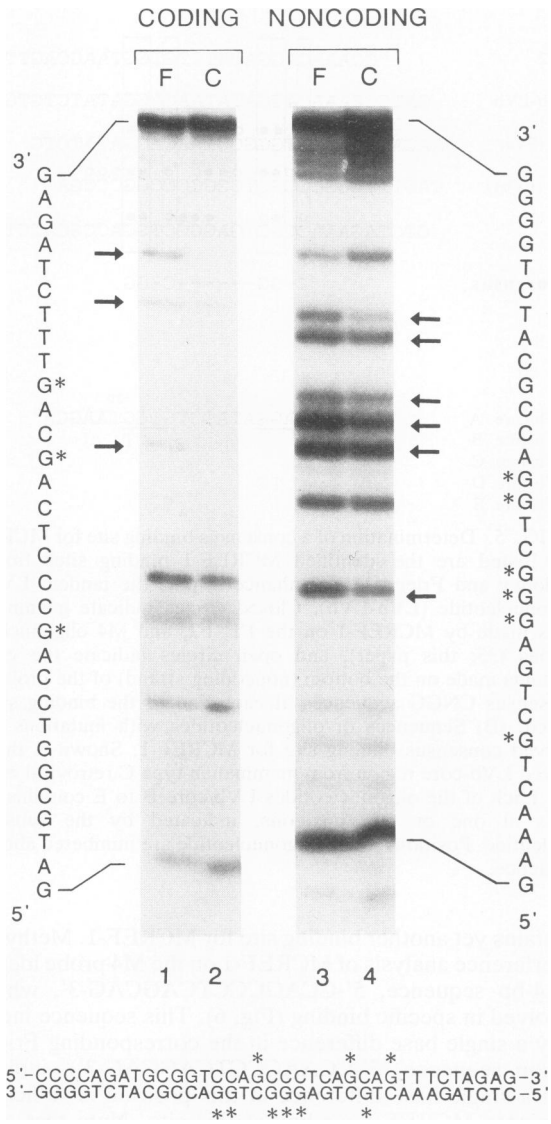

FIG. 6. Methylation interference analysis of MCREF-1 on the M4 probe (the Moloney enhancer GC-rich region). Shown is the methylation interference pattern generated by MCREF-1 from MEL cell nuclear extracts on the M4 probe. Lanes 1 and 2, top (coding) strand of the M4 probe; lanes 3 and 4, bottom (noncoding) strand. Lanes F and C correspond to the free DNA bands and the protein-DNA complex bands, respectively. Arrows indicate the locations of guanines whose methylation specifically inhibits the binding of MCREF-1 to the M4 probe. The sequence of the M4 probe is shown on each side of the panel and on the bottom. Asterisks next to the sequence indicate the guanine contacts.

DISCUSSION

We have identified two factors, LVt and MCREF-1, that bind to the conserved LVb-core region of mammalian type C retroviral enhancers. LVt requires sequences in the LVb site for binding, but not in the core site. MCREF-1 binding requires sequences in both the LVb and core sites. Both of these proteins also bind sites in less conserved regions of the Moloney and Friend MLV enhancers. The distinct binding activities for the LVb-core region were revealed by using different gel systems in the electrophoretic mobility shift assay. Tris-borate and Tris-acetate protocols clearly favored the formation and/or stability of CBF, while the Tris-glycine protocol preferentially stabilized LVt binding. This property is potentially useful for uncovering DNA-binding activities in impure protein samples that might otherwise be obscured.

Factors defined here and in previous studies that bind the Friend and Moloney virus enhancer regions are summarized in Fig. 7.

Comparison of LVt with previously identified binding activities, and possible identity to the Ets proteins. LVt binds both the previously defined LVb and LVc sites in the Moloney MLV enhancer (Fig. 7). The LVb and LVc sites have also been recently identified as binding sites for the *ets-1* and *ets-2* proteins (21, 38). Members of the Ets protein family include *ets-1* and *ets-2* (33, 37, 52, 53), *erg* (40), *elk-1* (39), E74 (6), PU.1 (28), GABPα (30), Fli-1 (1), Elf-1 (47), PEA3 (55), and SAP-1 (10). The DNA-binding domain of the Ets proteins is conserved (27), and all members of the family that have been studied bind the sequence GGA (10, 24, 27, 38, 47, 48, 55).

Some aspects of the binding properties of LVt suggest that it is a member of the Ets family. The preferential expression of LVt activity in T-cell lines is consistent with the tissue distribution of *ets-1* (2, 3, 9, 18). In addition, two guanines that LVt contacts in the LVb site (positions 8 and 9) are the two guanines in the conserved GGA sequence recognized by the members of the Ets protein family. Formation of the LVt protein-DNA complex can be specifically competed with an *ets-1* binding site from the Moloney murine sarcoma virus promoter (AGCGGAAGCG [9a]) (22). The guanine contacts made by LVt on the LVb site are only a subset of those made by *ets-1* (38). Nonetheless, both LVt and *ets-1* contact guanines at positions 8 and 9 on the top strand of the LVb site. Methylation of purine bases flanking guanines 8 and 9, which do not appear to affect LVt binding, show less severe effects on *ets-1* binding than methylation of guanines 8 and 9 (38). If LVt is *ets-1*, one would have to argue that the differences in the contacts made outside of the conserved GGA sequence in the LVb site might reflect differences in the purity of the two proteins and/or the conditions that were used in the methylation interference analysis. LVt could also be one of the other members of the Ets family of proteins, which may make different contacts in the sequences flanking the GGA consensus site. Either purification of LVt followed by microsequence analysis or antibodies that distinguish among proteins in the Ets family would be necessary to establish the identity of LVt as one of the Ets proteins.

Both the 5′ and 3′ LVt binding sites identified in this study overlap with previously identified binding sites in the Moloney virus enhancer. Differences in competition and methylation interference results clearly differentiate the 5′ LVt site from the previously identified LVb binding site which it overlaps extensively (43). The 3′ LVt binding site overlaps that of the LVc binding activity. Although there are slight differences in the reported methylation interference patterns of the two binding activities, differences in the methodology used to characterize the LVt and LVc binding activities make it impossible to exclude the possibility that they are in fact the same.

MCREF-1 binds multiple sites with a GGN_8GG consensus in mammalian type C retrovirus enhancers. MCREF-1 binds to several sites in the Friend virus enhancer: the previously defined FVa and FVb1 sites and a site in the conserved LVb-core region (Fig. 7). Three binding sites for MCREF-1 are present on the Moloney virus enhancer and adjacent GC-rich region. Originally, the degeneracy of the sequence

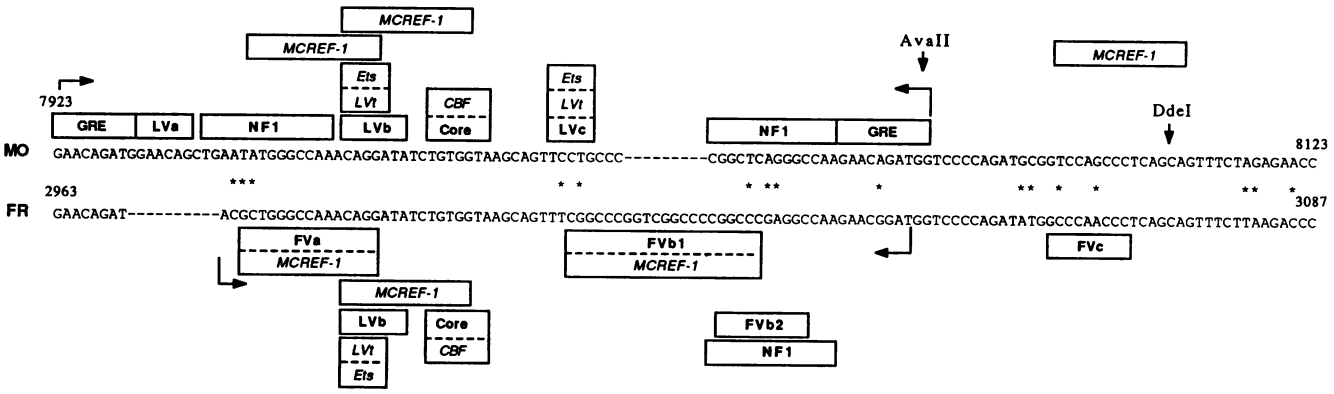

FIG. 7. Summary of proteins that bind the Friend and Moloney enhancers. All binding sites are indicated by boxes. The binding sites defined by Speck and Baltimore (43) and Manley et al. (35) are indicated by boxes containing letters in plain type, while the names of proteins characterized since these original descriptions are indicated in italics. The sequences and numbering of the sequences are the same as in Fig. 1.

of the first two MCREF-1 binding sites identified, the FVa and FVb1 sites, made it difficult to define a consensus binding site (35). This difficulty was compounded by what appeared to be significant differences in the purine contacts made by MCREF-1 on the FVa and FVb1 sites as defined by methylation interference analysis. Identification of additional MCREF-1 binding sites (this paper) and comparison of the purine and phosphate contacts made by MCREF-1 on these diverse sites (35, 46) have now allowed us to derive a consensus sequence of GGN_8GG (or $CNGGN_6CNGG$). The extensive methylation interference pattern generated by MCREF-1 in some of its sites, notably the FVb1 site, is still an interesting and inexplicable feature of the protein.

The 5′ MCREF-1 site (the FVa site) overlaps the LVt/Ets binding site in both the Moloney and Friend virus enhancers, and the MCREF-1 binding site in the LVb-core region overlaps the binding sites for both LVt or Ets and CBF (SEF1, PEBP2/3) (Fig. 7). Thus, there exists the potential for complex regulation of the occupancy of these sites by the various proteins that bind them.

Transcriptional regulation of viral pathogenesis. Our goal has been to identify nuclear factors whose binding to the Friend and Moloney virus enhancers might explain how these elements encode viral disease specificity. Genetic studies of viruses with recombinant enhancers revealed that erythroleukemia induction or T-cell leukemogenesis are encoded in at least three determinants that flank the central region of sequence identity within the Friend and Moloney virus enhancer and that act additively or cooperatively to produce strong determinants of disease specificity (19). These studies imply that to identify nuclear factors that shape disease specificity, one should look to the regions flanking the LVb-core sequence. In contrast, genetic studies in which mutations were introduced into the LVb and core sites of Moloney virus suggested that it is this region of sequence identity between Friend and Moloney viruses that is important in disease specificity, since mutations in the LVb and core sites can convert the Moloney virus to one that induces a significant percentage of erythroleukemias (44). Studies of the SL3 virus also pointed to the core site as a particularly important determinant of T-cell leukemogenesis (23, 31). How can these apparently contradictory observations be reconciled, and could the proteins now known to bind the Friend and Moloney viral enhancers explain the various genetic observations?

The finding of T-cell-enriched nuclear factors, LVt, ets-1, and CBF, that bind the central highly conserved region of mammalian type C virus enhancers may well explain why T-cell leukemia is the most commonly observed disease among leukemogenic mouse type C retroviruses. Since mutations introduced into the 5′ LVt/Ets site (the LVb site) and the CBF binding site convert Moloney virus from an almost exclusively T-cell leukemia-inducing virus to one that induces both T-cell and erythroleukemias, one explanation of why the Friend virus enhancer encodes erythroleukemogenicity might be that the assembly of factors that bind this enhancer prevents binding of either LVt/Ets or CBF to the LVb-core region, in effect producing an equivalent to mutations in these sites. One difference in binding sites between the Moloney and Friend enhancers in a region shown by genetics to be an important determinant of disease specificity is the presence of the 5′ NF-1 site in the Moloney virus enhancer overlapping the 5′ MCREF-1/FVa site (Fig. 7) (19, 35). Perhaps in Friend virus MCREF-1 binds the MCREF-1/FVa site and thereby blocks the binding of LVt/Ets over the LVb site, resulting in partial erythroleukemogenicity. In the Moloney virus enhancer, occupancy of the 5′ NF-1 site might preclude binding of MCREF-1, thus allowing assembly of LVt/Ets and CBF on the LVb-core site. Alternatively, of course, it remains possible that the Friend virus enhancer may contain binding sites, not yet identified, for proteins specifically expressed in erythroid cells.

On the 3′ side of the core site, also shown by genetics to be a region that encodes a determinant of disease specificity, is the 3′ LVt/Ets/LVc site which is present in Moloney but not Friend virus. Although a 2-base mutation introduced into this LVc/LVt/Ets site did not convert Moloney to an erythroleukemia-inducing virus (44), a naturally occurring 2 bp difference in this region between Friend and Moloney viruses does apparently contribute to the ability of Friend virus to induce a low percentage of erythroleukemias (19). This observation, that a different 2-bp mutation in the same protein binding site results in different pathogenic phenotypes, underscores the complexity of the retroviral enhancer. It also illustrates the importance of testing biochem-

ical observations by constructing the appropriate viral mutants and testing them in mice.

ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We thank Lucinia Pilapil for outstanding technical assistance. We thank Patty Culp for helpful discussions and for communicating unpublished data.

This work was supported by Public Health Service grant R01-CA19308 from the National Institutes of Health to N. Hopkins, by Public Health Service grant CA51065-01A1 from the National Cancer Institute to N. A. Speck, in part by grants DMB-9009098 from the National Science Foundation to N. A. Speck and P01-CA42063 (CORE) to P. A. Sharp, and by a Cancer Center CORE grant (CA23018) from the National Cancer Institute to the Norris Cotton Cancer Center. N.A.S. is a recipient of a Junior Faculty Research Award from the American Cancer Society.

REFERENCES

1. **Ben-David, U., E. B. Giddens, K. Letwin, and A. Bernstein.** 1991. Erythroleukemia induction by Friend murine leukemia virus: insertional activation of a new member of the ets gene family, Fli-1, closely linked to c-ets-1. Genes Dev. **5**:908–918.
2. **Bhat, N. K., F. J. Fisher, S. Fujiwara, R. Ascione, and T. S. Papas.** 1987. Temporal and tissue-specific expression of mouse ets genes. Proc. Natl. Acad. Sci. USA **84**:3161.
3. **Bhat, N. K., K. L. Komschlies, S. Fujiwara, R. J. Fisher, B. J. Mathieson, T. A. Gregorio, H. A. Young, J. W. Kasik, K. Ozato, and T. S. Papas.** 1989. Expression of ets genes in mouse thymocyte subsets and T cells. J. Immunol. **142**:672–678.
4. **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.
5. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
6. **Burtis, K. C., C. S. Thummel, C. W. Jones, F. D. Karim, and D. S. Hogness.** 1990. The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell **61**:85–99.
7. **Chatis, P. A., C. A. Holland, J. W. Hartley, W. P. Rowe, and N. Hopkins.** 1983. Role for 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.
8. **Chatis, P. A., C. A. Holland, J. E. Silver, T. N. Frederickson, N. Hopkins, and J. W. Hartley.** 1984. A 3′ end fragment encompassing the transcriptional enhancer of nondefective Friend virus confers erythroleukemogenicity on Moloney murine leukemia virus. J. Virol. **52**:248–259.
9. **Chen, J. H.** 1985. The proto-oncogene c-*ets* is preferentially expressed in lymphoid cells. Mol. Cell. Biol. **5**:2993–3000.

9a. **Culp, P., and N. Hopkins.** Unpublished data.

10. **Dalton, S., and R. Treisman.** 1992. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. Cell **68**:597–612.
11. **DesGroseillers, L., and P. Jolicoeur.** 1984. Mapping the viral sequences conferring leukemogenicity and disease specificity in Moloney and amphotropic murine leukemia viruses. J. Virol. **52**:448–456.
12. **DesGroseillers, L., and P. Jolicoeur.** 1984. The tandem direct repeats within the long terminal repeat of murine leukemia viruses are the primary determinant of their leukemogenic potential. J. Virol. **52**:945–952.
13. **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.
14. **DesGroseillers, L., R. Villemur, and P. Jolicoeur.** 1984. The high leukemogenic potential of Gross passage A murine leukemia virus maps in the region of the genome corresponding to the long terminal repeat and to the 3′ end of *env*. J. Virol. **47**:24–32.
15. **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.
16. **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.
17. **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.
18. **Ghysdael, J., A. Gegonne, P. Pognonec, D. Dernis, D. Leprince, and D. Stehelin.** 1986. Identification and preferential expression in thymic and bursal lymphocytes of a c-ets oncogene-encoded Mr 54,000 cytoplasmic protein. Proc. Natl. Acad. Sci. USA **83**:1714–1718.
19. **Golemis, E., Y. Li, T. N. Frederickson, J. W. Hartley, and N. Hopkins.** 1989. Distinct segments within the enhancer region collaborate to specify the type of leukemia induced by nondefective Friend and Moloney viruses. J. Virol. **63**:328–337.
20. **Golemis, E., N. A. Speck, and N. Hopkins.** 1990. Alignment of U3 region sequences of mammalian type C viruses: identification of highly conserved motifs and implications for enhancer design. J. Virol. **64**:534–542.
21. **Gunther, C. V., and B. J. Graves.** Unpublished data.
22. **Gunther, C. V., J. A. Nye, R. S. Bryner, and B. J. Graves.** 1990. Sequence-specific DNA binding of the proto-oncoprotein ets-1 defines a transcriptional activator seqauence within the long terminal repeat of the Moloney murine sarcoma virus. Genes Dev. **4**:667–679.
23. **Hallberg, B., J. Schmidt, A. Luz, F. S. Pedersen, and T. Grundstrom.** 1991. SL3-3 enhancer factor 1 transcriptional activators are required for tumor formation by SL3-3 murine leukemia virus. J. Virol. **65**:4177–4181.
24. **Hipskind, R. A., V. N. Rao, C. G. F. Mueller, E. S. P. Reddy, and A. Nordheim.** 1991. Ets-related protein Elk-1 is homologous to the c-fos regulatory factor p62TCF. Nature (London) **354**: 531–534.
25. **Ishimoto, A., A. Adachi, K. Sakai, and M. Matsuyama.** 1985. Long terminal repeat of Friend-MCF virus contains the sequence responsible for erythroid leukemia. Virology **141**:30–42.
26. **Kadonaga, J. T., and R. Tjian.** 1986. Affinity purification of sequence-specific DNA binding proteins. Proc. Natl. Acad. Sci. USA **83**:5889–5893.
27. **Karim, F. D., L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKercher, A. Celada, C. V. Van Beveren, R. A. Maki, C. V. Gunther, J. A. Nye, and B. J. Graves.** 1990. The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. Genes Dev. **4**:1451–1453.
28. **Klemsz, M. J., S. R. McKercher, A. Celada, C. Van Beveren, and R. A. Maki.** 1990. The macrophage and B cell specific transcription factor PU.1 is related to the ets oncogene. Cell **61**:113–124.
29. **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.
30. **LaMarco, K., C. C. Thompson, B. P. Byers, E. M. Walton, and S. L. McKnight.** 1991. Identification of ets- and notch-related subunits in GA binding protein. Science **253**:789–792.
31. **Lenz, J. (Albert Einstein College of Medicine).** Personal communication.
32. **Lenz, J., D. Celander, R. L. Crowther, R. Patarca, D. E. 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.
33. **Leprince, D., A. Gegonne, J. Coll, C. de Taisne, A. Schneeberger, C. Lagrou, and D. Stehelin.** 1983. A putative second cell-derived oncogene of the avian leukaemia retrovirus E26. Nature (London) **306**:395–397.
34. **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**:696–700.
35. **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.

36. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
37. **Nunn, M. F., P. H. Seeberg, C. Moscovici, and P. H. Duesberg.** 1983. Tripartite structure of the avian erythroblastosis virus E26 transforming gene. Nature (London) **306**:391–395.
38. **Nye, J., J. M. Petersen, C. V. Gunther, M. D. Jonsen, and B. J. Graves.** 1992. Interaction of murine Ets-1 with GGA binding sites establishes the ETS domain as a new DNA binding motif. Genes Dev. **6**:975–990.
39. **Rao, V. N., K. Huebner, M. Isobe, A. Ar-Rushdi, C. M. Croce, and E. S. P. Reddy.** 1989. elk, tissue-specific ets-related genes on chromosomes X and 14 near translocation breakpoints. Science **244**:66–70.
40. **Reddy, E. S. P., V. N. Rao, and T. S. Papas.** 1987. The erg gene: a human gene related to the ets oncogene. Proc. Natl. Acad. Sci. USA **84**:6131–6135.
41. **Satake, M., M. Inuzuka, K. Shigesada, T. Oikawa, and Y. Ito.** 1992. Differential expression of subspecies of polyomavirus and murine leukemia virus enhancer core binding protein, PEBP2, in various hematopoietic cells. Jpn. J. Cancer Res. **83**:714–722.
42. **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.
43. **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.
44. **Speck, N. A., B. Renjifo, E. Golemis, T. N. Frederickson, J. W. Hartley, and N. Hopkins.** 1990. Mutation of the core or adjacent LVb elements of the Moloney murine leukemia virus enhancer alters disease specificity. Genes Dev. **4**:233–242.
45. **Speck, N. A., B. Renjifo, and N. Hopkins.** 1990. Point mutations in the Moloney murine leukemia virus enhancer identify a lymphoid-specific viral core motif and 1,3-phorbol myristate acetate-inducible element. J. Virol. **64**:543–550.
46. **Sun, W., M. O'Connell, and N. A. Speck.** 1993. Characterization of a protein that binds multiple sequences in mammalian type C retrovirus enhancers. J. Virol. **67**:1976–1986.
47. **Thompson, C. B., C. Wang, I.-C. Ho, P. R. Bohjanen, B. Petryniak, C. H. June, S. Meisfeldt, L. Zhang, G. J. Nabel, B. Karpinski, and J. M. Leiden.** 1992. *cis*-acting sequences required for inducible interleukin-2 enhancer function bind a novel Ets-related protein, Elf-1. Mol. Cell. Biol. **12**:1043–1053.
48. **Thompson, C. C., T. A. Brown, and S. L. McKnight.** 1991. Convergence of Ets- and notch-related structural motifs in a heteromeric DNA binding complex. Science **253**:762–768.
49. **Thornell, A., B. Hallberg, and T. Grundström.** 1988. Differential protein binding in lymphocytes to a sequence in the enhancer of the mouse retrovirus SL3-3. Mol. Cell. Biol. **8**:1625–1637.
50. **Thornell, A., B. Hallberg, and T. Grundström.** 1991. Binding of SL3-3 enhancer factor 1 transcriptional activators to viral and chromosomal enhancer sequences. J. Virol. **65**:42–50.
51. **Wang, S., and N. A. Speck.** 1992. Purification of core-binding factor, a protein that binds the conserved core site in murine leukemia virus enhancers. Mol. Cell. Biol. **12**:89–102.
52. **Watson, D. K., M. J. McWilliams, P. Lapis, J. A. Lautenberger, C. W. Schweinfest, and T. S. Papas.** 1988. Mammalian ets-1 and ets-2 genes encode highly conserved proteins. Proc. Natl. Acad. Sci. USA **85**:7862–7866.
53. **Watson, D. K., M. J. McWilliams-Smith, M. F. Nunn, P. H. Duesberg, S. J. O'Brien, and T. S. Papas.** 1985. The ets sequence from the transforming gene of avian erythroblastosis virus, E26, has unique domains on human chromosomes 11 and 21: both loci are transcriptionally active. Proc. Natl. Acad. Sci. USA **82**:7294–7298.
54. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1985. RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Xin, J., A. Cowie, P. Lachance, and J. A. Hassell.** 1992. Molecular cloning and characterization of PEA3, a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. Genes Dev. **6**:481–496.