JOURNAL OF VIROLOGY, Jan. 1991, p. 42–50
0022-538X/91/010042-09$02.00/0

Vol. 65, No. 1

Binding of SL3-3 Enhancer Factor 1 Transcriptional Activators to Viral and Chromosomal Enhancer Sequences

ANDERS THORNELL, BENGT HALLBERG, AND THOMAS GRUNDSTRÖM*

Department of Applied Cell and Molecular Biology, University of Umeå, S-901 87 Umeå, Sweden

Received 2 July 1990/Accepted 26 September 1990

Interactions between SL3-3 enhancer factor 1 (SEF1) proteins and the enhancer of the murine leukemia virus SL3-3 were analyzed. SEF1 proteins were found to interact with two different DNA sequences within the DNA repeat region of the enhancer; these two motifs cooperated in enhancing initiation of transcription in T lymphocytes. Using an electrophoretic mobility shift assay, we identified nucleotides that are important for the SEF1 binding, and we deduced a sequence, 5′-TTTGCGGTTA/T-3′ with highly improved binding of SEF1 proteins. We show that many different SEF1 binding sequences exist in the transcription control regions of different viral and cellular genes. The results indicate a general role of SEF1 proteins in T-cell gene expression.

Developmental and environmental signals regulate the expression of eucaryotic cellular genes transcribed by RNA polymerase II. The complex scheme of regulation is mediated in part by sequence-specific DNA binding proteins that interact with promoter and enhancer elements and form a complex for initiation of transcription. Enhancer elements are organized in many short modular elements, in which regulatory proteins interact in a specific manner to regulate gene expression (for reviews, see references 25, 32, 36, 44).

The long terminal repeats (LTR) are primary determinants of the leukemogenicity of many leukemia viruses (4, 10, 12, 13, 17, 24, 29, 43, 50). The LTR harbor sets of similar but distinct DNA sequence motifs, which determine the tropism of the viruses (17). SL3-3 is a murine retrovirus that has been shown to induce T-cell lymphomas in several strains of mice (41); induction of these lymphomas is independent on a determinant in the SL3-3 LTR (29), which preferentially activates transcription in T cells (3). This LTR contains a 72-bp duplicated DNA segment followed by a third copy of the first 34 bp of the repeat (29). We have shown that the enhancer region in the LTR is composed of several DNA domains, and we have localized the most important segments for the enhancer activity in T lymphocytes to the repeated sequences (21). We showed in a previous report (47) that a family of proteins, denoted SL3-3 enhancer factor 1 (SEF1), interacts specifically with a DNA segment in the repeated region of the enhancer. A mutation of the identified SEF1 binding DNA segment decreased transcription in vivo twofold in two T-cell lines but not in a B-cell line. SEF1 protein-DNA complexes were obtained in higher amounts with nuclear extracts from cell lines of T lymphocytes and from lymphoid organs than with extracts from other cell lines and organs tested.

In this study we report an analysis of the interactions between SEF1 proteins and the enhancer of SL3-3 by using an electrophoretic mobility shift assay. We show that SEF1 proteins interact with two different DNA sequences within the repeated region of the LTR and that the two sites cooperate in enhancement of transcription. The importance of each nucleotide for binding of SEF1 proteins to the two sites was determined by mutant analysis. We demonstrate differences in the DNA sequence specificities between SEF1 protein-DNA complexes of different mobilities, and we identify a sequence, 5′-TTTGCGGTTA/T-3′, with highly improved binding of SEF1 proteins. We show that many different SEF1 binding sequences exist in the transcription control regions of different viruses and cellular genes.

MATERIALS AND METHODS

Cell lines and cell culture. The following cell lines were grown as previously described (47): EL-4, a mouse T-helper cell lymphoma (19, 38); CTL, a monoclonal cytotoxic T-cell line (1); L691-6, (kind gift from F. S. Pedersen), a mouse T lymphoma cell line (35); J558L, a mouse myeloma cell line (40); and HeLa, a human carcinoma cell line.

Construction of mutants and in vivo transcription assay. Mutant plasmids pESG146 and pESG147 are derivatives of pESG134 (21). pESG146 contains a 3-bp mutation at positions 163 to 165, from GAC to TCA. pESG147 harbors this mutation and the mutation of pESG133 (47), a 3-bp mutation at positions 188 through 190 from GTT to TGG. The mutant plasmids were constructed as previously described (20, 47). The cotransfected reference plasmid pESG003 (21) has a structure corresponding to that of pESG134, except that the enhancer is from simian virus 40 and the promoter is from rabbit β-globin. Lymphocytes and HeLa cells were transfected as previously described (21) by using the DEAE-dextran and calcium phosphate techniques, respectively. RNA samples were prepared by using the LiCl-urea method, and the specific RNAs were quantified by S1 nuclease mapping as previously described (21).

Preparation of nuclear extracts, electrophoretic mobility shift assay, and methylation interference experiments. The preparation of nuclear extracts and the electrophoretic mobility shift assay were as previously described (47). Probes and competitors to be analyzed in parallel were prepared in parallel as previously described (47). Equal amounts of the oligonucleotides were used, and this resulted in very small variations in different labelings. Methylation interference experiments were as previously described (47).

Computer analysis. The EMBL nucleic acid sequence data library was searched for the DNA sequences 5′-TC/TA/TGC/TGGTC/TA/T-3′ and closely related sequences with efficient SEF1 binding by using the NESEARCH program in

* Corresponding author.

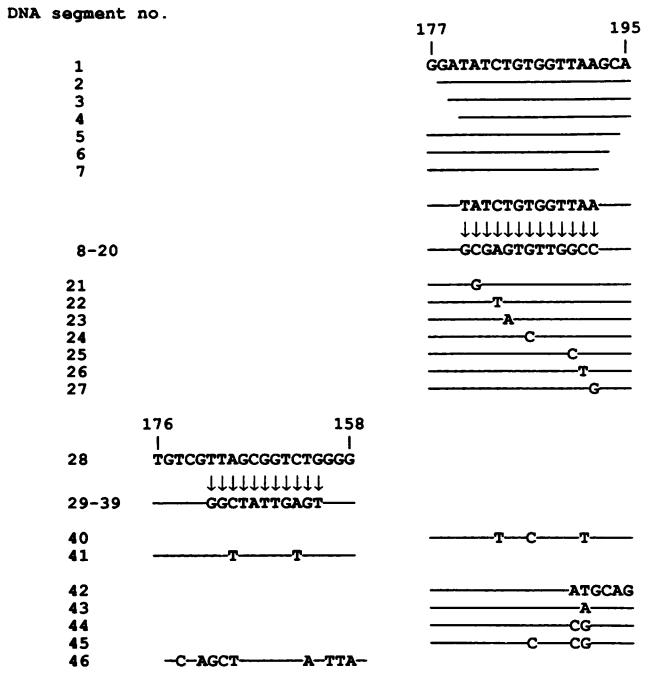


FIG. 1. Double-stranded DNA segments used in SEF1 binding experiments. DNA segments 1 and 28 harbor the wild-type SL3-3 sequences of SEF1 binding sites I and II, respectively. Only nucleotides different from those of the wild-type sites are indicated. DNA segments 8 through 20 and 29 through 39 contain the indicated C↔A and G↔T 1-bp substitutions, respectively. DNA segments 21 through 27 have base pair substitutions of site I that introduce the base pair of site II in each of the positions where sites I and II differ. DNA segments 40 and 41 are the improved SEF1 binding sites I and II, respectively. DNA segments 42 through 46 harbor sequences related to the SEF1 binding sites of SL3-3, derived from the transcriptional control regions of the viruses Moloney murine leukemia virus (positions 7956 through 7974 and 8031 through 8049), NFS murine leukemia virus, Soule murine leukemia virus, and polyomavirus, respectively (see Table 1).

the PC-gene software (Genofit S.A., Geneva, Switzerland). The sequences at known or potential transcriptional control regions of viruses or cellular genes are included in Table 1.

RESULTS

SEF1 proteins interact with two different sequences of the SL3-3 enhancer. During the course of a systematic electrophoretic mobility shift analysis of the repeated region of the LTR of SL3-3, formation of complexes with one labeled DNA segment (positions 158 through 176 and 230 through 248 of the SL3-3 enhancer; DNA segment 28 of Fig. 1) was specifically inhibited by a DNA segment containing the previously identified SEF1 site (47). The complexes obtained with the two DNA segments showed specific cross-competition by a 50-fold molar excess of the other DNA segment (Fig. 2A, lanes 4 and 9). The previously described SEF1 binding site, here designated site I, was about fivefold stronger than SEF1 binding site II, as calculated from multiple independent experiments. The difference in the binding strengths of the two SEF1 sites was present at multiple protein concentrations (data not shown). It is evident from Fig. 1A that several closely migrating, and possibly comigrating, SEF1 protein-DNA complexes were obtained in the electrophoresis. Note that, in addition to the SEF1 complexes, two much less competable complexes were obtained as previously described (47) (Fig. 2A, lanes 3, 8, and 9; Fig. 3A, lanes 3 through 9 and 12 through 18).

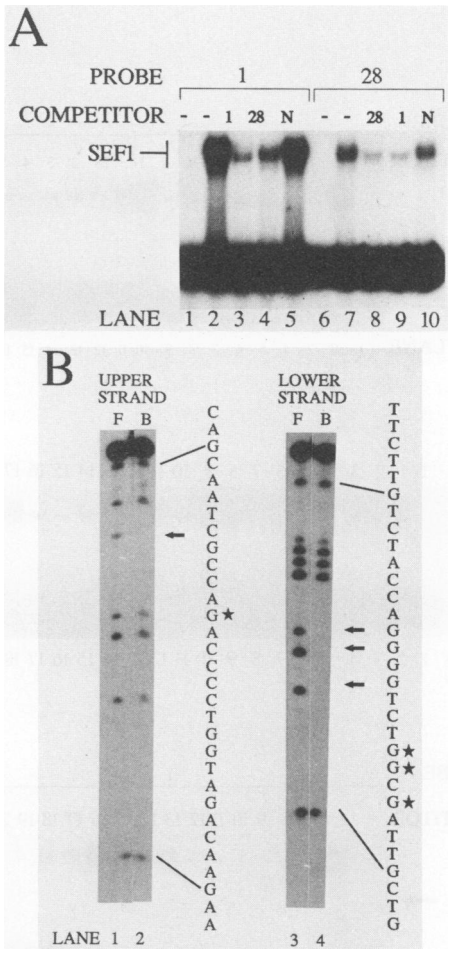
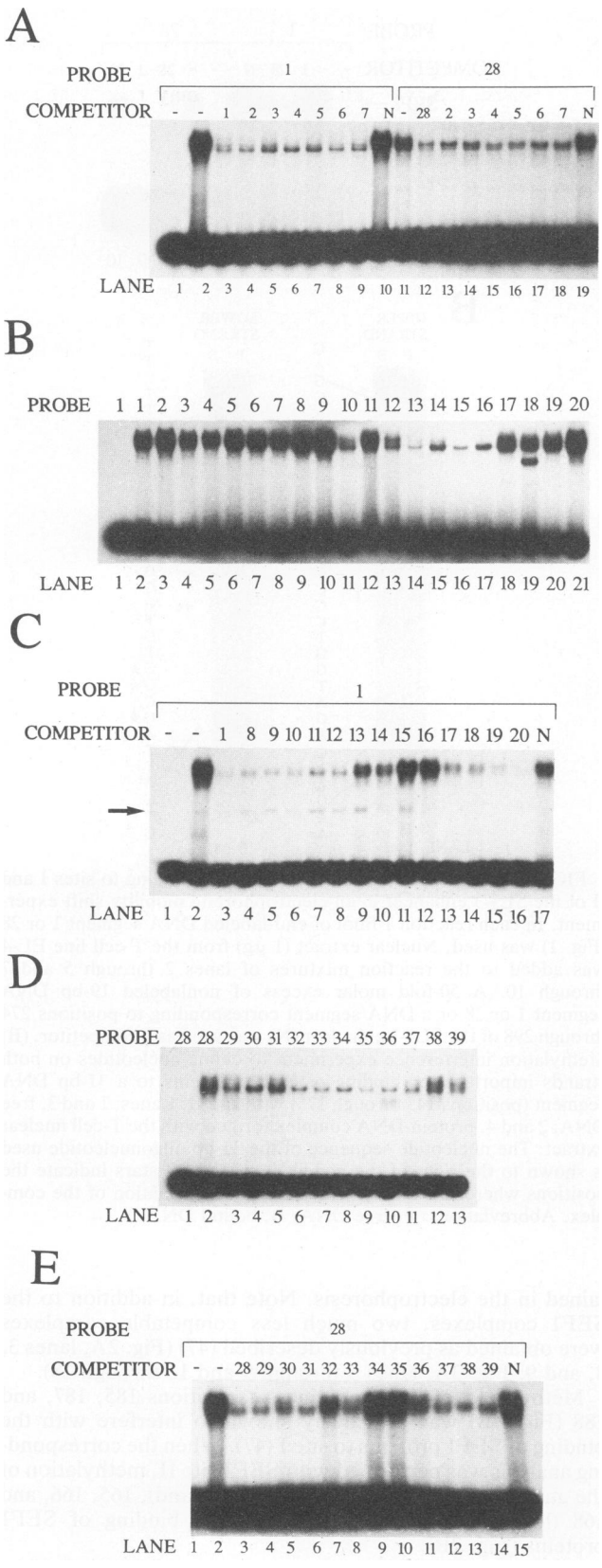


FIG. 2. (A) Competition analysis of SEF1 binding to sites I and II of the SL3-3 enhancer in an electrophoretic mobility shift experiment. In each reaction 1 fmol of end-labeled DNA segment 1 or 28 (Fig. 1) was used. Nuclear extract (1 μg) from the T-cell line EL-4 was added to the reaction mixtures of lanes 2 through 5 and 7 through 10. A 50-fold molar excess of nonlabeled 19-bp DNA segment 1 or 28 or a DNA segment corresponding to positions 274 through 298 of the SL3-3 enhancer (N) was used as a competitor. (B) Methylation interference experiment to define nucleotides on both strands important for binding of SEF1 proteins to a 31-bp DNA segment (positions 145 through 175) with site II. Lanes: 1 and 3, free DNA; 2 and 4, protein-DNA complex formed with the T-cell nuclear extract. The nucleotide sequence of the 31-bp oligonucleotide used is shown to the right of the panels. Arrows and stars indicate the positions where methylation interfered with formation of the complex. Abbreviations: F, free DNA; B, bound DNA.

Methylation of the guanidines at positions 185, 187, and 188 (Fig. 4B) was previously shown to interfere with the binding of SEF1 proteins to site I (47). When the corresponding analysis was performed with SEF1 site II, methylation of the guanidines at positions 163 (upper strand), 165, 166, and 168 (lower strand) interfered with the binding of SEF1 proteins (Fig. 2B).

Mutation of SEF1 binding sites affects transcription in vivo. We have previously shown that the SEF1 interaction with



site I is important for transcription in T-lymphocyte cell lines in vivo (47). To determine the in vivo importance of SEF1 binding to site II, we constructed a 3-bp substitution in site II and a double mutant that also contained the previously analyzed mutation of site I (47). All constructs are derivatives of plasmid pESG134 (21), which carries the SL3-3 enhancer in front of the early promoters of simian virus 40. This plasmid has one 72-bp repeat less than the LTR of SL3-3, a deletion that has only minor effects on the levels of transcription in the cell lines tested (21). The constructs were cotransfected with a control plasmid, followed by transient expression and quantification of the specific transcript relative to a reference transcript from the control plasmid (Materials and Methods). Typical quantitative S1 nuclease mappings of RNA from two independent transfections of a T-cell line, EL-4, with the different constructs are shown in Fig. 4A. The different mutations of the constructs and the average levels of transcription obtained in EL-4 and other cell lines are shown in Fig. 4B. The 3-bp substitution in SEF1 binding site I (pESG133) decreased transcription about twofold in the T-lymphocyte cell lines EL-4 and CTLL, whereas the transcription in the B-cell line X-63-1 was not affected (47). Here we show that the substitution decreased the transcription threefold in a third T-cell line, L691-6, whereas the decrease in a second B-cell line, J558L, was much smaller than that in all T-cell lines studied. The mutation of SEF1 binding site II (pESG146) decreased the transcription in all three T-cell lines about twofold, whereas in the B-cell line J558L and in HeLa cells the decreases in the levels of transcription were very small (Fig. 4B). Mutation of both SEF1 binding sites (pESG147) had a roughly multiplicative effect in the T cells, decreasing the transcription by three to fivefold. No decrease was seen with the

FIG. 3. Binding of SEF1 to 1 fmol of labeled DNA segments of different lengths or with different single-base-pair substitutions in site I or site II in electrophoretic mobility shift experiments. Samples of 1 μg of nuclear extract of the T-cell line EL-4 were used. (A) Competition of binding of SEF1 with two sets of DNA segments from site I which are shorter at either side; labeled DNA segments with SEF1 sites I (lanes 1 through 10) and II (lanes 11 through 19) were used. A 25-fold molar excess of the competitors was used. Lanes: 1, no nuclear extract added; 2 and 11, no competitor DNA added; 3 through 9, competitor DNA segments 1 through 7, respectively, added (Fig. 1); 12 through 18, competitor DNA segments 28 and 2 through 7, respectively, added; 10 and 19, nonspecific competitor DNA (positions 274 through 298 of SL3-3) added. (B) Binding of SEF1 to the two sets of shorter DNA segments and to DNA segments with single-base-pair substitutions of site I. Lanes: 1, no nuclear extract added; 2 through 21, DNA segments 1 through 20, respectively, added as labeled probes. (C) Competition of binding of SEF1 with DNA segments containing single-base-pair mutations of site I with labeled DNA segment 1 and a 50-fold molar excess of competitors. Lanes: 1, no nuclear extract added; 2, no competitor added; 3 through 16, competitor DNA segments 1 and 8 through 20, respectively, added; 17, nonspecific competitor added (as above). The arrow indicates a complex with a nucleotide sequence specificity distinct from that of SEF1. (D) Binding of SEF1 to DNA segments containing single-base-pair substitutions of site II. Lanes: 1, no nuclear extract added; 2 through 13, labeled DNA segments 28 through 39, respectively, added. (E) Competition of SEF1 binding to DNA segments containing single-base-pair mutations of site II with a 50-fold molar excess of competitors and labeled DNA segment 28. Lanes: 1, no nuclear extract added; 2, no competitor added; 3 through 14, competitor DNA segments 28 through 39, respectively, added; 15, nonspecific competitor DNA (N) added.

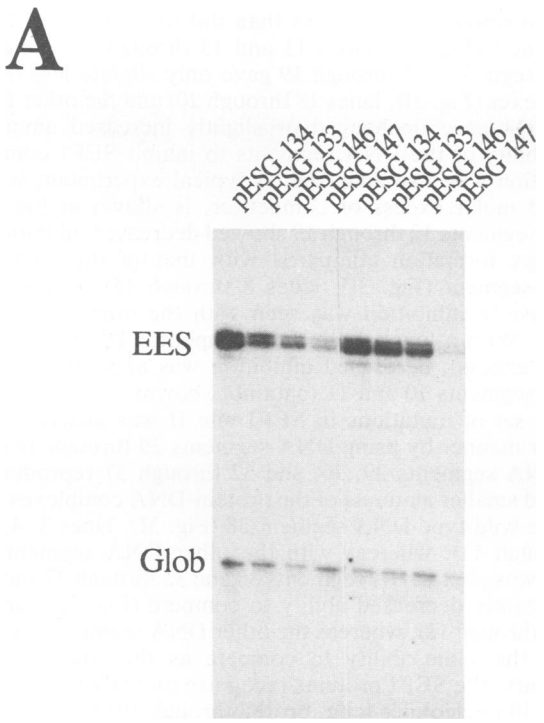
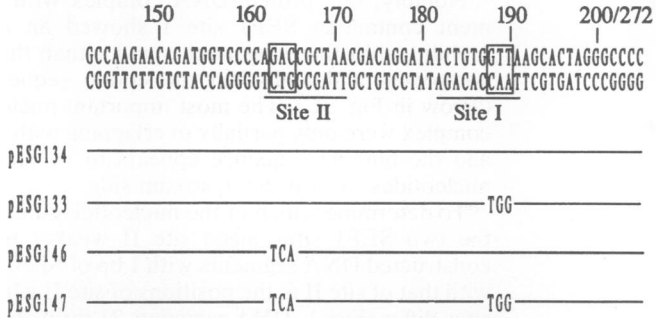


FIG. 4. Effect of mutated SEF1 binding sites on transient transcription in different lymphocyte cell lines and HeLa cells. Plasmid pESG134 and the derivatives carrying a mutation in SEF1 binding site I and/or II are described in Materials and Methods. Specific transcription was analyzed for plasmids with the SL3-3 enhancer in front of the early simian virus 40 promoter and a rabbit β-globin gene (see the text). (A) Quantitative S1 nuclease analysis of RNA prepared from two independent transient expression experiments with EL-4 cells and plasmids pESG134 (wild type), pESG133 (SEF1 site I mutant), pESG146 (SEF1 site II mutant), and pESG147 (both SEF1 sites mutated). (B) Summary of the results of the transient expression experiments in different cell lines. The values correspond to the amounts of RNA in each cell line, determined by quantitative S1 nuclease analysis, expressed relative to the value for pESG134, which was taken as 100%. The values represent the average (±20%) of several independent transfection experiments with different plasmid DNA preparations. Abbreviations: ND, not determined; EES, RNA transcribed from the early early simian virus 40 promoter start site; Glob, RNA transcribed from the cotransfected reference recombinant (see Materials and Methods).

B



RELATIVE TRANSCRIPTION (%)	L691-6	CTLL	EL-4	J558L	HeLa
pESG134	100	100	100	100	100
pESG133	34	43	56	76	ND
pESG146	59	58	49	79	84
pESG147	24	32	18	59	100

double mutant in HeLa cells, and a decrease much smaller than that in the T cells was seen in the B-cell line J558L.

Mutant analysis of the two SEF1 binding sites. To localize the nucleotides important for the binding of the SEF1 proteins, we constructed 16- to 18-bp double-stranded oligonucleotides, in which nucleotides outside of positions 180 through 192 of site I were deleted stepwise (Fig. 1; DNA segments 2 through 7). The DNA segments were used both as labeled probes and as competitors in electrophoretic mobility shift assays, since mutations could differentially affect the ability to form stable complexes and to compete for SEF1 binding. The studies were performed with nuclear protein extracts from EL-4 T cells, since in contrast to many other cell lines, no proteolytic attack on SEF1 proteins has been observed with EL-4, even in the absence of the eight protease inhibitors normally added to the preparations (47; data not shown). When the 16- to 18-bp DNA segments were used as competitors at a 25-fold molar excess, the inhibitions of SEF1 complex formation were about the same as those seen with the 19-bp DNA segments of either site I or site II (Fig. 3A, lanes 3 through 9 and 12 through 18). Correspondingly, no difference in the ability of the shorter DNA segments to form stable complexes with the SEF1 proteins could be seen (Fig. 3B, lanes 2 through 8). Thus, the DNA necessary for interaction with the SEF1 proteins is within a 13-bp DNA sequence in site I.

To evaluate the importance of each base pair in the two SEF1 binding sites, two sets of 19-bp double-stranded oligonucleotides were constructed. The mutations and their effects are summarized in Fig. 5A. Each DNA segment in one of the sets harbored a 1-bp substitution (A↔C or G↔T) within the 13-bp sequence containing site I (Fig. 1; DNA segments 8 through 20). The labeled DNA segments 10 and 12 through 16 produced reproducibly much lower amounts of

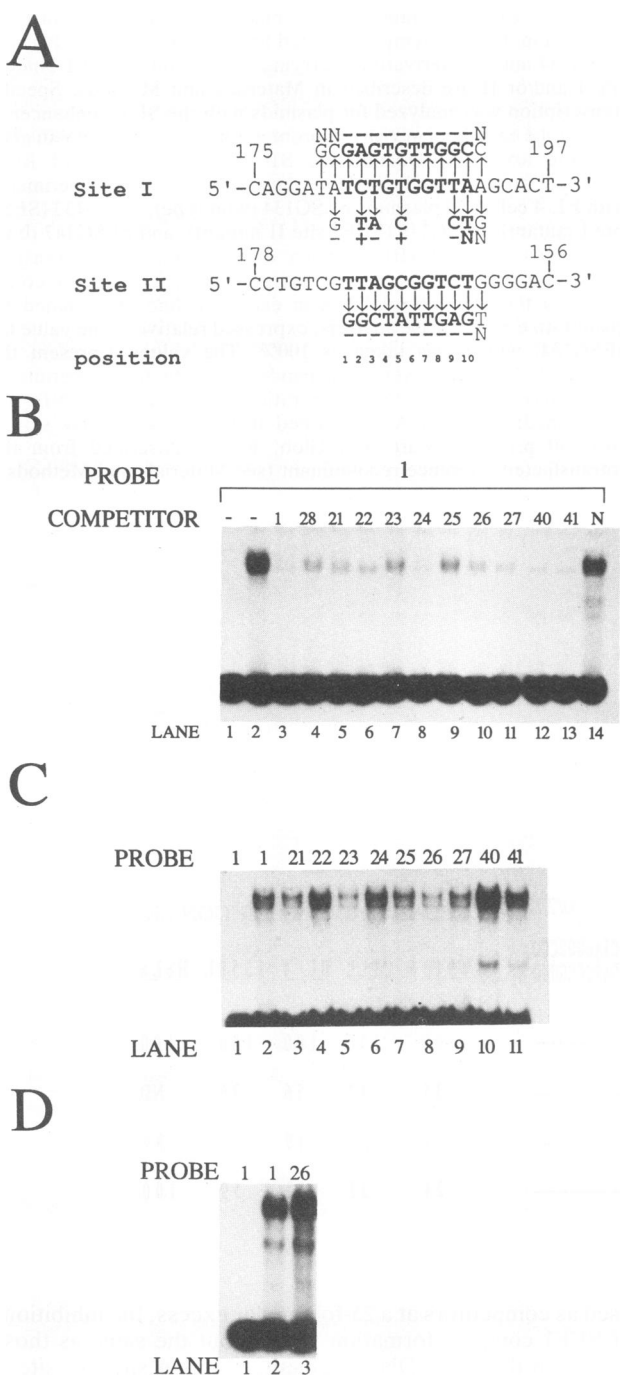


FIG. 5. Alignment of the SEF1 sites with summary of the effects of the mutations and binding of SEF1 to 1 fmol of labeled DNA segments with single-base-pair substitutions in site I at positions where site I and site II differ and with combinations of mutations introduced into site I (DNA segment 40) or site II (DNA segment 41). Experiments were performed as described in the legend to Fig. 3. (A) Alignment of the SEF1 sites and a summary of the effects of the mutations: +, increased SEF1 complex formation and/or ability to compete; −, decreased binding; N, no effect of the mutation. (B) Competition with DNA segment 1 as a labeled probe and a 50-fold molar excess of competitors. Lanes: 1, no nuclear extract added; 2, no competitor added; 3 through 13, competitor DNA segments 1, 28, 21 through 27, 40, and 41, respectively, added; 14, nonspecific competitor DNA segment added (as in Fig. 3). (C) Binding to labeled DNA segments. Lanes: 1, no nuclear extract added; 2 through 11, labeled DNA segments 1, 21 through 27, 40, and 41, respectively, added. (D) Binding to labeled DNA segments. Lanes: 1, no nuclear extract added; 2 and 3, DNA segments 1 and 26, respectively, added.

the protein-DNA complexes than did the wild type, DNA segment 1 (Fig. 3B, lanes 11 and 13 through 17), whereas DNA segments 17 through 19 gave only slightly less of the complexes (Fig. 3B, lanes 18 through 20) and the other DNA segments gave unchanged or slightly increased amounts. The ability of the DNA segments to inhibit SEF1 complex formation was also analyzed. A typical experiment, with a 50-fold molar excess of competitor, is shown in Fig. 3C. DNA segments 12 through 19 showed decreased inhibition of complex formation compared with that of the wild-type DNA segment (Fig. 3C, lanes 8 through 15), whereas no decrease in inhibition was seen with the other DNA segments. With a lower excess of competitor DNA (12.5-fold molar excess), decreased inhibition was apparent also for DNA segments 10 and 11 (data not shown).

The set of mutations in SEF1 site II was analyzed in a similar manner by using DNA segments 29 through 39 (Fig. 1). DNA segments 29, 30, and 32 through 37 reproducibly yielded smaller amounts of the protein-DNA complexes than did the wild-type DNA segment 28 (Fig. 3D, lanes 3, 4, and 6 through 11), whereas with the other DNA segments no effect was seen. DNA segments 30 and 32 through 37 showed significantly decreased ability to compete (Fig. 3E, lanes 5 and 7 through 12), whereas the other DNA segments showed about the same ability to compete as the wild type. In summary, the SEF1 proteins recognize two DNA sequences about 10 nucleotides long, bp 182 through 191 for site I and bp 163 through 171 for site II (Fig. 5A). The sequences of site I and site II (in inverted orientation) show homologies over a 10-nucleotide DNA sequence (positions 1 to 10 in Fig. 5A). At five of these positions the sequences are identical, at three positions they both contain pyrimidines, and at two positions one site contains an A and the other contains a T.

Notably, one protein-DNA complex with the DNA segment containing SEF1 site I showed an electrophoretic mobility shift that was much smaller than that of the SEF1 complexes and a distinct nucleotide sequence specificity (arrow in Fig. 3C). The most important nucleotides for this complex were only partially overlapping with those of SEF1, and the binding sequence appears to be juxtaposed a few nucleotides toward the upstream side.

To determine which of the nucleotide differences between the two SEF1 sites made site II weaker than site I, we constructed DNA segments with 1 bp of SEF1 site I replaced with that of site II at the positions of site II where the binding sites differ (Fig. 1, DNA segments 21 through 27). The SEF1 complexes with the wild-type site I DNA segment were efficiently inhibited by using a 50-fold molar excess of competitor DNA segments 22, 24 and 27 (Fig. 5B, lanes 6, 8, and 11, respectively), whereas DNA segments 21, 23, and 25 showed decreased ability to inhibit SEF1 complex formation (Fig. 5B, lanes 5, 7, and 9, respectively). The labeled DNA segments 22 and 24 showed increased amounts of complexes compared with those with the wild-type DNA segment (Fig. 5C, lanes 4 and 6), whereas the labeled DNA segments 21, 23, and 25 showed decreased amounts of SEF1 complexes (Fig. 5C, lanes 3, 5, and 7, respectively). Mutation of position 10 of the decamer (DNA segment 26) decreased binding of the SEF1 proteins in some experiments and gave

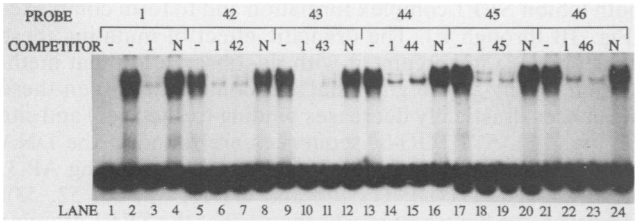


FIG. 6. Competition analysis with DNA segments of different viruses containing sequences 1 and 42 through 46 (Fig. 1) related to the SEF1 binding sites of SL3-3. Experiments were performed as described in the legend to Fig. 3, except that a 100-fold molar excess of the competitors was used. Lanes: 1, no nuclear extract added; 2, 5, 9, 13, 17, and 21, no competitor added; 4, 8, 12, 16, 20, and 24, nonspecific competitor DNA added (as in Fig. 3). For the other lanes the competitor DNA segment used is shown above each lane.

the same or slightly increased binding in other identical experiments (Fig. 5B, lane 10; Fig. 5C, lane 8; Fig. 5D, lane 3). We therefore favor the hypothesis that an A or a T in this position results in about equal SEF1 binding.

In summary, we find that substituting the nucleotides of site I that differ between the sites with those of site II resulted in up mutations at nucleotides 2 and 5, down mutations at nucleotides 3 and 9, and a close to neutral mutation at nucleotide 10 (Fig. 5A). This prompted us to introduce the two up mutations in site I simultaneously. We also introduced a T at position 10, since some experiments with this mutation had indicated a slight increase in SEF1 binding. We indeed found increased SEF1 binding and competition for the triple mutant site compared with that of the wild-type site I (Fig. 5C, compare lanes 2 and 10; Fig. 5B, compare lanes 3 and 12). In multiple experiments, the average increase in formation of complexes was about fourfold.

Although the most important nucleotides for SEF1 biding clearly are within the decanucleotide sequence, our results show that surrounding nucleotides are not all irrelevant for the binding strength. For example, DNA segment 21, in which a nucleotide directly 5′ to the decanucleotide sequence was substituted, showed decreased ability to compete and to bind SEF1 proteins (Fig. 5B, lane 5; Fig. 5C, lane 3). In addition, when the improved decanucleotide was introduced into a DNA segment with the flanking nucleotides of site II (DNA segment 41) instead of those of site I, only a 1.3-fold, as opposed to a 4-fold, increase in complex formation was obtained (Fig. 5C, compare lanes 2, 10, and 11). Nevertheless, this means that DNA segment 41 binds SEF1 six- to sevenfold more efficiently than does site II.

Corresponding mutant analyses with different amounts of nuclear proteins of CTL, J558L, and HeLa cells in several independent experiments yielded results that were essentially the same as those with nuclear proteins of EL-4 (data not shown). Notably, no mutation changed the electrophoretic mobilities of the complexes formed in any of the experiments. This indicates that the different complexes do not include multiple proteins with independent sequence-specific DNA binding ability.

SEF1 binding sites are present in many genes. A computer search of the EMBL nucleic acid sequence data library for potential SEF1 binding sites in other viruses and in cellular genes, by using different permutations of the decamer with efficient SEF1 binding (Materials and Methods), resulted in the identification of a large number of sequences. Viral and cellular genes found to harbor a sequence(s) identical or very similar to the known SEF1 binding sites in their transcriptional control region are presented in Table 1. In several of

TABLE 1. SEF1 binding sites in viral and cellular genes

Virus or cellular gene	Sequence[a]	Position[b]	Reference
Feline sarcoma virus	TCTGTGGTTA	122	22
Feline leukemia virus		2097	42
Gibbon leukemia virus		706	5, 48
Moloney murine leukemia virus	TCTGTGGTAA	8036	45
Murine sarcoma virus		252	15, 49
Simian sarcoma virus	TCTGTGGTCA	338, 582	14
Gibbon leukemia virus		807	5, 48
RFM/Un		154	30
Moloney murine leukemia virus	TCTGTGGTAT	7961	45
NFS murine leukemia virus	TCTGTGGTCG	145, 265	27
Soule murine leukemia viruses ST1 and ST4	TCTGCGGTCA	154, 243	10
		330, 417	10
Polyomavirus	GCTGCGGTCA	5156	11
Human interleukin-6	AGAGTGGTTC	−309	54
Human granulocyte-macrophage colony-stimulating factor	TTTGTGGTCA	569	26
Human tumor necrosis factor α	TCTGTGGTCT	253	39
Human tumor necrosis factor β	CCTGTGGTTC	649	39
Rat interleukin-3	TCTGTGGTTT	1212	9
Rat transin	TTTGTGGTTT	469	33
Mouse T-cell receptor β chain	TTAGTGGTTT	191	28
Human T-cell receptor CD3-ε	GCTGTGGTTA	168	8
Mouse T-cell receptor CD3-δ	GCTGTGGTAA	25	16
	TTTGTGGCTC	62	16
	TGTGTGGCAA	306	16

[a] The sequence is only given at the first listed one of identical SEF1 sequences. Nucleotides differing from the sequence of SEF1 site I of SL3-3 are underlined. Positions with nucleotides not shown in the present study to be compatible with efficient SEF1 binding are double underlined.

[b] Only the nucleotide number of the first position is shown.

the viruses the sequences occur more than once. For example, the Soule murine leukemia viruses ST1 and ST4 harbor as many as four such sequences each.

Many of the identified DNA sequences are indeed SEF1 binding sites, because they show either sequence identity with a site of SL3-3 virus (including nucleotides flanking the decamer) or carry a nucleotide difference shown here to be compatible with SEF1 binding. We decided to test whether SEF1 binding could also be seen for each of the other viral sequences identified during the search, those of Moloney murine leukemia virus (positions 8036 through 8045 and 7961 through 7970), NFS murine leukemia virus, polyomavirus, and Soule murine leukemia viruses ST1 and ST4 (Table 1). DNA segments with sequences 42 to 46 of Fig. 1 were constructed and used in competition studies. All five different DNA segments resulted in protein-DNA complexes (Fig. 6). Each of these was efficiently inhibited by a 100-fold molar excess of SL3-3 SEF1 site I DNA segment or the homologous DNA segment but not by the nonhomologous DNA segment. Interestingly, competition with SEF1 site I for the DNA segments of NFS and ST1-ST4 (segments 44 and 45) resulted in a complex that was inhibited by the homologous DNA segments (Fig. 6, lanes 14, 15, 18, and 19). Thus, although the sequence specificities of different SEF1 protein complexes are very similar, these results show that there are differences in the sequence specificities of different complexes formed with SEF1 or related proteins. Each nucleotide in the potential SEF1 binding sequences found in the known or potential transcriptional control regions of seven of the nine cellular genes in Table 1 is compatible with SEF1 binding. On the basis of the information about the nucleotide sequence specificities of SEF1 binding gained from this analysis of the 46 different DNA segments (above), it is very likely that the sequences of tumor necrosis factor β and the T-cell receptor CD3-δ genes listed in Table 1 would also function as SEF1 binding sites.

DISCUSSION

We have demonstrated that the SEF1 nuclear proteins specifically interact with two different DNA sequences of the enhancer element of the murine retrovirus SL3-3. Mutation of SEF1 binding site II, identified in this study, was found to decrease in vivo enhancement of transcription in several T-lymphocyte cell lines by about twofold. This is approximately the same level as we previously reported for mutation of SEF1 binding site I. However, SEF1 site I displayed fivefold higher binding of SEF1 proteins in vitro than did SEF1 site II. The in vivo effects of mutations within these two sites therefore show no correlation with the binding strengths determined in vitro. One explanation could be that the binding to the weaker site is strong enough to reach above a threshold level necessary for activity. Alternatively, a particular SEF1 protein(s) may exist that has a binding strength in vitro correlating with the importance of the sequences in vivo, or the effects of the mutations in vivo may reflect the importance of a context in which the SEF1 proteins function in concert with each other and with other proteins.

We have localized and identified the most important nucleotides for binding of SEF1 proteins within a DNA sequence of 10 bp. Interestingly, there appears to be a strong requirement for 5′-GTGG-3′ (site I) or 5′-GCGG-3′ (site II) in the center of the binding site. DNA segments harboring substitutions at any position of the 5′-GTGG-3′ or 5′-GCGG-3′ sequences show drastically decreased abilities to both inhibit SEF1 complex formation and to form complexes (Fig. 3B through E). The dramatic effect of mutating these nucleotides is in agreement with the observation that methylation of any of the guanidines in common between these sequences drastically decreases binding to site I (47) and site II (Fig. 1B). 5′-GTGG-3′ sequences are found in the DNA binding sites of many transcription factors, including AP-3, AP-4, GT-IA, GT-IB, GT-IC, and GT-IIC (6, 37, 52, 53). However, we have previously shown that the SEF1 proteins do not specifically recognize these binding sites and are a separate family of transcription factors (47).

One protein-DNA complex with the DNA segment containing SEF1 site I showed a very different nucleotide sequence specificity (Fig. 3C). The most important nucleotides for this complex only partially overlapped with those of SEF1, and the binding sequence appeared to be juxtaposed a few nucleotides toward the upstream side. The much smaller electrophoretic mobility shift of this complex compared with that of the SEF1 complexes indicates that it could represent the S-CBF factor, recently reported by Boral et al. (2), that binds to the same DNA segment as S/A-CBF, which could correspond to SEF1. However, the nucleotide sequence specificity of the additional protein-DNA complex appears to be distinct also from that of S-CBF (2), indicating that the complex could represent a factor separate from S-CBF, S/A-CBF, and SEF1. The possibility that S/A-CBF corresponds to SEF1 is supported by a recent report, published after this study was finished, that S/A-CBF can bind to a second DNA region in the SL3-3 enhancer (31) containing the DNA sequence identified here as SEF1 binding site II.

By substituting nucleotides in site I with those of site II at positions where the two SEF1 binding sites differ, we found that substitutions at two positions clearly decreased binding, substitutions at two positions increased binding, and substitution at one position was rather neutral. The simultaneous introduction of C-to-T shifts at positions 2 and 10 and a T-to-C shift at position 5 of the decamer sequence of site I yielded an SEF1 binding site with a fourfold-increased binding capacity. From these findings we deduce a sequence, 5′-TTTGCGGTTA/T-3′, for improved binding of SEF1 proteins. Results from purification of SEF1 proteins have shown that an affinity column with a DNA matrix containing the improved SEF1 binding sequence is much more efficient and specific than one with the native SEF1 site I (unpublished observation).

Interestingly, different SEF1 complexes show differences in their DNA sequence specificities. They show high recognition specificity for 5′-GCGGT-3′ or 5′-GTGGT-3′ and, to some extent, different nucleotide preferences around these nucleotides. Preliminary results from purification of SEF1 proteins show that they can be chromatographically separated by sequence-specific DNA affinity columns (unpublished observation). The differences between the set of proteins with separable binding preferences and electrophoretic mobility could be created at many different levels. We cannot exclude the possibility of proteolytic attack during protein preparation. However, we find it more likely that the differences reflect in vivo differences among the proteins, based on the existence of a family of closely related genes, differential gene rearrangements, differential mRNA splicing, or posttranslational modification(s).

Many sequences with many fewer discrepancies compared with SEF1 site I than those between sites I and II were found when we searched through the EMBL nucleic acid sequence data library for sequences closely related to the

improved SEF1 binding site. Interestingly, SEF1 sites were found in the transcriptional control regions of many viruses. The majority of these are leukemia- or sarcoma-causing agents. Clark et al. (7) have reported that the DNA element that we show here to be recognized by SEF1 proteins is found in the LTR of several murine retroviruses, and Golemis et al. (18) recently reported that this element is one of the highly conserved sequence elements in the enhancer sequences of mammalian type C retroviruses. We show that the SEF1 proteins actually bind to DNA segments containing the sequences found in these viruses. A weak band remained when the binding to the SEF1-like sequences from the Soule leukemia viruses and NFS virus was competed with SEF1 site I (Fig. 6). This band did not remain with the homologous DNA segments as a competitor. This probably reflects that the range of proteins binding to the sequences of Soule leukemia viruses and NFS virus includes a protein(s) in addition to the SEF1 proteins binding to the SL3-3 sequence.

The identified SEF1-like binding sites (Table 1) are in DNA regions believed to be important for transcription control. However, the importance of most of these sites has not been tested by point mutations in vivo. For the PEA2 site of polyomavirus, Wasylyk and co-workers, in studies that differ in the size and the number of copies of the enhancer segment used and in the choice of promoter and cell type, reported that point mutations could lead to both decreased (23) and increased (51) transcription. Interesting questions arising as to which condition(s) determines the reported ability to either repress or activate transcription and whether it is a property of the same or different proteins binding to the PEA2 site. A mutation in the SEF1 binding site of Moloney murine leukemia virus has recently been shown to alter the disease specificity of the virus (46).

SEF1 binding sites were found in the transcriptional control region of many cellular genes. Interestingly, many of these were genes expressed in different cell types of hematopoietic origin, including T cells. Furthermore, other genes identified are expressed only in cells of the T-lymphocyte lineage. These genes with T cell-specific expression are the β chain of the T-cell receptor and the δ and ε subunits of the T cell-CD3 complex. For these genes we find the SEF1 sites in the short enhancer segments, which have been shown to mediate T-cell-specific expression of the genes and which are localized downstream of the last exon of the gene (8, 16, 28, 34). Thus, SEF1 proteins appear to be important components for creating efficient activation of transcription in T cells, not only for viral genes but also for many chromosomal genes.

ACKNOWLEDGMENTS

We thank Christine Grundström for skillful technical assistance and Peter Brodin for help with the illustrations.

This work was supported by grants from the Swedish Natural Science Research Council (Dnr 4892-306), the Swedish Cancer Society (Dnr 2235-B89-03XA), the Board for Technological Development (Dnr 512-89-574), and the Research foundation of the oncologic clinic of Umeå (549/88). T. Grundström was the recipient of a special researcher position from the Swedish Natural Science Research Council (Dnr 4892-307).

REFERENCES

1. **Baker, P. E., S. Gillis, and K. A. Smith.** 1979. Monoclonal cytolytic T-cell lines. J. Exp. Med. **149**:273–278.
2. **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.
3. **Celander, D., and W. A. Haseltine.** 1984. Tissue-specific transcription preference as a determinant of cell tropism and leukaemogenic potential of murine retroviruses. Nature (London) **312**:159–162.
4. **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.
5. **Chen, S. J., N. J. Holbrook, K. F. Mitchell, C. A. Vallone, J. S. Greengard, G. R. Crabtree, and Y. Lin.** 1985. A viral long terminal repeat in the interleukin 2 gene of a cell line that constitutively produces interleukin 2. Proc. Natl. Acad. Sci. USA **82**:7284–7288.
6. **Chiu, R., M. Imagawa, R. J. Imbra, J. R. Bockoven, and M. Karin.** 1987. Multiple *cis*- and *trans*-acting elements mediate the transcriptional response to phorbol esters. Nature (London) **329**:648–651.
7. **Clark, S. P., R. Kaufhold, A. Chan, and T. W. Mak.** 1985. Comparison of the transcriptional properties of the Friend and Moloney retrovirus long terminal repeats: importance of tandem duplications and of the core enhancer sequence. Virology **144**:481–494.
8. **Clevers, H., N. Lonberg, S. Dunlap, E. Lacy, and C. Terhorst.** 1989. An enhancer located in a CpG-island 3′ to the TCR/CD3-ε gene confers T lymphocyte-specificity to its promoter. EMBO J. **8**:2527–2535.
9. **Cohen, D. R., A. J. Hapel, and I. G. Young.** 1986. Cloning and expression of the rat interleukin-3 gene. Nucleic Acids Res. **14**:3641–3658.
10. **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.
11. **Deininger, P. L., A. Esty, P. la Porte, H. Hsu, and T. Friedman.** 1980. The nucleotide sequence and restriction enzyme sites of the polyoma genome. Nucleic Acids Res. **8**:855–860.
12. **DesGroseillers, L., and P. Jolicoeur.** 1984. The tandem direct repeats within the long terminal repeat of murine leukemia viruses are the primary determinants 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. **Devare, S. G., E. P. Reddy, J. D. Law, K. C. Robbins, and S. A. Aaronson.** 1983. Nucleotide sequence of the simian sarcoma virus genome; demonstration that its acquired cellular sequences encode the transforming gene product p28sis. Proc. Natl. Acad. Sci. USA **80**:731–735.
15. **Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA **77**:3937–3941.
16. **Georgopoulos, K., P. van den Elsen, E. Bier, A. Maxam, and C. Terhorst.** 1988. A T cell-specific enhancer is located in a DNase I-hypersensitive area at the 3′ end of the CD3-δ gene. EMBO J. **7**:2401–2407.
17. **Golemis, E., Y. Li, T. N. Fredrickson, 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.
18. **Golemis, E. A., N. A. Speck, and N. Hopkins.** 1990. Alignment of U3 region sequences of mammalian type C retroviruses: identification of highly conserved motifs and implications for enhancer design. J. Virol. **64**:534–542.
19. **Gorer, P. A.** 1950. Studies in antibody response of mice to tumour inoculation. Br. J. Cancer **4**:372–379.
20. **Grundström, T., M. W. Zenke, M. Wintzerith, H. W. D. Matthes, A. Staub, and P. Chambon.** 1985. Oligonucleotide-directed mutagenesis by microscale "shot-gun" gene synthesis. Nucleic Acids Res. **13**:3305–3316.
21. **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.
22. **Hampe, A., M. Gobet, J. Even, C. J. Sherr, and F. Galibert.**

1983. Nucleotide sequences of Feline sarcoma virus long terminal repeats and 5′ leaders show extensive homology to those of other mammalian retroviruses. J. Virol. **45**:466–472.

23. **Imler, J. L., C. Schatz, C. Wasylyk, B. Chatton, and B. Wasylyk.** 1988. A Harvey-ras responsive transcription element is also responsive to a tumour-promoter and to serum. Nature (London) **332**:275–278.
24. **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.
25. **Jones, N. C., P. W. J. Rigby, and E. B. Ziff.** 1988. *trans*-Acting protein factors and the regulation of eukaryotic transcription: lessons from studies on DNA tumor viruses. Genes Dev. **2**:267–281.
26. **Kaushansky, K., P. J. O'Hara, K. Berkner, G. M. Segal, F. S. Hagen, and J. W. Adamson.** 1986. Genomic cloning, characterization, and multilineage growth-promoting activity of human granulocyte-macrophage colony-stimulating factor. Proc. Natl. Acad. Sci. USA **83**:3101–3105.
27. **Khan, A. S., and M. A. Martin.** 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. Proc. Natl. Acad. Sci. USA **80**:2699–2703.
28. **Krimpenfort, P., R. de Jong, Y. Uematsu, Z. Dembic, S. Ryser, H. von Boehmer, M. Steinmetz, and A. Berns.** 1988. Transcription of T cell receptor β-chain genes is controlled by a downstream regulatory element. EMBO J. **7**:745–750.
29. **Lenz, J., D. Celander, 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.
30. **Liou, R. S., L. R. Boone, J. O. Kiggans, D. M. Yang, T. W. Wang, R. W. Tennant, and W. K. Yang.** 1983. Molecular cloning and analysis of the endogenous retrovirus chemically induced from RFM/Un mouse cell culture. J. Virol. **46**:288–292.
31. **LoSardo, J. E., A. L. Boral, and J. Lenz.** 1990. Relative importance of elements within the SL3-3 virus enhancer for T-cell specificity. J. Virol. **64**:1756–1763.
32. **Maniatis, T., S. Goodbourn, and J. A. Fischer.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236**: 1237–1245.
33. **Matrisian, L. M., P. Leroy, C. Ruhlmann, M. C. Gensel, and R. Breathnach.** 1986. Isolation of the oncogene and epidermal growth factor-induced transin gene: complex control in rat fibroblasts. Mol. Cell. Biol. **6**:1679–1686.
34. **McDougall, S., C. L. Peterson, and K. Calame.** 1988. A transcriptional enhancer 3′ of Cβ2 in the T cell receptor β locus. Science **241**:205–208.
35. **McGrath, M. S., E. Pillemur, D. Kooistra, and I. L. Weissman.** 1980. The role of MuLV receptors on T-lymphoma cells in lymphoma cell proliferation, p. 157–184. *In* N. L. Warner (ed.), Contemporary topics in immunology. Plenum Publishing Corp., New York.
36. **McKnight, S., and R. Tjian.** 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795–805.
37. **Mermod, N., T. J. Williams, and R. Tjian.** 1988. Enhancer binding factors AP-4 and AP-1 act in concert to activate SV40 late transcription *in vitro*. Nature (London) **332**:557–561.
38. **Mohit, B., and K. Fan.** 1971. Hybrid cell line from a cloned immunoglobulin-producing mouse myeloma and a non-producing mouse lymphoma. Science **171**:75–77.
39. **Nedwin, G. E., S. L. Naylor, A. Y. Sakaguchi, D. Smith, J. Jarret-Nedwin, D. Pennica, D. V. Goeddel, and P. W. Gray.** 1985. Human lymphotoxin and tumor necrosis factor genes: structure, homology and chromosomal localization. Nucleic Acids Res. **13**:6361–6373.
40. **Oi, V. T., S. L. Morrison, L. A. Herzenberg, and P. Berg.** 1983. Immunoglobulin gene expression in transformed lymphoid cells. Proc. Natl. Acad. Sci. USA **80**:825–829.
41. **Pedersen, F. S., R. L. Crowther, D. Y. Tenney, A. M. Reimhold, and W. A. Haseltine.** 1981. Novel leukaemogenic retroviruses isolated from cell line derived from spontaneous AKR tumour. Nature (London) **292**:167–170.
42. **Riedel, N., E. A. Hoover, P. W. Gasper, M. O. Nicolson, and J. I. Mullins.** 1986. Molecular analysis and pathogenesis of the feline aplastic anemia retrovirus, feline leukemia virus C-SARMA. J. Virol. **60**:242–250.
43. **Rosen, C., W. Haseltine, J. Lenz, R. Ruprecht, and M. W. Cloyd.** 1985. Tissue selectivity of murine leukemia virus infection is determined by long terminal repeat sequences. J. Virol. **55**:862–866.
44. **Sassone-Corsi, P., and E. Borrelli.** 1986. Transcriptional regulation by *trans*-acting factors. Trends Genet. **2**:215–219.
45. **Shinnick, T. M., R. A. Lerner, and G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
46. **Speck, N., B. Renjifo, E. Golemis, T. N. Fredrickson, J. W. Hartley, and H. Hopkins.** 1990. Mutation of the core or adjacent LVb elements of the Moloney virus enhancer alters disease specificity. Genes Dev. **4**:233–242.
47. **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.
48. **Trainor, C. D., J. L. Scott, S. F. Josephs, K. E. Fry, and M. S. Reitz, Jr.** 1984. Nucleotide sequence of the large terminal repeat of two different strains of gibbon ape leukemia virus. Virology **137**:201–205.
49. **van Beveren, C., F. van Straaten, J. A. Gallenshaw, and I. M. Verma.** 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell **27**:97–108.
50. **Vogt, M., C. Haggblom, S. Swift, and M. Haas.** 1985. Envelope gene and long terminal repeat determine the different biological properties of Rauscher, Friend, and Moloney mink cell focus-inducing viruses. J. Virol. **55**:184–192.
51. **Wasylyk, B., J. L. Imler, B. Chatton, C. Schatz, and C. Wasylyk.** 1988. Negative and positive factors determine the activity of the polyoma virus enhancer α domain in undifferentiated and differentiated cell types. Proc. Natl. Acad. Sci. USA **85**:7952–7956.
52. **Xiao, J. H., I. Davidson, D. Ferrandon, R. Rosales, M. Vigneron, M. Macchi, F. Ruffenach, and P. Chambon.** 1987. One cell-specific and three ubiquitous nuclear proteins bind *in vitro* to overlapping motifs in the domain B1 of the SV40 enhancer. EMBO J. **6**:3005–3013.
53. **Xiao, J. H., I. Davidson, M. Macchi, R. Rosales, M. Vigneron, A. Staub, and P. Chambon.** 1987. *In vitro* binding of several cell-specific and ubiquitous nuclear proteins to the GT-1 motif of the SV40 enhancer. Genes Dev. **1**:794–807.
54. **Yasukawa, K., T. Hirano, Y. Watanabe, K. Muratani, T. Matsuda, S. Nakai, and T. Kishimoto.** 1987. Structure and expression of human B cell stimulatory factor-2 (BSF-2/IL-6) gene. EMBO J. **6**:2939–2945.