JOURNAL OF VIROLOGY, Apr. 1994, p. 2632–2648
0022-538X/94/$04.00+0

Vol. 68, No. 4

A Transcriptional Regulatory Element Is Associated with a Nuclease-Hypersensitive Site in the *pol* Gene of Human Immunodeficiency Virus Type 1

CARINE VAN LINT,[1] JACQUES GHYSDAEL,[2] PETER PARAS, JR.,[1] ARSÈNE BURNY,[3] AND ERIC VERDIN[1]*

Laboratory of Viral and Molecular Pathogenesis, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892[1]; Section de Biologie, Institut Curie, 91405 Orsay Cédex, France[2]; and Département de Biologie Moléculaire, Université Libre de Bruxelles, 1640 Rhode-St-Genèse, Belgium[3]

Received 15 November 1993/Accepted 7 January 1994

Analysis of the chromatin organization of the integrated human immunodeficiency virus type 1 (HIV-1) genome has previously revealed a major constitutive DNase I-hypersensitive site associated with the *pol* gene (E. Verdin, J. Virol. 65:6790–6799, 1991). In the present report, high-resolution mapping of this site with DNase I and micrococcal nuclease identified a nucleosome-free region centered around nucleotides (nt) 4490 to 4766. A 500-bp fragment encompassing this hypersensitive site (nt 4481 to 4982) exhibited transcription-enhancing activity (two- to threefold) when it was cloned in its natural position with respect to the HIV-1 promoter after transient transfection in U937 and CEM cells. Using in vitro footprinting and gel shift assays, we have identified four distinct binding sites for nuclear proteins within this positive regulatory element. Site B (nt 4519 to 4545) specifically bound four distinct nuclear protein complexes: a ubiquitous factor, a T-cell-specific factor, a B-cell-specific factor, and the monocyte/macrophage- and B-cell-specific transcription factor PU.1/Spi-1. In most HIV-1 isolates in which this PU box was not conserved, it was replaced by a binding site for the related factor Ets1. Factors binding to site C (nt 4681 to 4701) had a DNA-binding specificity similar to that of factors binding to site B, except for PU.1/Spi-1. A GC box containing a binding site for Sp1 was identified (nt 4623 to 4631). Site D (nt 4816 to 4851) specifically bound a ubiquitously expressed factor. These results identify a transcriptional regulatory element associated with a nuclease-hypersensitive site in the *pol* gene of HIV-1 and suggest that its activity may be controlled by a complex interplay of *cis*-regulatory elements.

Human immunodeficiency virus type 1 (HIV-1) is a lentivirus that is etiologically associated with AIDS, a slowly progressive degenerative disease affecting the immune and central nervous systems in humans (for reviews, see references 18, 24, 35, 52, and 79). Regulation of HIV-1 gene expression is controlled by the combined action of *trans*-acting cellular and viral regulatory factors (for a review, see references 18, 31, 35, 40, 64, and 79). The protein factors exert their regulatory functions by interacting with viral *cis*-acting elements, at both the DNA and the RNA levels. At the transcriptional level, the long terminal repeat (LTR), which is present at both extremities of the integrated viral genome, contains *cis*-acting elements necessary for transcription initiation (5′-LTR) and for polyadenylation of the viral transcripts (3′-LTR). Several inducible or constitutive transcription factors bind in vitro to each of the four functionally defined regions of the 5′-LTR: first, the negative regulatory element, a silencer, containing binding sites for the cellular factors AP-1, COUP, NF-AT1, USF, glucocorticoid receptor, and NF-IL6; second, the TCF-1α site and the enhancer region, which interacts with NF-κB, HIVEN86A, and EBP-1; third, the basal promoter containing three Sp1-binding sites, a TATA box, and an initiator element close to the transcription initiation site; and fourth, the 5′ untranslated leader region containing binding sites for UBP-1/LBP-1, UBP-2, and CTF/NF1 (for reviews, see references 18, 31, 35, 40). Furthermore, the last region corresponds to the *trans*-activation response (TAR) element whose RNA forms a stable stem-loop structure interacting with the viral transactivator Tat in a manner critical for high activity of the HIV-1 promoter (for reviews, see references 17, 29, 43, and 69).

In addition to these elements, a 12-*O*-tetradecanoylphorbol-13-acetate (TPA)-inducible intragenic enhancer was identified in the *pol* gene of HIV-1 (83). This element is composed of two functional subdomains encompassing nucleotides (nt) 4079 to 4342 and nt 4781 to 6026, both exhibiting TPA-inducible enhancing activity, on the herpes simplex virus thymidine kinase promoter in HeLa cells (83). However, when these elements were tested in T-lymphoid and monocyte/macrophage cell lines, they exhibited no significant activity in the context of either the heterologous thymidine kinase promoter or the homologous HIV-1 promoter (8). Nevertheless, analysis of the chromatin organization of integrated HIV-1 identified a single major DNase I-hypersensitive site in the 8-kb region located between the two LTRs (82). This site maps to the *pol* gene precisely between the two functional domains of the intragenic enhancer identified in HeLa cells (82). This constitutive hypersensitive site is present only in a cell line of monocytic origin and not in two cell lines of lymphoid origin, suggesting a cellular specificity associated with this intragenic element (82).

Since DNase I-hypersensitive sites are frequently associated with regulatory regions of genes such as promoters and enhancers (36), the presence of this open chromatin configuration in vivo prompted us to examine the potential regulatory role of the underlying DNA element in HIV-1 transcription. In this report, we have characterized the intragenic hypersensitive site and found that a region of the *pol* gene centered around nt

* Corresponding author. Mailing address: The Picower Institute for Medical Research, 350 Community Dr., Manhasset, NY 11030. Phone: (516) 562-9495. Fax: (516) 365-5090.

4490 to 4766 is nucleosome free in vivo. When cloned downstream and in the sense orientation relative to the HIV-1 promoter, the *pol* gene-hypersensitive site region reproducibly increases transcription mediated by the HIV-1 5′-LTR in U937 and CEM cells. This positive regulatory element contains several recognition sites for ubiquitous and cell-type-specific DNA-binding factors, suggesting their involvement in the control of the transcriptional activity of the region.

MATERIALS AND METHODS

Cell culture. Human cell lines were obtained from the AIDS Research and Reference Reagent Program (National Institute of Allergy and Infectious Diseases [NIAID], National Institutes of Health, Bethesda, Md.). All cell lines, except HeLaS3 and HeLa, were grown in RPMI 1640 medium (GIBCO-BRL) supplemented with 10% fetal bovine serum (HyClone), 50 U of penicillin and 50 μg of streptomycin per ml, and 2 mM glutamine at 37°C in a 95% air–5% CO_2 atmosphere and were maintained at a density of 0.25×10^6 to 1×10^6 cells per ml (exponential growth phase). HeLa cells were cultured in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum and HeLaS3 cells in minimal essential medium supplemented with 5% horse serum. When cells were cultured in spinner flasks, medium was supplemented with 15 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; pH 7.3). When indicated, cells were treated for 12 h with 10 nM TPA dissolved in dimethyl sulfoxide (final concentration, 0.01%) or with dimethyl sulfoxide alone (final concentration, 0.01%) as a control.

Indirect end-labeling technique. The indirect end-labeling technique was performed essentially as previously described by our laboratory (82, 84). Briefly, nuclei were purified from exponentially growing cells. DNase I and micrococcal nuclease digestions were performed in situ on intact nuclei, and DNA was extracted and purified. Control genomic DNA was purified, and nuclease treatment of naked DNA was performed as previously described (82, 84). Southern blotting was carried out as described elsewhere (82, 84), with the following modifications. Purified DNA (30 μg) was digested with *Eco*RI, and the fragments generated were separated by electrophoresis in 1.5% agarose gels in 1× TBE buffer (90 mM Tris-HCl, 90 mM boric acid, 2 mM EDTA [pH 8.3]) at 1.5 V/cm. Size markers were electrophoresed along with the samples. Each size marker was generated by digesting $HIV\text{-}1_{NL4\text{-}3}$ DNA (cloned in plasmid pNL4-3 [1]) with *Eco*RI and another enzyme chosen to generate a fragment of defined size and location in the region under study. Several of these markers were mixed together, added to 30 μg of cellular DNA, and coelectrophoresed with the samples. After electrophoresis, gels were incubated in denaturing solution and then in neutralizing solution and were transferred by capillarity to a nylon membrane. DNA was cross-linked to nylon membranes by exposure to UV light. The membranes were washed and prehybridized. A probe spanning nt 5063 to 5743 (coordinates with respect to the NY5 provirus, where nt +1 is the first nucleotide in the 5′-LTR U3 region) was synthesized by 15 cycles (95°C for 2 min, 55°C for 2 min, and 72°C for 3 min) of PCR using 1 ng of a plasmid containing a complete molecular clone of HIV-1 (pBRU2) as a template and primers EV173 and EV174. The sequences of the primers were as follows: EV173, 5′ (nt 5743)-CTTATTA TGGCTTCCACTCCTGC-3′; and EV174, 5′ (nt 5063)-TGAT TGTGTGGCAAGTAGACAGG-3′. The amplified DNA fragment was labeled by the random primer reaction (25) and was purified on a Sephadex G-50 column. The denatured DNA probe was added to the prehybridization buffer and allowed to hybridize for at least 16 h at 42°C. After hybridization, the membranes were washed and exposed for autoradiography.

Plasmid constructs. Recombinant DNA techniques were performed according to the methods described by Sambrook et al. (71) and Ausubel et al. (6). To create plasmids used in DNase I footprinting analysis, five overlapping fragments (nt 4346 to 4521, nt 4391 to 4564, nt 4491 to 4668, nt 4601 to 4782, and nt 4731 to 4962 [coordinates with respect to the NY5 provirus]) were generated by PCR amplification with the $HIV\text{-}1_{HXB2}$ genome as a template and were directly cloned in the vector pCR1000 according to the specifications of the manufacturer (TA cloning kit; Invitrogen Corporation). Recombinants with fragments inserted in the antisense orientation were selected and designated pHXB2(4346–4521), pHXB2(4391–4564), pHXB2(4491–4668), pHXB2(4601–4782), and pHXB2(4731–4962), respectively.

A 793-bp fragment containing the $HIV\text{-}1_{LAI}$ 5′-LTR (nt 1 to 792) was generated by PCR amplification with pBRU2 as a template. The amplified fragment was digested by *Pst*I (site added in the 5′ primer) and *Xba*I (site added in the 3′ primer) after gel purification and was ligated to the *Pst*I-*Xba*I-restricted vector pCATBASIC (Promega). The resulting plasmid was designated pLTR.

PCR was used to amplify the *pol* gene fragment (nt 4481 to 4982) corresponding to the nuclease-hypersensitive region from the provirus integrated in U1 cells. *Bam*HI sites were introduced in the PCR primers, and the *Bam*HI-restricted PCR fragment was cloned in the unique *Bam*HI site of pLTR, placing the fragment downstream of the 5′-LTR-*cat* transcriptional unit. The 5′ primer oligonucleotide encompassed the coding strand sequence from nt 4481 to 4505 and contained added overlapping *Bam*HI and *Xho*I restriction sites (underlined and designed in boldface, respectively) at the 5′ end (5′-CG**GGATCCTCGA**[nt 4481]**G**AAGCAGAAGTAATTC CAGCAGAG-3′). The 3′ primer oligonucleotide encompassed the complementary sequence of the *pol* gene from nt 4957 to 4982 and contained an added *Bam*HI site (underlined) at the 5′ end (5′-CGGGATCCT[nt 4982]ATTACTACTGC CCCTTCACCTTTCC-3′). Amplification reactions were conducted with 1 μg of U1 genomic DNA according to the protocol provided with the AmpliTaq DNA polymerase (Perkin-Elmer Cetus), using a Perkin-Elmer Cetus thermal cycler. Recombinant plasmids containing the U1 fragment (nt 4481 to 4982) inserted in the sense (S) and antisense (AS) orientations with respect to the transcriptional unit were designated pLTR-$U1_{Lg}$(S) and pLTR-$U1_{Lg}$(AS), respectively. To create pLTR-$U1_{Sm}$(S) and pLTR-$U1_{Sm}$(AS), which contain a smaller *pol* gene fragment from the U1 provirus (nt 4481 to 4777) in the S and AS orientations, respectively, the U1 fragment containing this region was amplified by PCR from U1 genomic DNA. The 5′ primer was as described above. The 3′ primer oligonucleotide encompassed the complementary sequence of the *pol* gene from nt 4752 to 4777 and contained added overlapping *Bam*HI and *Nde*I restriction sites (underlined and designed in boldface, respectively) at the 5′ end (5′-CG**GGATCCATAT**[nt 4777]GGATGAATACTGCCATTTGTACTGC-3′).

To test the potential promoter activity of the U1 *pol* gene-hypersensitive region, the large fragment extending from nt 4481 to 4982 was also cloned in pCATBASIC. The 5′ primer was as described above, and the 3′ primer encompassed the complementary sequence of the *pol* gene from nt 4957 to 4982 with an additional *Xba*I restriction site (underlined) at the 5′ end (5′-GCTCTAGAT[nt 4982]ATTACTACTGCCCCT TCACCTTTCC-3′). Amplification reactions were carried out

as described above. The amplified fragment was digested with *Xho*I and *Xba*I and then ligated into the *Sal*I-*Xba*I sites of pCATBASIC to generate pIP.

DNA sequencing. All PCR-amplified fragments were sequenced on both strands by the dideoxynucleotide termination sequencing method (72) with Sequenase version 2.0 (U.S. Biochemicals). The nucleotide sequence of the nuclease-hypersensitive region in U1 cells was identical to that of the homologous region in the HIV-1 isolate NY5. For this reason, the base pair coordinates of the U1 intragenic region were numbered according to the NY5 sequence (where nt +1 is the start of U3 in the 5′-LTR).

Transfection and CAT assays. Plasmids were grown in *Escherichia coli* HB101 and purified by the QUIAGEN procedure (QUIAGEN, Inc., Chatsworth, Calif.) and then by two phenol-chloroform extractions and ethanol precipitation. Transient transfection assays were carried out by using the DEAE-dextran procedure (6) with a Rous sarcoma virus (RSV) LTR-luciferase construct (a gift from L. Hudson) as an internal control for transfection efficiency in early experiments. Five micrograms of control plasmid (pLTR or pCATBASIC) or equimolar amounts of the different constructs were used to transfect 10^7 cells. When indicated, 0.5 μg of Tat expression vector pSV_2tat (AIDS Research Reagent Program, NIAID) containing the *tat* gene under the control of the simian virus 40 (SV40) early promoter region was cotransfected with the plasmids. At 42 to 48 h after transfection, the cells were harvested, washed once with phosphate-buffered saline, and resuspended in 150 μl of 0.25 M Tris-HCl (pH 7.8). Cell extracts were prepared by three cycles of freezing-thawing (−70°C→37°C) and then by incubation at 67°C for 7 min. Protein concentrations were determined with the Bio-Rad protein determination kit and gamma globulin as a standard. CAT (chloramphenicol acetyltransferase) enzyme activities were determined by the diffusion CAT assay method (59) with equivalent amounts of protein extract (1 to 5 μg, depending on the cell line used) in 250 μl of 100 mM Tris-HCl (pH 7.8) containing 0.5 μCi of [^{3}H]acetyl coenzyme A (200 mCi/mmol; Dupont-New England Nuclear) and 1 mM chloramphenicol. Each sample was placed under 5 ml of Econofluor (Dupont-New England Nuclear) at room temperature, and scintillation counting was performed every hour for 4 h. Six to eight independent transfections were performed with two separately prepared plasmid stocks.

DNA-protein interaction analyses. (i) Electrophoretic mobility shift assays (EMSAs). Nuclear extracts were prepared according to the method described by Dignam et al. (19) with modifications (2). To minimize proteolysis, all buffers included the following protease inhibitors (Boehringer Mannheim): 10 μg of antipain, 2 μg of aprotinin, 10 μg of chymostatin, 1 μg of leupeptin, and 1 μg of pepstatin per ml. Protein concentrations were determined by the Bradford method (10) by using the Bio-Rad protein assay with bovine plasma gamma globulin as the standard.

Oligonucleotides were synthesized on an Applied Biosystems A380 synthesizer by the phosphoroamidate method and were purified on denaturing polyacrylamide gels and Sep-Pak cartridges for solid-phase extraction (Waters, Millipore, Mass.) (71). The concentration of single-stranded oligonucleotides was determined according to the method described by Sambrook et al. (71) after measurement of the A_{260}. The DNA sequences of the coding strands of oligonucleotides used in this study are listed in Table 1.

EMSAs were performed with modifications of previously described procedures (2, 81). Double-stranded oligonucleotides were 5′ end labeled with [γ-^{32}P]ATP (>5,000 Ci/mmol; Amersham) and T4 polynucleotide kinase, purified after isolation from polyacrylamide gels, and used as probes. The typical protein-DNA binding reactions were carried out as follows. Nuclear extract (10 to 15 μg of protein) was first incubated on ice for 10 min in the absence of probe and specific competitor DNA in a 16-μl reaction mixture containing 10 μg of DNase-free bovine serum albumin (Pharmacia), 1 to 2 μg of poly(dI-dC) (Pharmacia) as nonspecific competitor DNA, 50 μM $ZnCl_2$, 0.25 mM dithiothreitol, 20 mM HEPES (pH 7.3), 60 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA, and 10% (vol/vol) glycerol. A total of 10,000 to 20,000 cpm of probe (10 to 40 fmol) was then added to the mixture with or without a molar excess of an unlabeled competitor, and the mixture was incubated for 15 min on ice. The reaction mixture was loaded directly onto a 6% nondenaturing polyacrylamide gel which had been pre-electrophoresed for 1 h at 150 V. Samples were electrophoresed at room temperature at 150 V for 2 to 3 h in 1× TGE buffer (25 mM Tris-acetate [pH 8.3], 190 mM glycine, 1 mM EDTA). Gels were dried, and autoradiographic exposures were carried out with Kodak XAR-Omat films for 16 to 24 h at −70°C with two intensifying screens. For each oligonucleotide probe tested, a range of protein and poly(dI-dC) concentrations was used in preliminary experiments.

For analysis of Ets1 binding, whole-cell extracts of *Spodoptera frugiperda* SF9 cells infected with a recombinant Ac-NPV-Ets1 baculovirus (9) were used as a source for Ets1 protein, and extracts from uninfected insect cells were used as a control. Conditions for EMSA were as previously described (9).

For analysis of PU.1 binding, whole-cell lysates of COS1 cells transfected with ΔEB-PU.1, an SV40 promoter-based expression vector for human PU.1/Spi-1 (68), were utilized. At 48 h after transfection, pelleted cells were lysed on ice for 15 min in 3 volumes of a solution containing 10 mM HEPES (pH 7.9), 0.3 M NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, 0.5% Triton X-100, 10 μg of leupeptin per ml, 1% aprotinin (Sigma), and 100 μg of phenylmethylsulfonyl fluoride per ml and centrifuged for 15 min at 12,000 × *g*. An aliquot of the supernatant fraction (approximately 2 μg of total protein) was used in EMSAs as previously described (68). Extracts of cells transfected with the empty ΔEB expression vector were used as a control.

(ii) DNase I footprinting assays. End-labeled DNA probes were prepared either by filling in with the Klenow fragment of DNA polymerase I or by 5′-end phosphorylation with T4 polynucleotide kinase after dephosphorylation with calf intestinal alkaline phosphatase. Probes were purified by polyacrylamide gel electrophoresis, elution, and passing through a NACS 52 column as previously described (2). DNase I footprinting reactions were performed as described by Dynan (21), with the following modifications. A typical reaction mixture contained 20,000 cpm of probe, 50 μg of crude nuclear extract, and 4 μg of poly(dI-dC) (Pharmacia) in a total volume of 50 μl. The final ionic composition was 10 mM HEPES (pH 7.9), 50 mM KCl, 0.5 mM dithiothreitol, 50 μM $ZnCl_2$, 0.1 mM EDTA, and 10% (vol/vol) glycerol. This mixture was incubated on ice for 15 min. After binding, an equal volume of a solution containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$ was added to the reaction mixture, and the mixture was incubated for 1 min at room temperature. A freshly diluted solution of DNase I (Sigma Chemical Co.) was added at a final concentration of 0.1 to 20 μg/ml. DNase I digestions were performed for 2 min at room temperature and were stopped by the addition of 90 μl of a solution containing 0.2 M NaCl, 0.03 M EDTA, 1% sodium

TABLE 1. Oligonucleotides used in EMSA experiments

Oligonucleotide	Sequence	Source (genomic position)[a]	Source or reference
AP-1	5′-TGTTATAAAGCATGAGTCAGACACCTCTGGCTTTCTG-3′	Collagenase gene enhancer (nt −84 to −48)	3
NF-κB	5′-AGTTGAGGGGACTTTCCCAGGC-3′	SV40 enhancer	49
TCF-1	5′-GACTGAGAACAAAGCGCTCTCACAC-3′	CD3ε enhancer (nt 74 to 96)	80
TCF-1α	5′-GATCTAGGGCACCCTTTGAAGCTCT-3′	T-cell receptor α enhancer (nt 57 to 81)	87
Ets1 (HXB2)	5′-AACAGGGCAGGAAACAGCATATT-3′	HIV-1_{HXB2} (nt 4505 to 4527)	This study
Ets1 (U1)	5′-GACAGGGCAAGAAACAGCATACT-3′	HIV-1_{U1} (nt 4505 to 4527)	This study
ETS1	5′-ATAAACAGGAAGTGGT-3′		32
ETS1mut	5′-ATAAACACCAAGTGGT-3′		32
SV40PU.1	5′-GATCTGAAATAACCTCTGAAAGAGGAACTTGGTTAGGTA-3′	Variant enhancer-minus of SV40	45
SV40PU.1mut	5′-GATCTGAAATAACCTCTGAAAGACCAACTTGGTTAGGTA-3′	Variant enhancer-minus of SV40	45
U1 site B short	5′-ATACTTCCTCTTAAAATTAG-3′	HIV-1_{U1} (nt 4523 to 4542)	This study
U1 site B	5′-CAGCATACTTCCTCTTAAAATTAGCAG-3′	HIV-1_{U1} (nt 4519 to 4545)	This study
SF2 site B	5′-CAGCATATTTTCTCTTAAAATTAGCAG-3′	HIV-1_{SF2} (nt 4519 to 4545)	This study
HXB2 site B	5′-CAGCATATTTTCTTTTAAAATTAGCAG-3′	HIV-1_{HXB2} (nt 4519 to 4545)	This study
LAI site B	5′-CAGCATACTTTCTTTTAAAATTAGCAG-3′	HIV-1_{LAI} (nt 4519 to 4545)	This study
Site D	5′-CAGGGGAAAGAATAGTAGACATAATAGCAACAGACA-3′	HIV-1_{U1} (nt 4816 to 4851)	This study
Site C	5′-TAGAATCTATGAATAAAGAAT-3′	HIV-1_{U1} (nt 4681 to 4701)	This study

[a] Genomic positions of all HIV-1 sequences are given according to the numbering of the HIV-1_{NY5} genome (where nt +1 is the start of U3 in the 5′-LTR) independently of the considered isolate.

dodecyl sulfate, and 100 μg of yeast tRNA per ml. Proteinase K was added to a final concentration of 200 μg/ml at 56°C for 1 h. The DNA was extracted with phenol-chloroform, precipitated with ethanol, and analyzed by electrophoresis on 6 to 8% polyacrylamide sequencing gels. Binding reactions with purified Sp1 were performed in the absence of nonspecific competitor DNA according to the instructions provided by the manufacturer (Promega). Maxam and Gilbert purine sequencing reaction (51) products of the probes were used as size markers.

(iii) Methylation interference assays. Double-stranded oligonucleotides end labeled on one strand only were partially methylated with dimethyl sulfate by standard procedures (51) and were incubated with nuclear extract. DNA-protein complexes were separated from the free probe by EMSAs. The entire gel was electrophoretically transferred to a DEAE membrane (NA45; Schleicher & Schuell) at 28 V (0.8 A) for 30 min. After electrophoretic transfer, the DEAE membrane was exposed for autoradiography, and the relevant bands were cut. DNA was eluted at 65°C for 1 h in 200 μl of elution buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA [pH 8.0], 1 M NaCl). DNA was purified by phenol-chloroform extraction and ethanol precipitation, resuspended in water, and cleaved at the methylated purines with 1 M piperidine for 30 min at 90°C. Cleavage products were separated on a 15% sequencing gel.

RESULTS

High-resolution mapping of the hypersensitive site located in the *pol* gene of HIV-1. Previous study of HIV-1 proviral chromatin organization in chronically infected cell lines at low resolution has identified a major constitutive DNase I-hypersensitive site in the *pol* gene of the U1 cell line but not in the ACH2 and 8E5 cell lines (82). The fact that this intragenic hypersensitive site was observed only in cells of monocytic origin (U1) (27) and not in two cell lines of lymphoid origin (ACH2 and 8E5) (15, 28) suggested that a cellular specificity was associated with this element. To map the intragenic DNase I-hypersensitive site at higher resolution, this experiment was repeated with a new probe that was located more closely to the hypersensitive site. Purified nuclei from TPA-treated and untreated U1 cells were digested in vitro with DNase I or micrococcal nuclease (Fig. 1A or B, respectively), and digestion products were analyzed by Southern blotting with indirect end labeling (82, 84, 92). U1 cells were examined both with and without TPA treatment, since this agent was previously shown to induce viral expression in these cells (27). The efficiency of viral induction following TPA treatment was monitored by measuring the secreted p24 antigen in the culture medium. DNase I digestion of U1 nuclei resulted in the appearance of two hypersensitive sites (one major and one minor, indicated in Fig. 1A by arrows I and II, respectively) mapping to the *pol* gene. The exact positions of these hypersensitive sites in the viral genome were determined by using markers containing double digestions of HIV-1_{NL4-3} DNA by *Eco*RI and several restriction enzymes throughout the region. Site I mapped to nt 4450 to 4566, and site II, which was more clearly detectable after a longer exposure of the same gel (data not shown), mapped to nt 4701 to 4773 (values averaged from two independent experiments). To prove that the hypersensitive sites observed were a consequence of chromatin organization and not secondary to sequence-directed cleavage preference by DNase I, the pattern of digestion of U1 nuclei by DNase I was compared with that of U1 naked DNA digested in vitro. The range of DNase I concentrations was chosen to generate similar levels of digestion under in vitro and in vivo conditions. No sequence preference for cutting by DNase I was observed in vitro (Fig. 1A, lanes 9 to 10). As previously described (82), TPA induction resulted in a threefold reduction in the intensity of these sites (by densitometry scanning) (Fig. 1A, lanes 5 to 8).

To confirm these results with another nuclease, we used micrococcal nuclease. This enzyme preferentially cuts DNA present in nucleosome-free regions in chromatin, as well as DNA present in linker domains between adjacent nucleosomes. Although in vitro micrococcal nuclease exhibited sequence preference for cutting naked DNA (Fig. 1B, lanes 9 to 10), two hypersensitive sites mapping to the same region as the DNase I-hypersensitive sites were noted only in vivo (indicated by arrows in Fig. 1B), supporting the hypothesis that they were secondary to chromatin organization. Site I mapped to nt 4490 to 4529, and site II mapped to nt 4585 to 4621. Similar results were obtained after TPA treatment (Fig. 1B, lanes 5 to 8).

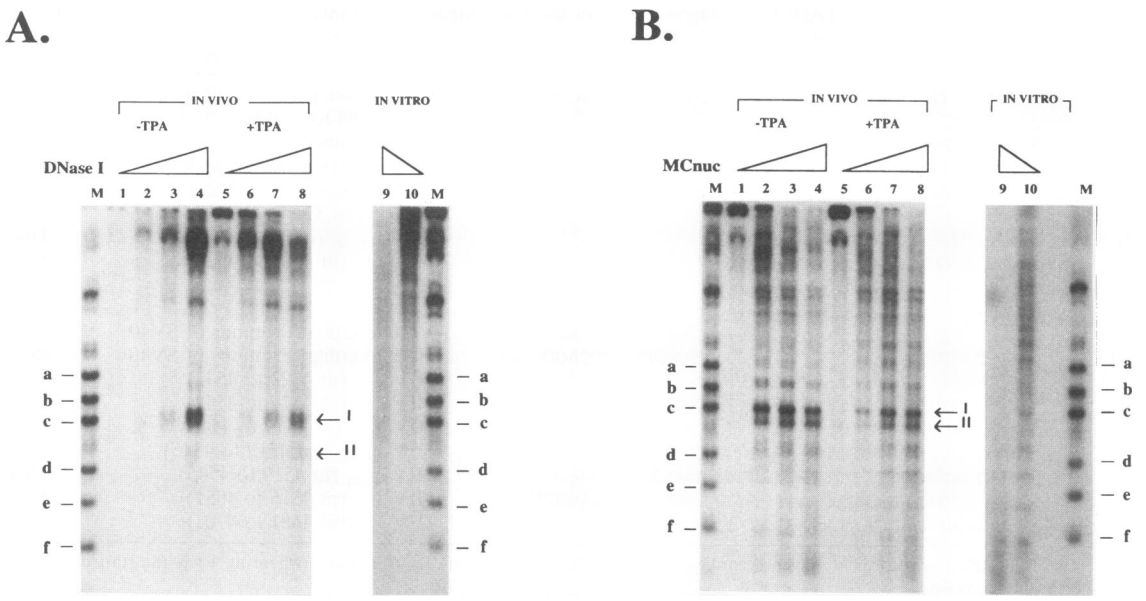

FIG. 1. High-resolution analysis of a nuclease-hypersensitive site in the *pol* gene of HIV-1. Nuclei from U1 cells that were untreated or treated with TPA (10 nM) for 12 h were incubated with either DNase I on ice for 10 min (A) or micrococcal nuclease (MCnuc) at room temperature for 20 min (B). DNAs were then purified, digested in vitro with *Eco*RI, and analyzed by indirect end labeling with a probe (nt 5063 to 5743) abutting the unique restriction site *Eco*RI (nt 5743). (A) The following doses of DNase I were used in vivo: 0 (lanes 1 and 5), 10 (lanes 2 and 6), 15 (lanes 3 and 7), and 20 (lanes 4 and 8) U/ml. As a control, naked U1 genomic DNA was digested in vitro with DNase I (0.5 and 0.25 U/ml [lanes 9 and 10, respectively]). Molecular weight markers are double digests of HIV-1_{NL4-3} DNA with *Eco*RI (nt 5743) and *Kpn*I (nt 4158 [marker a]), *Pvu*II (nt 4345 [marker b]), *Xmn*I (nt 4492 [marker c]), *Dra*I (nt 4784 [marker d]), *Ava*II (nt 4938 [marker e]), and *Nde*I (nt 5123 [marker f]). Hypersensitive sites are indicated by arrows and roman numerals. (B) The following doses of micrococcal nuclease were used: 0 (lanes 1 and 5), 0.015 (lanes 2 and 6), 0.03 (lanes 3 and 7), and 0.06 (lanes 4 and 8) U/ml. As a control, naked U1 genomic DNA was digested in vitro with micrococcal nuclease (0.04 and 0.02 U/ml [lanes 9 and 10, respectively]). Hypersensitive sites identified in vivo are indicated by arrows.

High-resolution mapping of the intragenic hypersensitive site in U1 cells identified two hypersensitive sites after both DNase I and micrococcal nuclease digestion. Since nuclease-hypersensitive sites usually indicate the presence of nonnucleosomal DNA in vivo, we conclude that the region of the *pol* gene centered around nt 4490 to 4766 is nucleosome free in vivo.

The intragenic hypersensitive site region increases 5′-LTR-directed gene expression in U937 and CEM cells in transient transfection assays. Nuclease-hypersensitive sites have been found in association with regulatory regions of genes such as enhancers, promoters, upstream activating sequences, silencers, terminators, recombination loci, telomeres, and centromeres (36). To address the potential role of the *pol*-hypersensitive region, we tested its potential enhancer activity on the transcription rate of the HIV-1 promoter. A region extending from nt 4481 to 4982, encompassing the nuclease-hypersensitive region, was amplified by PCR from U1 genomic DNA and sequenced. Its nucleotide sequence was identical to that of the homologous region in the HIV-1 isolate NY5. The fragment was subcloned into the construct pLTR in the S and AS orientations with respect to the transcriptional unit, downstream of the *cat* reporter gene [pLTR-$U1_{Lg}$(S) and pLTR-$U1_{Lg}$(AS), respectively]. The pLTR construct contains the complete 5′-LTR (plus the leader sequence up to the ATG of *gag*) driving the expression of the *cat* gene. Constructs were transiently transfected into the $CD4^+$ T-lymphoid cell line CEM and the promonocytic cell line U937 with or without the Tat expression vector pSV_2TAT. In preliminary experiments, transfections were normalized for the levels of luciferase activity by including as an internal control an RSV LTR-luciferase construct in each transfection mixture. Since no significant variations were noted in transfection efficiencies between samples from a single experiment, the RSV LTR-luciferase construct was not further utilized. In the absence of Tat, low CAT activity was detected with all constructs in the two cell lines tested (data not shown). In the presence of Tat, a 46-fold increase in CAT activity was observed with pLTR compared with the absence of Tat. All values were normalized to the amount of CAT activity of the control pLTR in the presence of Tat to obtain relative CAT activity. In U937 cells, transfection of plasmid pLTR-$U1_{Lg}$(S) caused a 2.4-fold increase in CAT activity compared with that of the control plasmid pLTR tested under the same conditions (Fig. 2). This effect was strongly orientation dependent, since the same fragment cloned in the AS orientation exhibited no enhancing activity [pLTR-$U1_{Lg}$(AS); Fig. 2]. Similar results were obtained for CEM cells [2.4- and 1.5-fold for pLTR-$U1_{Lg}$(S) and pLTR-$U1_{Lg}$(AS), respectively; Fig. 2]. These results indicate that the region associated with the intragenic hypersensitive site functions as a weak T-lymphoid and monocytic transcriptional enhancer when it is cloned in a position similar to that observed within the viral genome (downstream and in the S orientation relative to the 5′-LTR-*cat* transcriptional unit).

To define further the region contributing to transcriptional enhancing activity, a smaller fragment (nt 4481 to 4777) was also tested after cloning into pLTR in the S and AS orientations [pLTR-$U1_{Sm}$(S) and pLTR-$U1_{Sm}$(AS), respectively]. As shown in Fig. 2, the smaller fragment exhibited enhancing activity similar to that of the larger fragment in both U937 cells and CEM cells. This enhancing activity was strongly dependent on the orientation of the fragment with respect to the promoter, as described for the larger fragment (Fig. 2).

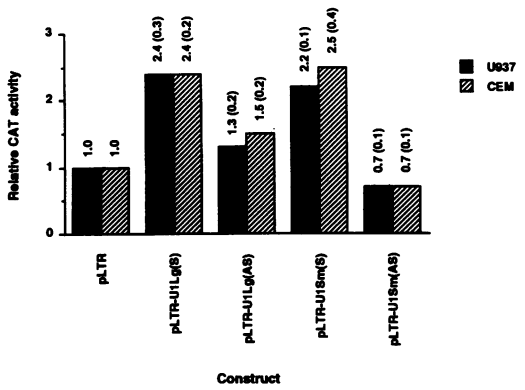

FIG. 2. Transcriptional enhancing activity of the intragenic hypersensitive-site region. The enhancer activity of the region associated with the *pol* gene-hypersensitive site was tested after cloning in pLTR and transfection of the plasmids in U937 and CEM cells with the Tat expression vector pSV_2TAT. CAT assays were performed as described in Materials and Methods. The results are presented as histograms indicating CAT activities relative to that of vector pLTR, which was assigned a value of 1. Means and standard errors of the means from 6 to 8 independent transfections performed with two different DNA preparations are indicated.

We also tested the ability of the nuclease-hypersensitive region to function as a promoter. The large fragment was cloned upstream of the *cat* gene in pCATBASIC. Very low CAT activity was measured after transfection of the resulting plasmid pIP in CEM and U937 cells (data not shown). We also considered the possibility that the promoter activity is inducible or dependent on the expression of a viral protein(s). The same construct pIP was cotransfected in CEM and U937 cells, along with a Tat expression vector or the plasmid pNL4-3 (1) containing the complete HIV-1 genome in the presence or the absence of TPA. None of these conditions significantly increased CAT activity in any of the cell lines tested (data not shown). In additional experiments, we cotransfected the construct pLTR along with the plasmid pBRU2, which was digested by *Pst*I (cutting at nt 1419) and *Sac*I (cutting at nt 492 and 682), thereby preventing elongation of the transcripts initiated in the 5′-LTR. Since the level of expression of the reporter *cat* gene present in pLTR is dependent on the transactivation of the HIV-1 promoter by the viral protein Tat, the presence in the digested pBRU2 plasmid of an intragenic promoter able to produce Tat should be revealed by an increase in the CAT expression of the reporter system pLTR. No transactivation of the HIV-1 promoter present in pLTR was detected. We conclude from these experiments that the intragenic nuclease-hypersensitive region does not exhibit any promoter activity.

In summary, when cloned downstream and in the sense orientation with respect to the HIV-1 promoter, a *pol* region associated with a nuclease-hypersensitive domain in U1 cells was found to reproducibly increase (two- or threefold) transcription mediated by the viral 5′-LTR in U937 and CEM cells in the presence of Tat.

Identification of sites of interaction with nuclear proteins within the hypersensitive site region by DNase I footprinting analysis. The transcriptional control of gene expression is the result of interactions between *cis*-acting DNA sequences and *trans*-acting factors in regulatory regions (for reviews, see references 39 and 55). An open chromatin configuration can indicate that the underlying DNA is occupied by DNA-binding proteins, preventing its incorporation into nucleosomes (26, 46, 91). To identify transcription factors susceptible to interaction with DNA within the *pol* gene-hypersensitive site, we first performed a screening of the region using DNase I footprinting analysis. Five overlapping DNA subfragments encompassing the nuclease-hypersensitive region were used as probes in footprinting reactions with crude nuclear extracts from CEM and U937 cells. This initial screening identified four potential nuclear protein binding sites, called sites A, B, C, and D (Fig. 3A through C). Site A footprint was partial and not reproducible (data not shown) and was not further studied. Site B footprint encompassed nt 4521 to 4539 and was observed with extracts from U937 (Fig. 3A) and CEM cells (data not shown). Site C footprint encompassed nt 4682 to 4715 and was detected with extracts from U937 and CEM cells (Fig. 3B, lanes 4 and 6, respectively) but not with extracts from HeLa cells (Fig. 3B, lane 5). Site D footprint encompassed nt 4817 to 4841 and was observed with extracts from U937 (Fig. 3C) and CEM cells (data not shown). All footprints were flanked by sites of increased DNase I sensitivity (indicated by arrows in Fig. 3), a phenomenon often observed as consequence of protein binding to DNA. In addition, when the monovalent cation concentration (Na^+ or K^+) was increased, the protected regions became sensitive to DNase I attack, resulting in a digestion pattern indistinguishable from that observed in the control lane (Fig. 3A [compare lanes 3, 4, and 5 and 3, 6, and 7 for site B], Fig. 3C [compare lanes 3, 4, and 5 for site D], and data not shown [for site C]). To examine the specificity of the DNase I protection in site C, competition experiments were performed with a double-stranded 55-bp oligonucleotide encompassing the site C footprint. This competitor was found to abolish the site C footprint when present at a molar excess between 10- and 100-fold (Fig. 3B, lanes 9 and 10, respectively), whereas an oligonucleotide with similar size and base composition but with an unrelated sequence failed to inhibit footprint formation (data not shown).

Each site was further characterized by EMSA and methylation interference analysis.

Characterization of nuclear factors binding to site B. Comparison of the nucleotide sequences of site B among several HIV-1 isolates revealed that site B displayed a high degree of sequence polymorphism (Fig. 4A). To determine whether this polymorphism generated different binding specificities, four sites B with different sequences (U1 [= NY5], SF2, LAI, and HXB2) were tested as probes in EMSAs with nuclear extracts from the following human cell lines with different origins: SupT1, Jurkat, and A3.01 (a clonal derivative of the CEM cell line) ($CD4^+$ T-lymphocytic cell lines); HeLa (an epithelial cell line); U937 and KG-1 (promonocytic/macrophage cell lines); and Raji and Namalwa (B-lymphocytic cell lines). Two different patterns were observed, and these are illustrated in Fig. 4B by the U1 and SF2 sites B. A pattern of four major retarded DNA-protein complexes was observed with U1 site B as a probe (Fig. 4B): a ubiquitous complex (B1), a B-cell-specific complex (B2), a T-cell-specific complex (B3), and a monocyte/macrophage- and B-cell-specific complex (B4). With SF2 site B as a probe (Fig. 4B), only retarded complexes B1, B2, and B3 were generated. Oligonucleotides containing the site B sequence from LAI or HXB2 behaved like SF2 site B (data not shown). Thus, the macrophage- and B-cell-specific complex (B4) was detected only with U1 site B as a probe. Binding of proteins in complexes B1, B2, B3, and B4 was shown to be sequence specific by competition EMSAs (Fig. 4C and D). U1 site B oligonucleotide probe was incubated with nuclear extract from Raji or U937 cells in the presence of different competitor oligonucleotides (Fig. 4C). All complexes were inhibited by an excess of unlabeled homologous oligonucleo-

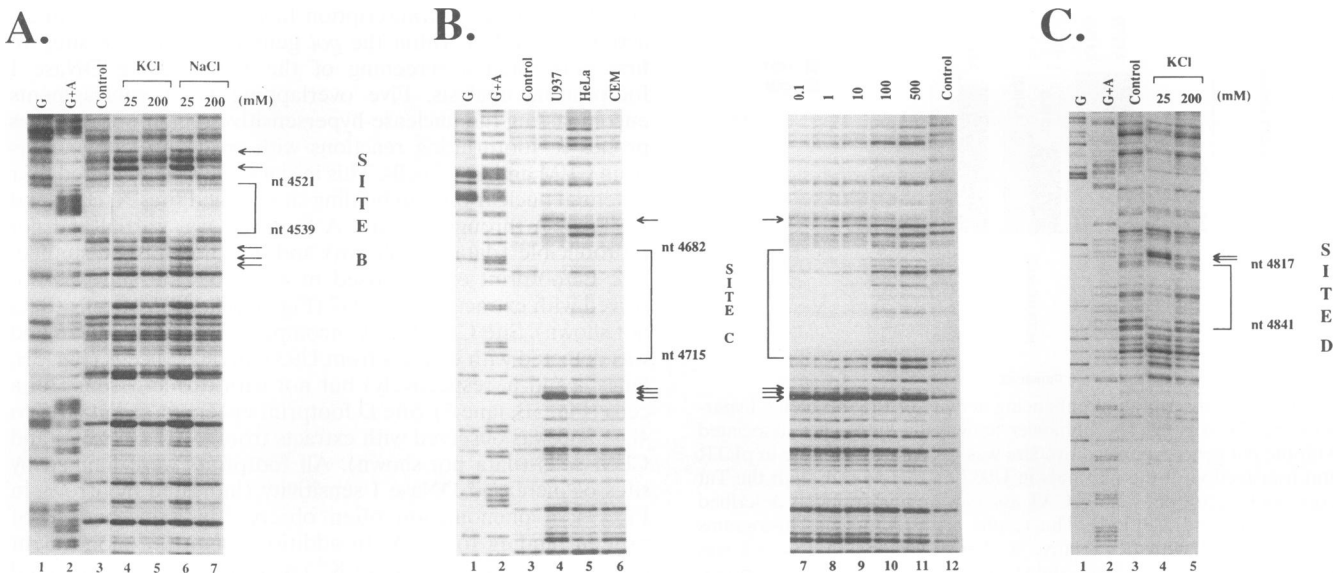

FIG. 3. DNase I footprinting analysis of the hypersensitive site. Overlapping DNA fragments encompassing the nuclease-hypersensitive region (nt 4346 to 4962) were used as probes in footprinting reactions. Probes were partially digested with DNase I in the absence of protein (control lanes) or after incubation with 50 μg of total nuclear extract from different cell lines. Protected areas are indicated by brackets. Numbers at the tops and bottoms of the brackets represent the positions of the nucleotides in the HIV-1_{NY5} genome. Maxam and Gilbert purine sequencing reaction products of the same probes were run in parallel (lanes 1 and 2) as size markers. Arrows indicate hypersensitive bands. (A) A 215-bp *Hin*dIII-*Sac*II restriction fragment of plasmid pHXB2(4491–4668) was 5′ end labeled at the *Hin*dIII site on the noncoding strand and was used as a probe in footprinting reactions with a nuclear extract from U937 cells at a final KCl concentration of 25 (lane 4) or 200 (lane 5) mM or at a final NaCl concentration of 25 (lane 6) or 200 (lane 7) mM. (B) A 221-bp *Hin*dIII-*Sac*II fragment of plasmid pHXB2(4601–4782) was 5′ end labeled at the *Hin*dIII site on the noncoding strand and was used as a probe in footprinting reactions with a nuclear extract from U937 (lane 4), HeLa (lane 5), or CEM (lane 6) cells. In lanes 7 to 11, increasing amounts of a double-stranded 55-bp oligonucleotide (nt 4670 to 4724) homologous to site C were incubated with a U937 nuclear extract before addition of the probe. The number at the top of each lane represents the molar ratio of the 55-bp oligonucleotide competitor to the HIV-1 probe. (C) A 269-bp *Hin*dIII-*Sac*II fragment of plasmid pHXB2(4731–4962) was 5′ end labeled at the *Hin*dIII site on the noncoding strand and was used as a probe in footprinting reactions with a U937 nuclear extract at a final NaCl concentration of 25 (lane 4) or 200 (lane 5) mM.

tide (Fig. 4C, lanes 2 to 4) but not by an oligonucleotide with an unrelated sequence containing a NF-κB binding site (Fig. 4C, lanes 5 to 7). Complexes B1, B2 (Fig. 4C), and B3 (data not shown) were inhibited by SF2 site B oligonucleotide (Fig. 4C, lanes 8 to 10) and site C oligonucleotide (Fig. 4C, lanes 11 to 13; see below). Similar competition experiments with SF2 site B as a probe and nuclear extracts from Raji and CEM cells confirmed these results (Fig. 4D). Similar competition EMSAs with LAI site B and HXB2 site B as probes showed similar results (data not shown).

The cell specificity of complex B4 and the presence of a PU box on the noncoding strand in U1 site B (Fig. 4A) suggested that B4 could result from binding of the macrophage- and B-cell-specific transcription factor Spi-1/PU.1 (33, 45, 56, 63). To determine whether the protein(s) present in complex B4 was related to the transcription factor Spi-1/PU.1, an oligonucleotide containing the PU box from the enhancer of an SV40 variant (45) was used as a competitor in EMSAs. The SV40PU.1 oligonucleotide inhibited specifically the formation of complex B4 at a 20-fold molar excess, whereas no competition of binding in complexes B1 and B2 was observed at the same molar excess (Fig. 4C, lane 15). A shorter oligonucleotide (U1 site B short [nt 4523 to 4542]) which was centered on the putative PU box present in site B was used in EMSAs and was found to generate only complex B4 when used as a probe with the different nuclear extracts (data not shown). The same oligonucleotide (U1 site B short) inhibited only the formation of complex B4 when used as a competitor (Fig. 4C, lanes 17 to 19). To demonstrate directly that the PU.1 protein could bind to U1 site B, the U1 site B short probe was incubated with human PU.1/Spi-1 protein, derived from lysates of COS1 cells transfected with a PU.1/Spi-1 expression vector (68). One retarded complex was generated (Fig. 5A, lane 2), whereas incubation with lysates of COS1 cells transfected with the control vector showed no retarded band (Fig. 5A, lane 1). Competition experiments showed that this protein-DNA complex was inhibited by the unlabeled homologous oligonucleotide (Fig. 5A, lanes 3 to 5) and by the SV40PU.1 oligonucleotide (Fig. 5A, lanes 6 to 8), whereas a mutated version of this oligonucleotide (SV40PU.1mut), in which two guanines crucial for the binding activity of the PU box were changed to cytosines, failed to compete when used at a 100-fold molar excess (Fig. 5A, lane 9), demonstrating the specificity of the complex. Thus, PU.1 transiently expressed in transfected COS1 cells specifically bound to the HIV-1 PU.1/Spi-1 binding site.

Interestingly, HIV-1 isolates which do not contain the PU box (HXB2, LAI, and SF2) do contain, just upstream of site B, a potential binding site for Ets1 and closely related factors of the Ets family of DNA-binding proteins to which PU.1 belongs (Fig. 4A). EMSAs with an oligonucleotide probe containing the potential Ets1 site of HIV-1_{HXB2} (nt 4505 to 4527) and extracts from SF9 cells infected with a recombinant baculovirus expressing the Ets1 protein (9) resulted in a retarded band (Fig. 5B, lane 2), whereas a control extract generated no shift (Fig. 5B, lane 1). This retarded band was specific since it was inhibited by an excess of unlabeled $Ets1_{HXB2}$ oligonucleotide (Fig. 5B, lane 5) or by an oligonucleotide corresponding to a

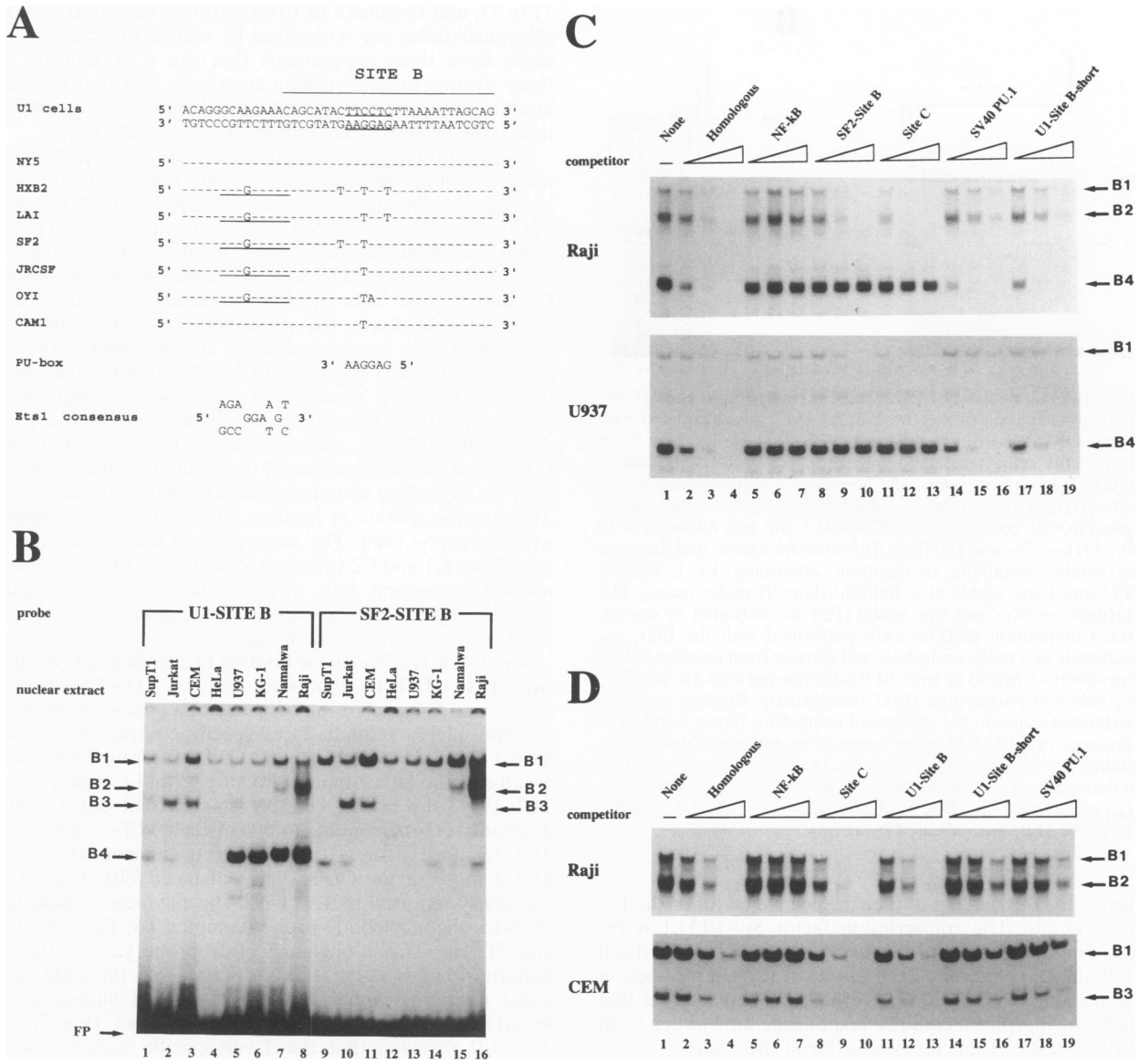

FIG. 4. EMSA analysis of nuclear factors binding to site B. (A) Alignment of the site B regions from several sequenced HIV-1 isolates. Data were retrieved from reference 57. The site B oligonucleotide (nt 4519 to 4545) is indicated by a line above the sequence. Nucleotides different from the U1 sequence are indicated. The potential Ets1 consensus binding site and PU box are indicated. (B) U1 site B and SF2 site B oligonucleotide probes (lanes 1 to 8 and 9 to 16, respectively) were incubated with nuclear extracts from human cell lines of different origins (indicated above the lanes). Retarded DNA-protein complexes (B1, B2, B3, and B4) and the free probes (FP) are indicated by arrows. (C) The U1 site B oligonucleotide probe was incubated with nuclear extract from Raji or U937 cells. Binding assays were performed in the absence of competitor (lanes 1) or in the presence of a 4 (lanes 2, 5, 8, 11, 14, and 17)-, 20 (lanes 3, 6, 9, 12, 15, and 18)-, or 100 (lanes 4, 7, 10, 13, 16, and 19)-fold molar excess of unlabeled specific competitor oligonucleotides. The competitors used are indicated above the lanes. The figure shows only the specific retarded bands of interest. (D) The SF2 site B oligonucleotide probe was incubated with nuclear extract from Raji or CEM cells. Binding assays were performed in the absence of competitor (lanes 1) or in the presence of a 4 (lanes 2, 5, 8, 11, 14, and 17)-, 20 (lanes 3, 6, 9, 12, 15, and 18)-, or 100 (lanes 4, 7, 10, 13, 16, and 19)-fold molar excess of unlabeled specific competitor oligonucleotides. The competitors used are indicated at the top of each lane. The figure shows only the specific retarded bands of interest.

previously described high-affinity Ets1 binding site (32) (Fig. 5B, lane 3) but not by the same oligonucleotide containing a mutation that eliminates Ets1 binding (ETS1mut) (Fig. 5B, lane 4). No complex was formed in similar EMSAs with the homologous region of NY5 as a probe (data not shown).

The recognition core sequence 5′-GGA(A/T)-3′ is present in the Ets-binding sites described so far for Ets1, Erg, Elf-1, E74, PEA3, Spi-1/PU.1, and Spi-B and appears to be essential for Ets proteins to bind DNA (42, 60, 85). However, multiple Ets family members display distinct DNA-binding specificities that are determined in part by the bases flanking this sequence (11, 74, 85, 86). For this reason, the binding of Ets1 to the PU box of U1 site B and the binding of PU.1/Spi-1 to the Ets1 site were also tested. The Ets1 site was found to bind only Ets1 and not PU.1/Spi-1, whereas the PU box bound only PU.1/Spi-1 and not Ets1 (data not shown). Thus, the two sites identified

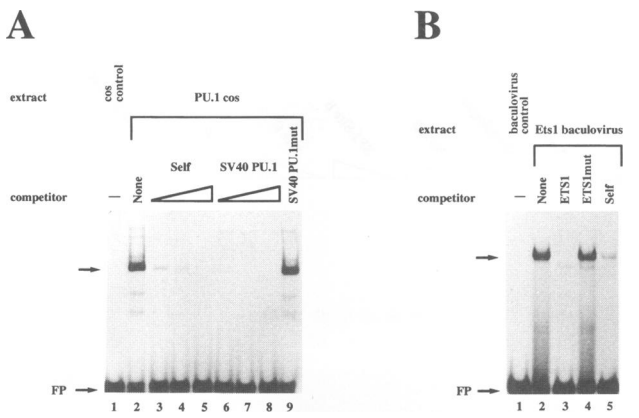

FIG. 5. Binding analysis of PU.1 and Ets1. (A) PU.1. The U1 site B short oligonucleotide probe (nt 4523 to 4542) was incubated with lysates of COS1 cells transfected with ΔEB-PU.1 (lanes 2 to 9) or lysates of COS1 cells transfected with the ΔEB vector (cos control [lane 1]). In competition assays, the homologous oligonucleotide was added at a 10 (lane 3)-, 40 (lane 4)-, and 100 (lane 5)-fold molar excess; an oligonucleotide containing the SV40PU.1 site was added at a 10 (lane 6)-, 40 (lane 7)-, and 100 (lane 8)-fold molar excess; and the same oligonucleotide containing a mutation abolishing PU.1 binding (SV40PU.1mut) was added at a 100-fold (lane 9) molar excess. The DNA-protein complex and free probe (FP) are indicated by arrows. (B) Ets1. Competition EMSAs were performed with the $Ets1_{HXB2}$ oligonucleotide as a probe and whole-cell extracts from uninfected SF9 cells (baculovirus control) or from SF9 cells infected with an Ac-NPV-Ets1 recombinant baculovirus (Ets1 baculovirus). Binding reactions were performed either in the absence of competitor (lanes 1 and 2) or in the presence of a 100-fold molar excess of an optimized high-affinity Ets1 binding site oligonucleotide (lane 3), in the presence of a mutant version derived from it (lane 4), or in the presence of the homologous $Ets1_{HXB2}$ oligonucleotide (lane 5). The positions of the Ets1 complex and free probe (FP) are indicated by arrows.

are specific for the binding of their respective factor of the Ets oncogene family. The transcription factor Spi-1/PU.1 is expressed in macrophages and B cells but not in T-lymphoid cell lines (30, 45, 68), whereas Ets1 is known to be expressed in thymocytes and T cells (8a). Therefore, our data suggest that the site B polymorphism could be responsible for a switch from a macrophage/B-cell specificity to a T-cell specificity.

Characterization of nuclear factors binding to site C. Interactions of nuclear factors with the site C region were analyzed by EMSAs with an oligonucleotide probe (site C [nt 4681 to 4701]) centered on the site C footprint described above. Incubation of this probe with nuclear extracts from different cell lines generated a pattern of three retarded complexes (Fig. 6A): a ubiquitous complex (complex C1 [lanes 1 to 9]), a B-cell-specific complex (complex C2 [lanes 7 and 8]), and a T-cell-specific complex (complex C3 [lanes 2 and 3]). This pattern was highly reminiscent of that observed with site B, except for the B4 complex, suggesting that site B and site C bind the same nuclear factors. The specificities of the complexes were demonstrated by competition experiments (Fig. 6B). When a CEM nuclear extract was used, complexes C1 and C3 were outcompeted by the homologous site C oligonucleotide (Fig. 6B, lanes 2 to 4) or by an oligonucleotide containing site B (HXB2) (Fig. 6B, lanes 8 to 10) but not by the NF-κB oligonucleotide (Fig. 6B, lanes 5 to 7). Similar results were obtained in competition EMSAs with a Raji or U937 nuclear extract (data not shown). In addition, reverse experiments in which a labeled probe containing site B was used and inhibited by unlabeled oligonucleotide containing site C (Fig. 4C, lanes 11 to 13, and D, lanes 8 to 10) confirmed that site B and site C oligonucleotides are recognized by related proteins. We conclude from these experiments that site C specifically binds three distinct nuclear protein complexes and that the factors binding to site C have a DNA-binding specificity similar to those of factors B1, B2, and B3 binding to site B.

Methylation interference analysis of site C was performed to identify contact guanine and adenine residues in the DNA-binding site. An oligonucleotide consisting of site C (nt 4683 to 4716) was 5′ end labeled on the coding strand, partially methylated with dimethyl sulfate, incubated with extract, and subjected to EMSAs. DNAs were isolated from complexes C1, C2, and C3 and from the free probe and were treated with piperidine to cleave the phosphodiester backbone at the position of each methylated base. The positions in the DNA sequence at which methylation interfered with formation of the complex were revealed by comparison of the cleavage patterns of DNAs from the retarded and from the free probes. As shown in Fig. 6C, analysis of DNA purified from complex C3 showed that methylation of the guanine residue at position 4691 on the coding strand abolished formation of this complex. The guanine residue at position 4683 could not be examined with the probe used. The same analysis was carried out with complexes C1 and C2 (data not shown), and the same guanine residue at position 4691 was identified to be important for binding, suggesting the presence of overlapping binding sites in site C or competition among factors for a same site.

The T-cell-specific factor binding to sites B and C is distinct from TCF-1/TCF-1α. Computer analysis of DNA sequences of sites B and C revealed the presence of a potential binding site for two highly related T-cell-specific transcription factors, TCF-1 and TCF-1α (48) (Fig. 7A). To determine whether the T-cell-specific factor binding to sites B and C belongs to the TCF-1/TCF-1α family, EMSAs were carried out with oligonucleotides corresponding to two distinct TCF-1 elements: the TCF-1α motif of the T-cell receptor α enhancer (87) and the TCF-1 motif of the CD3ε gene enhancer (80) (Fig. 7A). As previously reported (87), a T-cell-specific factor bound to the TCF-1α oligonucleotide probe (complex Tα; Fig. 7B, lanes 2 and 3). The TCF-1 oligonucleotide probe yielded the same pattern of bands as the TCF-1α probe, with the addition of a faster migrating complex detected with all nuclear extracts tested (data not shown). Comparison with site C (Fig. 7B, lanes 7 to 12) revealed that the T-cell-specific complex that we identified (C3) migrated more slowly than complex Tα (Fig. 7B, compare lanes 2 to 3 with lanes 8 to 9) and that the doublet characteristic of the TCF-1α and TCF-1 probes with the CEM nuclear extract was not observed with the site C probe (Fig. 7B, compare lane 2 with lane 8), indicating that the T-cell-specific factor present in complex C3 is different from TCF-1/TCF-1α. Competition experiments with TCF-1α and site C probes and a CEM nuclear extract confirmed this prediction (Fig. 7C). Complex Tα was efficiently inhibited by unlabeled homologous oligonucleotide (Fig. 7C, lane 2) but not by unrelated oligonucleotides containing a NF-κB or an AP-1 binding site (Fig. 7C, lanes 3 and 4). The TCF-1 oligonucleotide was a effective competitor of binding to the TCF-1α probe (Fig. 7C, lane 5), confirming that the same T-cell factor binds to both sites. Both site C and site B (HXB2) oligonucleotides partially competed for formation of complex Tα (Fig. 7C, lanes 6 and 7, respectively). These results are consistent with the presence of a low-affinity TCF-1 binding site in site B and site C (Fig. 7A). However, the T-cell factor present in complex C3 was distinct from TCF-1/TCF-1α, since it was not inhibited by either the TCF-1α or the TCF-1 oligonucleotide (Fig. 7C, lanes 12 and 13, respectively) but was well inhibited by site B (HXB2) (Fig.

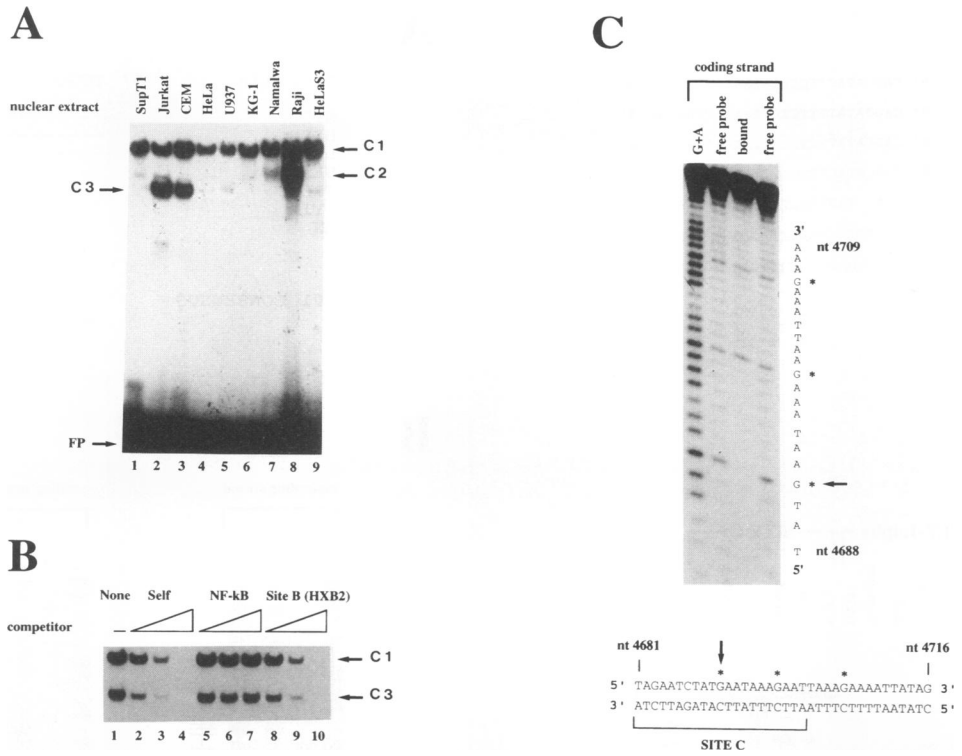

FIG. 6. EMSA and methylation interference analysis of site C. (A) The site C oligonucleotide probe was incubated with nuclear extracts from different cell lines (indicated at the top of each lane). The positions of complexes C1, C2, and C3 and of the free probe (FP) are indicated by arrows. (B) The site C probe was incubated with a CEM nuclear extract in the absence of competitor (lane 1) or in the presence of increasing molar excesses (10-, 40-, and 100-fold) of the homologous oligonucleotide (lanes 2, 3, and 4), of an NF-κB binding site oligonucleotide (lanes 5, 6, and 7) or of HXB2 site B oligonucleotide (lanes 8, 9, and 10). Retarded complexes are indicated (C1 and C3). The figure shows only the specific retarded bands of interest. (C) Methylation interference analysis of site C. An oligonucleotide (nt 4683 to 4716) consisting of site C was 5′ end labeled on the coding strand, partially methylated with dimethyl sulfate, and subjected to EMSA with a CEM nuclear extract. Complexes C3 (bound) and free probe were isolated and chemically cleaved by piperidine. Cleavage products were separated on a 15% denaturing polyacrylamide gel. Asterisks indicate the guanine residues. The position of the guanine residue that was contacted by the protein is indicated by an arrow. At the bottom, the sequence of the site C region is shown along with the methylation interference data. The numbers above the DNA sequence represent the positions of the nucleotides in the HIV-1_{NY5} genome.

7C, lane 14). Similar competition experiments with a Jurkat nuclear extract or the TCF-1 oligonucleotide probe showed identical results (data not shown). We conclude that sites B and C contain a high-affinity binding site for a T-cell factor that is distinct from TCF-1/TCF-1α.

The intragenic hypersensitive site contains a binding site for Sp1. A computer search identified a potential binding site for transcription factor Sp1 in the hypersensitive site region (Fig. 8A). In vitro DNase I footprinting analysis with affinity-purified human Sp1 resulted in protection of a GC-rich region extending from nt 4617 to 4636 on the noncoding strand (Fig. 8B, left panel) and from nt 4618 to 4637 on the coding strand (Fig. 8B, right panel). Footprints on both strands were flanked by sites of increased DNase I sensitivity (marked by the arrows in Fig. 8B). Since the Sp1 site was not perfectly conserved among different HIV-1 isolates (Fig. 8A), we also examined the DNase I protection pattern of the potential HXB2 Sp1 site on both strands and detected the same two regions protected from DNase I (data not shown). These experiments demonstrate that the DNA sequence identified by homology to the Sp1 consensus sequence in the U1-hypersensitive site binds Sp1 in vitro.

Characterization of nuclear factors binding to site D. Interactions of nuclear factors with the site D region were analyzed by EMSAs with the site D oligonucleotide probe (nt 4816 to 4851). Incubation of this probe resulted in one major nuclear protein complex (labeled D in Fig. 9A) with all nuclear extracts tested. The relative intensity of complex D varied among cell lines (Fig. 9A). Complex D was specific, since its formation was suppressed in the presence of an excess of unlabeled homologous oligonucleotide (Fig. 9B, lanes 2 to 5) but not by an excess of an unrelated oligonucleotide, even at a 400-fold molar excess (Fig. 9B, lanes 6 to 9). Methylation interference analysis of site D was performed for both strands to identify the guanine residues that are important for binding. Methylation of two guanines residues at positions 4830 and 4833 on the coding strand (Fig. 9C, left panel) and of one guanine residue at position 4835 on the noncoding strand (Fig. 9C, right panel) was shown to strongly interfere with binding to site D. These three guanine residues are positioned in the center of the region identified by DNase I footprinting analysis. In addition, the T residue at position 4838 showed enhanced intensity in the bound-probe lane relative to the free-probe lane (Fig. 9C). We conclude that a ubiquitously expressed factor interacts specifically with site D. No homology to any of the published DNA motifs has been found for this site.

DISCUSSION

In the present report, we describe the identification and physical characterization of a transcriptional regulatory ele-

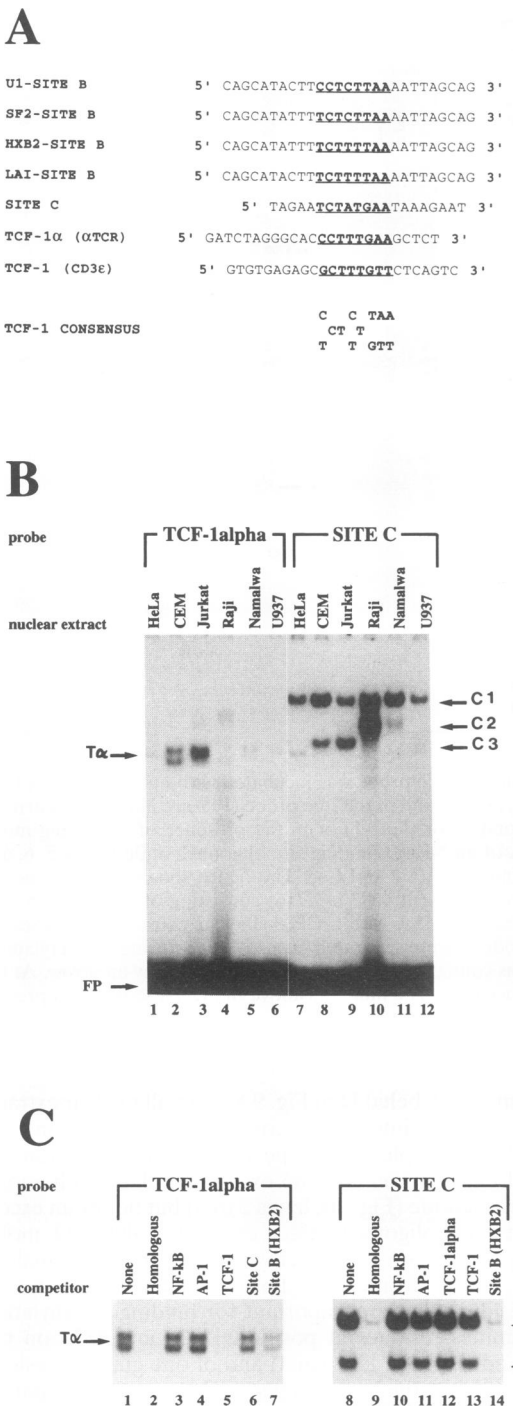

FIG. 7. The T-cell-specific factor binding to sites B and site C is distinct from TCF-1/TCF-1α. (A) The sequences of sites B and of site C are aligned with the TCF-1 consensus DNA-binding site (48). (B) TCF-1α (lanes 1 to 6) and site C (lanes 7 to 12) oligonucleotide probes were incubated with different nuclear extracts (indicated at the top of each lane). The positions of the free probe (FP), of the TCF-1α complex (Tα), and of complexes C1, C2, and C3 are indicated by arrows. (C) Competition EMSAs were performed with oligonucleotide TCF-1α or site C as probe and CEM nuclear extract. Binding reactions were performed either in the absence of competitor (lanes 1 and 8) or in the presence of a 50-fold molar excess of competitor oligonucleotides. The competitors used are indicated at the top of each lane. Complexes Tα, C1, and C3 are indicated. The figure shows only the specific retarded bands of interest.

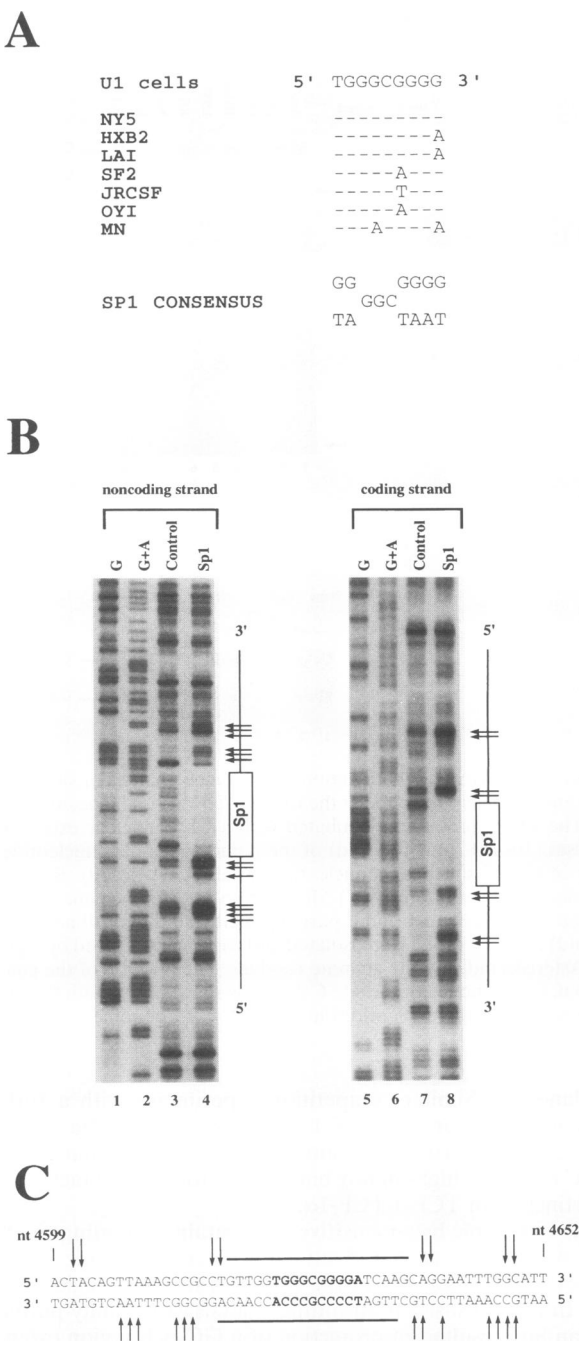

FIG. 8. DNase I footprinting analysis of Sp1 binding. (A) The potential Sp1 site identified in the *pol* gene of U1 cells is aligned with the homologous sequences of other HIV-1 isolates and with the Sp1 consensus binding site. Nucleotides different from the U1 sequence are indicated. Data were retrieved from reference 57. (B) The 263-bp *Xho*I-*Afl*II restriction fragment (nt 4480 to 4743) of plasmid pLTR-$U1_{sm}$(S) was labeled at the *Afl*II site either on the 5′ end on the noncoding strand with [γ-^{32}P]ATP and T4 polynucleotide kinase (left panel) or on the 3′ end on the coding strand with [α-^{32}P]dTTP and the Klenow fragment of DNA polymerase I (right panel). The probes were partially digested with DNase I in the absence of protein (lanes 3 and 7) or after incubation with 1 footprinting unit of affinity-purified Sp1 (Promega) (lanes 4 and 8). The Sp1-protected areas are indicated by open boxes. Maxam and Gilbert purine sequencing reaction products of the probes are shown in the two leftmost lanes as size markers. Arrows indicate sites of increased DNase I sensitivity associated with

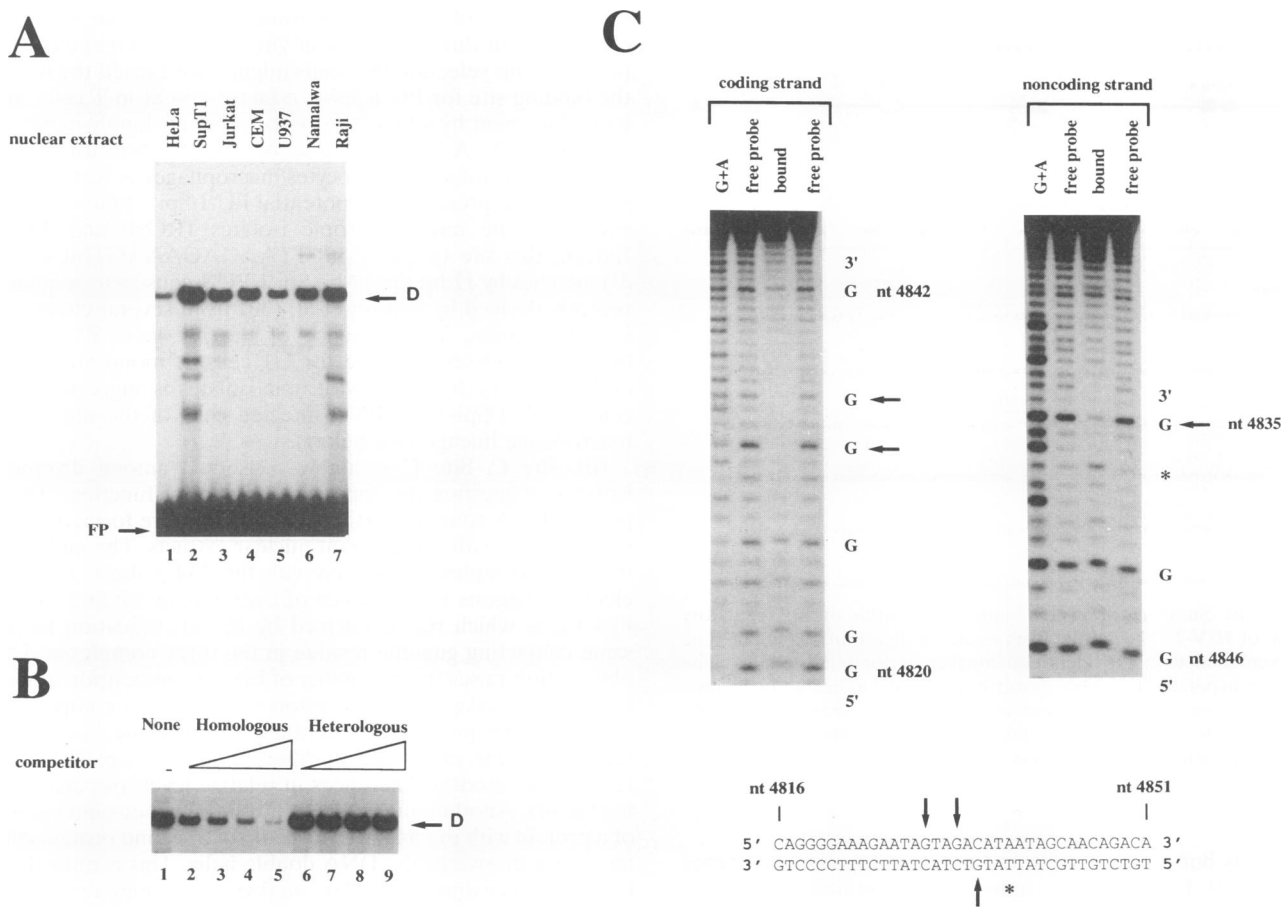

FIG. 9. EMSA and methylation interference analysis of site D. (A) Site D oligonucleotide probe was incubated with nuclear extracts from different cell lines (as indicated at the top of each lane). Free probe (FP) and the major nuclear protein complex (D) generated are indicated by arrows. (B) Competition EMSAs were performed with site D oligonucleotide as the probe and CEM nuclear extract. Binding reactions were performed either in the absence of competitor (lane 1) or in the presence of a 50 (lanes 2 and 6)-, 100 (lanes 3 and 7)-, 200 (lanes 4 and 8)-, or 400 (lanes 5 and 9)-fold molar excess of competitor oligonucleotides. The competitors used were either the homologous site D oligonucleotide (lanes 2 to 5) or the heterologous site C (nt 4683 to 4716) oligonucleotide (lanes 6 to 9). Complex D is indicated by an arrow. The figure shows only the specific retarded bands of interest. (C) Methylation interference analysis of site D. Site D oligonucleotide was 5′ end labeled either on the coding strand (left panel) or on the noncoding strand (right panel), partially methylated with dimethyl sulfate, and subjected to EMSA with a CEM nuclear extract. Complex D (bound) and free probe were isolated. After gel purification, DNA was chemically cleaved with piperidine. The cleavage products were separated on a 15% denaturing polyacrylamide gel. Guanine residues at positions where methylation interfered with binding are marked by arrows. The asterisk indicates a site of increased sensitivity to piperidine cleavage. At the bottom, the sequence of site D is shown along with a summary of the methylation interference data. Numbers above the DNA sequence represent the positions of the nucleotides in the HIV-1_{NY5} genome.

ment associated with a DNase I-hypersensitive site present in the *pol* gene of HIV-1. High-resolution mapping of the *pol*-hypersensitive region in U1 cells identified two hypersensitive sites after both DNase I and micrococcal nuclease digestion (Fig. 1; summarized in Fig. 10), indicating that a region centered around nt 4490 to 4766 is nucleosome free in vivo. A 500-bp fragment encompassing this hypersensitive region positively regulated transcription from the HIV-1 LTR in transient transfection assays. Several sequence motifs within the viral regulatory region which interact specifically with nuclear factors in vitro were defined physically by in vitro DNase I footprinting, EMSAs, and methylation interference experiments (summarized in Fig. 10).

DNA-binding motifs of the intragenic regulatory region. (i) Site B. Site B displays a high degree of sequence polymorphism; however, this polymorphism does not generate considerable variations of binding specificities. Four nuclear protein-DNA complexes (B1, B2, B3, and B4) were detected in EMSAs with U1 site B as a probe. Complex B4 was demonstrated to contain a protein(s) related to the transcription factor PU.1/Spi-1. This *PU.1/Spi-1* proto-oncogene-encoded factor is an *ets* gene family member which has been previously cloned by different methodologies (33, 45, 56, 63). PU.1/Spi-1 is expressed in erythroid cells, B cells, myelomonocytes, and

Sp1 binding. (C) The nucleotide sequences of the coding and noncoding strands within the Sp1 region are shown. Sites of increased DNase I sensitivity are indicated by arrows. Bars above and below the sequence indicate Sp1-protected areas of the coding and noncoding strands, respectively. Numbers above the DNA sequence represent the positions of the nucleotides in the HIV-1_{NY5} genome.

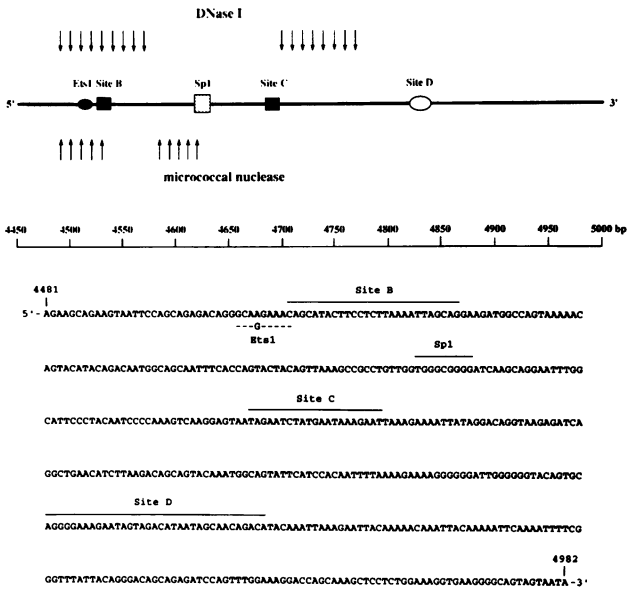

FIG. 10. Summary of protein binding sites within the *pol* regulatory region of HIV-1. The results from nuclease digestion of the *pol* gene are combined with the identified nuclear factor binding sites. The nuclease-hypersensitive sites shown in Fig. 1 are indicated by arrows. The sequence of the U1-hypersensitive site is shown, with bold overrules indicating the binding sites that were identified. The Ets1 site present in some HIV-1 isolates is indicated.

mast cells but not in T cells and nonhematopoietic cell types (30, 45, 68). DNA-binding studies have identified the consensus sequence for PU.1 binding as purine-rich segments containing a 5′-GGAA-3′ core sequence (42). PU.1/Spi-1 binding sites are present in a large number of B-lymphoid-, myeloid-, and erythroid cell-specific transcriptional elements, such as the enhancers of two SV40 variants (66), the promoter of the major histocompatibility complex class II gene I-Aβ (20), the mouse immunoglobulin κ3′ enhancer (67), the promoter of the human myeloid cell-specific CD11b gene (62), the lymphotropic papovavirus enhancer (23), and the B-cell-specific immunoglobulin μ heavy-chain gene enhancer (58). In many of these systems, the PU.1 sites have been shown to be crucial for the activity of the regulatory element. Importantly, a PU.1 binding site has recently been identified in the LTR of equine infectious anemia virus (EIAV), a lentivirus related to HIV-1 (12). In addition, this site has been shown to be the primary determinant for the transcriptional activity of the EIAV promoter in monocytes (12).

Our results demonstrate that the PU box of U1 site B and the Ets1 site identified just upstream of site B in several HIV-1 isolates are specific for binding of their respective factors, PU.1/Spi-1 and Ets1. Nevertheless, because proteins that bind DNA by means of an ETS domain share DNA recognition properties (74, 85, 86), we cannot rule out the possibility that other *ets* family members also bind to these sites. It is also worth noting that the Ets1 binding site is absent in U1 cells and present in the T-cell-tropic isolates HXB2, LAI, and SF2, which do not contain the PU.1 site (Fig. 4A).

The PU box is detected in a minority of HIV-1 isolates (NY5 and U1), and most isolates sequenced to date contain a site for other *ets* family members such as Ets1. However, most laboratory isolates of HIV-1 have been selected in vitro for their ability to replicate in T-cell lines and are known not to be truly representative of viral strains found in vivo, since sequence changes occur during passage of virus in vitro in tissue culture (47, 53). This selection in T cells might have caused the loss of the binding site for PU.1/Spi-1, a factor absent in T cells, and its replacement by a binding site for other *ets* family members such as Ets1. A potential role of the PU box for HIV-1 regulation in infected monocytes/macrophages is further suggested by the presence of a potential PU.1/Spi-1 binding site in site B of the monocytotropic isolates JRCSF and JRFL. Indeed, this site (5′-[nt 4536]-TTAAGAGAAAGT[nt 4525]-3′) matches by 11 bp the 12-bp Spi-1/PU.1 consensus sequence recently derived by Galson et al. (30) from several characterized PU boxes. The presence of the *pol* gene PU.1/Spi-1 binding site described here for U1 cells of monocytic origin and potentially for monocytotropic isolates is suggestive of a role of PU.1/Spi-1 in HIV-1-infected cells of the monocyte/macrophage lineage (see below).

(ii) Site C. Site C is highly conserved among divergent isolates, suggesting an important regulatory function. Three protein-DNA complexes (C1, C2, and C3) were formed at site C in EMSAs with the different nuclear extracts. The multiplicity of the complexes observed with the 23-bp site C oligonucleotide suggests the presence of overlapping binding sites, a hypothesis which was confirmed by the identification of the same contacting guanine residue in the three complexes. This observation raises the possibility of in vivo competition among factors at overlapping recognition sites. In vivo competition among factors provides a potentially sensitive mechanism for regulating gene expression in different cell types or tissues in response to modest differences in relative levels of participating factors. Another alternative is the simultaneous interaction of a protein with the major groove and of a second protein with the minor groove of the DNA double helix. This is unlikely in this case, since dimethyl sulfate methylates guanine residues at the N-7 position in the major groove. It is also possible that the different binding factors do not all interact with the DNA at the same time, depending on the cellular conditions. Our competition experiments demonstrate that the factors present in complexes C1, C2, and C3 have a DNA-binding specificity similar to that of factors B1, B2, and B3, respectively, identified within site B.

Sites B and C contain a sequence homologous to the TCF-1 consensus sequence (Fig. 7A). The results of our EMSA experiments showed that site B and site C contain a low-affinity TCF-1 binding site, since they were able to compete for TCF-1 binding to the TCF-1α probe (Fig. 7C). Site B competed more efficiently for binding than site C, presumably because the latter matches the TCF-1 consensus sequence less well (Fig. 7A). However, sites B and C bind with a much higher affinity a T-cell-specific factor (complexes B3 and C3) distinct from TCF-1/TCF-1α. Our results do not exclude the possibility that this T-cell factor is a member of the TCF-1 family. Further purification and isolation of a cDNA clone encoding this T-cell factor will clarify the answer to this question.

In addition to the potential TCF-1 binding site, sites B and site C share another sequence homology, from which we derived the consensus sequence 5′-CATANWTYY-3′ (see below), in which N represents any nucleotide, W represents T or A, and Y represents C or T. This sequence could constitute the DNA-binding site of the nuclear factors binding to both site B and site C. A data base search indicated that this consensus sequence is present in several cellular and viral genes, including the LTR of Gardner-Arnstein feline leukemia virus, the mouse immunoglobulin κ gene, the human interferon α gene, the chicken c-*myc* gene, the rat interleukin 3 gene, and in the genomes of human papillomavirus type 18,

SV40, human Epstein-Barr virus, human herpes simplex virus type 1, and human hepatitis B virus (data not shown).

Site C,	CATAGATTC
LAI site B,	CATACTTTC
SF2 site B,	CATAATTTT
U1 site B,	CATACTTCC
HXB2 site B	CATATTTTC
consensus sequence,	CATANWTYY

(iii) Sp1 site. The *pol* regulatory element contains a binding site for Sp1, a ubiquitous transcription factor that binds to GC-rich sequence elements in a wide variety of viral and cellular promoters (22, 41). Sp1 can mediate the formation of a DNA loop in vitro by self-association of DNA-bound Sp1 molecules joining together distant DNA fragments (50, 77). Furthermore, in vivo, a functional interaction between upstream and downstream Sp1 binding sites results in a marked synergistic activation of transcription of the modified thymidine kinase promoter (16). The presence of three tandemly arranged Sp1 sites in the HIV-1 5′-LTR and of one Sp1 site in the *pol* nuclease-hypersensitive site suggests a putative interaction between these elements. Such an association could bring in close proximity other factors bound to the 5′-LTR and to the *pol* regulatory element, and this arrangement could play a role in transcriptional regulation of the HIV-1 promoter.

(iv) Site D. A ubiquitously expressed factor interacts specifically with the site D sequence, which is well conserved among many HIV-1 isolates. The relative intensity of complex D varied among cell lines, suggesting quantitative differences of the factor, although other parameters such as contamination of the nuclear contents with cytoplasm may alter relative factor concentrations. Computer analysis revealed no homology between this binding site and the recognition sequences of known transcription factors.

DNA regulatory regions in eukaryotic genomes frequently adopt a nuclease-hypersensitive configuration (36). This hypersensitivity is thought to result from the disruption of the packaging of DNA into nucleosomes by DNA-bound regulatory proteins (26, 46, 91). The absence of a hypersensitive site in the *pol* gene of HIV-1 integrated in two chronically infected T-cell lines could be due to differences in nucleotide sequence among HIV-1 isolates or to the presence of different regulatory factors in different cell types (lymphoid versus monocyte/macrophage cells). It is also possible that the *pol* region is nucleosome free in ACH2 and 8E5 cells but is occupied by bound factors in such a way as to adopt a configuration resistant to nuclease digestion. In vivo footprinting of the protein-DNA interactions occurring in intact U1, ACH2, and 8E5 cells will provide additional information on this issue. An apparent contradiction lies in the observation that the hypersensitive site is observed only in U1 cells (derived from U937) and not in lymphoid cells (ACH2 and 8E5) (82), whereas transient functional studies presented here show the fragment to be equally active in both cell types. However, the regulation of transiently transfected DNA may not completely reflect that found in the natural chromosomal context, since it has been shown that the bulk of DNA is not assembled into chromatin under these experimental conditions (4). Recent experiments by Kim and colleagues (44) support this latter point. These authors studied linker-scanning mutations in the LTR of HIV-1 and compared the effects of each of these mutations in vitro (transient transfection of plasmid constructs) and in vivo (effect on viral growth of viral mutants). Several mutants demonstrated effects on viral replication that were more severe than what could have been anticipated from the effects of the same mutations in transient transfection assays (44). A similar paradox has been observed for the myoD promoter/enhancer element; in transient transfection, this element is active in myogenic as well as in many nonmyogenic cell lines, whereas in transgenic mice its activity is exclusively restricted to the skeletal muscle lineage (34).

These observations suggest that despite its weak activity in transient transfection assays, the *pol* gene element described here may exert a more important effect in the context of an integrated provirus. Experiments with viral strains containing mutations in each of the binding sites identified in this report will be necessary to assess the true biological significance of each of these sites for HIV-1 replication.

The positive regulatory element identified here could play an important role in the macrophage tropism of HIV-1. Previous studies have demonstrated that HIV-1 macrophage tropism is primarily determined at the level of the entry, by regions of the surface envelope glycoprotein gp120, including the V3 loop region (14, 37, 61, 75, 89, 90, 93). However, it is also clear that other determinants of tropism may manifest themselves after viral entry, as illustrated by the *vpr* and *vpu* genes, which are required for efficient viral replication in primary monocytes/macrophages (7, 89). The presence in the *pol* gene of HIV-1 of a regulatory region containing a PU.1 site and associated with a monocyte-specific DNase I-hypersensitive site suggests that an additional determinant of macrophage tropism might be present at the level of transcription. Whereas a large number of studies have examined the lymphoid specificity of the HIV promoter (for reviews, see references 18, 31, 35, 40, 64, and 79), little is known regarding its monocyte/macrophage specificity or other *cis*-acting elements responsible for this specificity. The tropism of other viruses for their target tissues is determined in part at the transcriptional level. In murine leukemia viruses, a switch in tropism to T cells was found to be caused by alterations in the U3 region of the LTR (13). The enhancer element of lymphotropic papovavirus, which contains a PU box required for its activity, contributes to the restricted tropism of the virus for primate B lymphocytes (23). Importantly, the PU.1/Spi-1 binding site situated in the EIAV LTR is a primary determinant of promoter activity in transfected monocyte cell lines, suggesting that it may be relevant to the macrophage tropism of the lentivirus EIAV (12). Most studies of transcriptional regulatory mechanisms of retroviruses have focused on the 5′-LTR. However, studies by several groups have identified elements located outside the LTR that play a significant role in determining the transcriptional rate and tissue specificity of retroviral promoters (5, 54, 65, 70). The element described here in the *pol* gene of HIV-1 offers another example of this concept.

The novel regulatory element reported here is located between the two functional domains of the intragenic enhancer previously identified in HeLa cells (83). Although each of these subdomains is not active in T-lymphoid and monocytoid cell lines when isolated (8), each might nevertheless contribute to the regulatory activity of the *pol*-hypersensitive region. It is possible that the complete functional unit of the HIV-1 intragenic regulatory region is constituted of these three elements: the regulatory region described here and the two domains described previously (83). Such a situation has been reported for the SV40 enhancer, which can be dissected into multiple segments, each with a characteristic cell type specificity and distinct in some cases from that of the complete enhancer (73). Specific expression of immunoglobulin heavy-chain genes in B cells is transcriptionally controlled by a core enhancer that is equally active in both lymphoid and nonlymphoid cells and a flanking region upstream of this core which

increases immunoglobulin heavy-chain enhancer activity in B cells but represses this activity in nonlymphoid cells (38). The potential importance of the two previously identified subdomains is suggested by the presence of three functional AP-1 binding sites in one of the fragments (81) and the recent identification in the other fragment of a functional glucocorticoid response element (76).

The observations reported here identify a novel regulatory element associated with the *pol* gene of HIV-1 and putative *cis*-acting elements contributing to its transcriptional activity. Further work will be necessary to determine the biological significance of the *pol*-hypersensitive region in the virus life cycle, with a special emphasis on the potential role of this region in macrophage tropism. Site-directed mutations of the protein binding sites identified will be examined in the context of an infectious HIV-1 provirus for their potential effects on virus expression and replication. Studying the molecular events that regulate HIV-1 transcription should lead to a better understanding of AIDS pathogenesis.

ACKNOWLEDGMENTS

We thank Monique Dubois-Dalcq for support, encouragement, and suggestions during the course of this work. We thank Aboubaker El Kharroubi for nuclear extracts samples and Anthony Fauci, Tom Folks, and Guido Poli for making their cell lines available through the AIDS Research and Reference Reagent Program (Division of AIDS, NIAID, NIH). We are grateful to Monique Dubois-Dalcq, Jin Kim, Louis Droogmans, and Richard Currie for critical reading and discussion of the manuscript.

C.V.L. is an aspirant of the Fonds National de la Recherche Scientifique (Brussels, Belgium) and was on leave from the Free University of Brussels.

REFERENCES

1. **Adachi, A., H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin.** 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. **59:**284–291.
2. **Amemiya, K., R. Traub, L. Durham, and E. O. Major.** 1992. Adjacent nuclear factor-1 and activator protein binding sites in the enhancer of the neurotropic JC virus. J. Biol. Chem. **267:**14204–14211.
3. **Angel, P., I. Baumann, B. Stein, H. Delius, H. J. Rhamsdorf, and P. Herrlich.** 1987. 12-*O*-Tetradecanoyl-phorbol-13-acetate induction of the human collagenase gene is mediated by an inducible enhancer element located in the 5′-flanking region. Mol. Cell. Biol. **7:**2256–2266.
4. **Archer, T. K., P. Lefebvre, R. G. Wolford, and G. L. Hager.** 1992. Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. Science **255:**1573–1576.
5. **Arrigo, S., M. Yun, and K. Beemon.** 1987. *cis*-acting regulatory elements within *gag* genes of avian retroviruses. Mol. Cell. Biol. **7:**388–397.
6. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Strühl.** 1987. Current protocols in molecular biology. Wiley, New York.
7. **Balotta, C., P. Lusso, R. Crowley, R. C. Gallo, and G. Franchini.** 1993. Antisense phosphorothioate oligodeoxynucleotides targeted to the *vpr* gene inhibit human immunodeficiency virus type 1 replication in primary human macrophages. J. Virol. **67:**4409–4414.
8. **Becker, N., A. Burny, and E. Verdin.** Unpublished observation.

8a. **Bhat, N. K., R. 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:**3161–3165.

9. **Bosselut, R., J. F. Duvall, A. Gégonne, M. Bailly, A. Hémar, J. Brady, and J. Ghysdael.** 1990. The product of the c-ets-1 proto-oncogene and the related Ets2 protein act as transcriptional activators of the long terminal repeat of human T cell leukemia virus HTLV-1. EMBO J. **9:**3137–3144.
10. **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.
11. **Brown, T. A., and S. L. McKnight.** 1992. Specificities of protein-protein and protein-DNA interaction of GABPα and two newly defined *ets*-related proteins. Genes Dev. **6:**2502–2512.
12. **Carvalho, M., and D. Derse.** 1993. The PU.1/Spi-1 proto-oncogene is a transcriptional regulator of a lentivirus promoter. J. Virol. **67:**3885–3890.
13. **Celander, D., and W. Haseltine.** 1984. Tissue-specific transcription preference as a determinant of cell tropism and leukaemogenic potential of murine retroviruses. Nature (London) **312:**159–162.
14. **Cheng-Mayer, C., M. Quiroga, J. W. Tung, D. Dina, and J. A. Levy.** 1990. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. J. Virol. **64:**4390–4398.
15. **Clouse, K. A., D. Powell, I. Washington, G. Poli, K. Strebel, W. Farrar, P. Barstad, J. Kovacs, A. S. Fauci, and T. M. Folks.** 1989. Monokine regulation of human immunodeficiency virus type 1 expression in a chronically infected human T cell clone. J. Immunol. **142:**431–438.
16. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59:**827–836.
17. **Cullen, B. R.** 1990. The HIV-1 Tat protein: an RNA sequence-specific processivity factor? Cell **63:**655–657.
18. **Cullen, B. R.** 1991. Regulation of human immunodeficiency virus replication. Annu. Rev. Microbiol. **45:**219–250.
19. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
20. **Dorn, A., H. J. Fehling, W. Koch, M. Le Meur, P. Gerlinger, C. Benoist, and D. Mathis.** 1988. B-cell control region at the 5′ end of a major histocompatibility complex class II gene: sequences and factors. Mol. Cell. Biol. **8:**3975–3987.
21. **Dynan, W. S.** 1987. DNase I footprinting as an assay for mammalian gene regulatory proteins, p. 75–87. *In* J. K. Setlow (ed.), Genetic engineering principles and methods, vol. 9. Plenum Publishing Corp., New York.
22. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35:**79–87.
23. **Erselius, J. R., B. Jostes, A. K. Hatzopoulos, L. Mosthaf, and P. Gruss.** 1990. Cell-type-specific control elements of the lymphotropic papovavirus enhancer. J. Virol. **64:**1657–1666.
24. **Fauci, A. S.** 1988. The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. Science **239:**617–622.
25. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
26. **Felsenfeld, G.** 1992. Chromatin as an essential part of the transcriptional mechanism. Nature (London) **355:**219–224.
27. **Folks, T. M., J. Justement, A. Kinter, S. Schnittman, J. Orenstein, G. Poli, and A. S. Fauci.** 1988. Characterization of a promonocytic clone chronically infected with HIV and induction by 13-phorbol-12-myristate acetate. J. Immunol. **140:**1117–1122.
28. **Folks, T. M., D. Powell, M. Lightfoote, S. Koenig, A. S. Fauci, S. Benn, A. Rabson, D. Daugherty, H. E. Gendelman, M. D. Hoggan, S. Venkatesan, and M. A. Martin.** 1986. Biological and biochemical characterization of a cloned leu-3$^-$ cell surviving infection with the acquired immune deficiency syndrome retrovirus. J. Exp. Med. **164:**280–290.
29. **Frankel, A. D.** 1992. Activation of HIV transcription by Tat. Curr. Opin. Genet. Dev. **2:**293–298.
30. **Galson, D. L., J. O. Hensold, T. R. Bishop, M. Schalling, A. D. D'Andrea, C. Jones, P. E. Auron, and D. E. Housman.** 1993. Mouse β-globin DNA-binding protein B1 is identical to a proto-oncogene, the transcription factor *Spi-1*/PU.1, and is restricted in expression to hematopoietic cells and the testis. Mol. Cell. Biol. **13:**2929–2941.
31. **Gaynor, R.** 1992. Cellular transcription factors involved in the

regulation of HIV-1 gene expression. AIDS **6:**347–363.
32. **Gégonne, A., B. Punyammalee, B. Rabault, R. Bosselut, S. Seneca, M. Crabeel, and J. Ghysdael.** 1992. Analysis of the DNA binding and transcriptional activation properties of the Ets1 oncoprotein. New Biol. **4:**512–519.
33. **Goebl, M. G.** 1990. The PU.1 transcription factor is the product of the putative oncogene Spi-1. Cell **61:**1165–1166.
34. **Goldhamer, D. J., A. Faerman, M. Shani, and C. P. Emerson, Jr.** 1992. Regulatory elements that control the lineage-specific expression of myoD. Science **256:**538–542.
35. **Greene, W. C.** 1990. Regulation of HIV-1 gene expression. Annu. Rev. Immunol. **8:**453–475.
36. **Gross, D. S., and W. T. Garrard.** 1988. Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. **57:**159–197.
37. **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253:**71–74.
38. **Imler, J.-L., C. Lemaire, C. Wasylyk, and B. Wasylyk.** 1987. Negative regulation contributes to tissue specificity of the immunoglobulin heavy-chain enhancer. Mol. Cell. Biol. **7:**2558–2567.
39. **Johnson, P. F., and S. L. McKnight.** 1989. Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. **58:**799–839.
40. **Jones, K. A.** 1989. HIV trans-activation and transcription control mechanisms. New Biol. **1:**127–135.
41. **Kadonaga, J. T., K. R. Carner, F. R. Masiarz, and R. Tjian.** 1987. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell **51:**1079–1090.
42. **Karim, F. D., L. D. Urness, C. S. Thummel, M. J. Klemsz, S. R. McKercher, A. Celada, C. 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.
43. **Karn, J., and M. A. Graeble.** 1992. New insights into the mechanism of HIV-1 transactivation. Trends Genet. **8:**365–368.
44. **Kim, J. Y. H., F. Gonzales-Scarano, S. L. Zeichner, and J. C. Alwine.** 1993. Replication of type 1 human immunodeficiency viruses containing linker substitution mutations in the −201 to −130 region of the long terminal repeat. J. Virol. **67:**1658–1662.
45. **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.
46. **Kornberg, R. D., and Y. Lorch.** 1992. Chromatin structure and transcription. Annu. Rev. Cell Biol. **8:**563–587.
47. **Kusumi, K., B. Conway, S. Cunningham, A. Berson, C. Evans, A. K. N. Iversen, D. Colvin, M. V. Gallo, S. Coutre, E. G. Shpaer, D. V. Faulkner, A. de Ronde, S. Volkman, C. Williams, M. S. Hirsch, and J. I. Mullins.** 1992. Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. J. Virol. **66:**875–885.
48. **Leiden, J. M.** 1992. Transcriptional regulation during T-cell development: the αTCR gene as a molecular model. Immunol. Today **13:**22–30.
49. **Lenardo, M. J., and D. Baltimore.** 1989. NF-κB: a pleiotropic mediator of inducible and tissue-specific gene control. Cell **58:** 227–229.
50. **Mastrangelo, I. A., A. J. Courey, J. S. Wall, S. P. Jackson, and P. V. C. Hough.** 1991. DNA looping and Sp1 multimer links: a mechanism for transcriptional synergism and enhancement. Proc. Natl. Acad. Sci. USA **88:**5670–5674.
51. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
52. **McCune, J. M.** 1991. HIV-1: the infective process *in vivo*. Cell **64:**351–363.
53. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Månson, B. Asjö, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. Cell **58:**901–910.
54. **Miller, C. L., R. Garner, and V. Paetkau.** 1992. An activation-dependent, T-lymphocyte-specific transcriptional activator in the mouse mammary tumor virus *env* gene. Mol. Cell. Biol. **12:**3262–3272.
55. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. Science **245:**371–378.
56. **Moreau-Gachelin, F., D. Ray, M.-G. Mattei, P. Tambourin, and A. Tavitian.** 1989. The putative oncogene Spi-1: murine chromosomal localization and transcriptional activation in murine acute erythroleukemias. Oncogene **4:**1449–1456.
57. **Myers, G., J. A. Berzofsky, B. Korber, R. F. Smith, and G. N. Pavlakis.** 1992. Human retroviruses and AIDS. A compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, N. Mex.
58. **Nelsen, B., G. Tian, B. Erman, J. Gregoire, R. Maki, B. Graves, and R. Sen.** 1993. Regulation of lymphoid-specific immunoglobulin μ heavy chain gene enhancer by ETS-domain proteins. Science **261:**82–86.
59. **Neumann, J. R., C. A. Morency, and K. O. Russian.** 1987. A novel rapid assay for chloramphenicol acetyltransferase gene expression. BioTechniques **5:**444–447.
60. **Nye, J. A., J. M. Petersen, C. V. Gunther, M. D. Jonsen, and B. J. Graves.** 1992. Interaction of murine Ets1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. **6:**975–990.
61. **O'Brien, W. A., Y. Koyanagi, A. Namazie, J.-Q. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. Y. Chen.** 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature (London) **348:**69–73.
62. **Pahl, H. L., R. J. Scheibe, D.-E. Zhang, H.-M. Chen, D. L. Galson, R. A. Maki, and D. G. Tenen.** 1993. The proto-oncogene PU.1 regulates expression of the myeloid-specific CD11b promoter. J. Biol. Chem. **268:**5014–5020.
63. **Paul, R., S. Schuetze, S. L. Kozak, C. A. Kozak, and D. Kabat.** 1991. The *Sfpi-1* proviral integration site of Friend erythroleukemia encodes the *ets*-related transcription factor PU.1. J. Virol. **65:**464–467.
64. **Pavlakis, G. N., and B. K. Felber.** 1990. Regulation of expression of human immunodeficiency virus. New Biol. **2:**20–31.
65. **Payvar, F., D. DeFranco, G. L. Firestone, B. Edgar, O. Wrange, S. Okret, J.-A. Gustafsson, and K. R. Yamamoto.** 1983. Sequence-specific binding of glucocorticoid receptor to MTV DNA at sites within and upstream of the transcribed region. Cell **35:**381–392.
66. **Pettersson, M., and W. Schaffner.** 1987. A purine-rich DNA sequence motif present in SV40 and lymphotropic papovavirus binds a lymphoid-specific factor and contributes to enhancer activity in lymphoid cells. Genes Dev. **1:**962–972.
67. **Pongubala, J. M. R., S. Nagulapalli, M. J. Klemsz, S. R. McKercher, R. A. Maki, and M. L. Atchison.** 1992. PU.1 recruits a second nuclear factor to a site important for immunoglobulin κ3′ enhancer activity. Mol. Cell. Biol. **12:**368–378.
68. **Ray, D., R. Bosselut, J. Ghysdael, M.-G. Mattei, A. Tavitian, and F. Moreau-Gachelin.** 1992. Characterization of Spi-B, a transcription factor related to the putative oncoprotein Spi-1/PU.1. Mol. Cell. Biol. **12:**4297–4304.
69. **Rosen, C. A.** 1991. Regulation of HIV gene expression by RNA-protein interactions. Trends Genet. **7:**9–14.
70. **Ryden, T. A., M. de Mars, and K. Beemon.** 1993. Mutation of the C/EBP binding sites in the Rous sarcoma virus long terminal repeat and *gag* enhancers. J. Virol. **67:**2862–2870.
71. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1982. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
72. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
73. **Schirm, S., J. Jiricny, and W. Schaffner.** 1987. The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. Genes Dev. **1:**65–74.
74. **Seth, A., R. Ascione, R. J. Fisher, G. J. Mavrothalassitis, N. K. Bhat, and T. S. Papas.** 1992. The *ets* gene family. Cell Growth Differ. **3:**327–334.
75. **Shioda, T., J. A. Levy, and C. Cheng-Mayer.** 1991. Macrophage and T-cell line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature (London) **349:**167–169.

76. **Soudeyns, H., R. Geleziunas, G. Shyamala, J. Hiscott, and M. A. Wainberg.** 1993. Identification of a novel glucocorticoid response element within the genome of the human immunodeficiency virus type 1. Virology **194:**758–768.
77. **Su, W., S. Jackson, R. Tjian, and H. Echols.** 1991. DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. **5:**820–826.
78. **Tesmer, V. M., A. Rajadhyaksha, J. Babin, and M. Bina.** 1993. NF-IL6-mediated transcriptional activation of the long terminal repeat of the human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **90:**7298–7302.
79. **Vaishnav, Y. N., and F. Wong-Staal.** 1991. The biochemistry of AIDS. Annu. Rev. Biochem. **60:**577–630.
80. **van de Wetering, M., M. Oosterwegel, D. Dooijes, and H. Clevers.** 1991. Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J. **10:**123–132.
81. **Van Lint, C., A. Burny, and E. Verdin.** 1991. The intragenic enhancer of human immunodeficiency virus type 1 contains functional AP-1 binding sites. J. Virol. **65:**7066–7072.
82. **Verdin, E.** 1991. DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1. J. Virol. **65:** 6790–6799.
83. **Verdin, E., N. Becker, F. Bex, L. Droogmans, and A. Burny.** 1990. Identification and characterization of an enhancer in the coding region of the genome of human immunodeficiency virus type 1. Proc. Natl. Acad. Sci. USA **87:**4874–4878.
84. **Verdin, E., P. Paras, Jr., and C. Van Lint.** 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. **12:**3249–3259.
85. **Wang, C.-Y., B. Petryniak, I.-C. Ho, C. B. Thompson, and J. M. Leiden.** 1992. Evolutionarily conserved ets family members display distinct DNA binding specificities. J. Exp. Med. **175:**1391–1399.
86. **Wasylyk, B., S. L. Hahn, and A. Giovane.** 1993. The Ets family of transcription factors. Eur. J. Biochem. **211:**7–18.
87. **Waterman, M. L., and K. A. Jones.** 1990. Purification of TCF-1α, a T-cell-specific transcription factor that activates the T-cell receptor Cα gene enhancer in a context-dependent manner. New Biol. **2:**621–636.
88. **Westervelt, P., H. E. Gendelman, and L. Ratner.** 1991. Identification of a determinant within the human immunodeficiency virus type 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. USA **88:**3097–3101.
89. **Westervelt, P., T. Henkel, D. B. Trowbridge, J. Orenstein, J. Heuser, H. E. Gendelman, and L. Ratner.** 1992. Dual regulation of silent and productive infection in monocytes by distinct human immunodeficiency virus type 1 determinants. J. Virol. **66:**3925–3931.
90. **Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66:**2577–2582.
91. **Wolffe, A.** 1992. Chromatin, structure and function. Academic Press, Inc., San Diego, Calif.
92. **Wu, C.** 1989. Analysis of hypersensitive sites in chromatin. Methods Enzymol. **170:**269–289.
93. **York-Higgins, D., C. Cheng-Mayer, D. Bauer, J. A. Levy, and D. Dina.** 1990. Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. J. Virol. **64:**4016–4020.