JOURNAL OF VIROLOGY, Apr. 1992, p. 2514–2518
0022-538X/92/042514-05$02.00/0

Vol. 66, No. 4

Probing Protein-DNA Interactions at the Long Terminal Repeat of Human Immunodeficiency Virus Type 1 by In Vivo Footprinting

FRANCESCA DEMARCHI,[1] PIERLANFRANCO D'AGARO,[2] ARTURO FALASCHI,[1] AND MAURO GIACCA[1]*

International Centre for Genetic Engineering and Biotechnology, AREA Science Park, Padriciano 99, 34012 Trieste,[1] and Institute of Hygiene, University of Trieste, 34127 Trieste,[2] Italy

Received 25 October 1991/Accepted 13 December 1991

We have analyzed protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1 in a productively infected T-cell line by in vivo dimethyl sulfate footprinting. Major footprints are evident at the basal promoter and enhancer elements. In particular, proteins appear to occupy the TATA box, the Sp1 sites, and the two repeats of the enhancer region. In the negative regulatory element, protections are detected over the USF/MLTF and NFAT-1 sites. Furthermore, two previously unrecognized sites, from nucleotides −260 to −275 and from nucleotides −205 to −216, respectively, appear to be involved in protein-DNA interactions. These two sites are purine rich and share a common sequence motif.

Gene expression and viral replication of human immunodeficiency virus type 1 (HIV-1) is controlled by complex regulatory mechanisms involving transcriptional as well as posttranscriptional events in which both cellular and virus-encoded proteins take part (5, 8). In particular, transcriptional regulation of HIV-1, exerted by the 5′ long terminal repeat (LTR) of the provirus, appears to be strictly related to the establishment of, and to the escape from, latency (31).

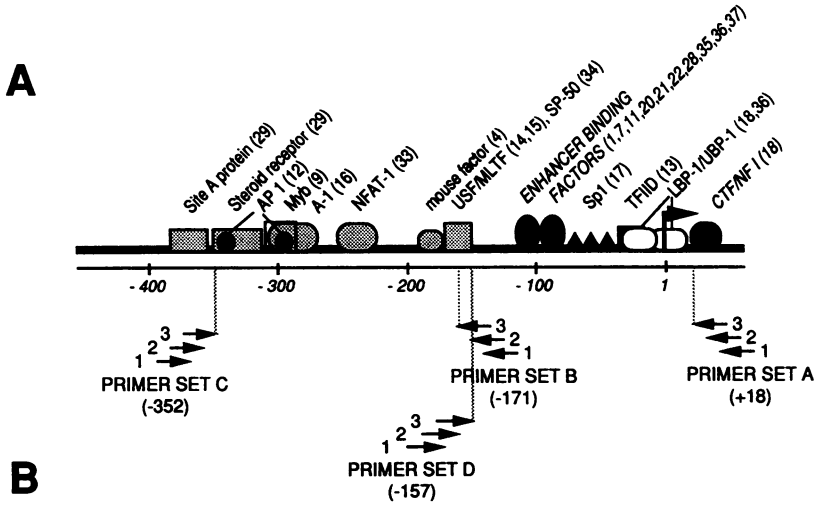
A large number of *cis*-acting viral regulatory elements has been identified in the LTR (Fig. 1 and references therein); from a functional point of view, the LTR can be divided into three main regions: the basal promoter region, the enhancer element, and the negative regulatory element (NRE). The basal promoter region, encompassing the transcription start site, includes a TATA box (13, 14), three Sp1 binding sites (17, 23), and sites for the UBP-1/LBP-1 factor (18, 19, 36); it exerts a positive basal effect on transcription. The enhancer region (containing two repeats of a sequence recognized by several proteins [1, 7, 11, 20–22, 28, 36, 37]) increases the effect of the basal region and renders the promoter inducible by different stimuli (6). The NRE, extending upstream of the enhancer site and having not precisely defined boundaries, contains binding sites for several factors with activating (9, 12), repressing (15, 24), or unknown (4, 16, 29) activity; the overall function of the NRE is to downregulate transcription (32).

Protein-DNA interactions at most of these sites have been recognized by in vitro binding studies and by in vitro transcription and transient transfection assays. However, the interaction of specific DNA-binding proteins with their corresponding sites may not always occur in the living cell (2, 27), where it is affected by several other parameters, such as chromatin structure, interactions with other proteins, conformational changes, and epigenetic modifications of DNA such as methylation.

To identify the DNA sites actually occupied in vivo in a chronically infected HIV-1-producing cell line, we analyzed H9 cells infected with HIV-$1_{HTLVIIIB}$ by in vivo dimethyl sulfate (DMS) footprinting, using the ligation-mediated polymerase chain reaction (PCR) technique (27).

HIV-$1_{HTLVIIIB}$-infected H9 cells were grown in RPMI 1640 supplemented with 10% fetal calf serum. Ten millions of exponentially growing cells were treated by adding to the medium freshly prepared DMS (0.1% final concentration) for 5 min. Under these conditions, partial methylation of guanine residues at position N-7 is obtained (10). The reaction was stopped by the addition of 10 volumes of ice-cold phosphate-buffered saline (PBS); cells were collected by centrifugation and washed twice with PBS. Nuclei were isolated by lysis with 0.1% Nonidet P-40, and DNA was extracted according to standard procedures. In parallel, 200 μg of naked genomic DNA (1 μg/μl in water) extracted from untreated cells was reacted in vitro with 1 μl of DMS for 30 s at room temperature. The reaction was stopped by the addition of β-mercaptoethanol (0.2 M final concentration), and DNA was precipitated twice with sodium acetate. G+A and C+T reactions were performed on naked genomic DNA according to the sequencing protocol of Maxam and Gilbert (25). Restriction enzyme digestion of the DNA preparations, strand scission with piperidine, and precipitation were performed according to Pfeifer and Riggs (30). Ligation-mediated PCR was performed essentially as described by Mueller and Wold (27) with four sets of oligonucleotides able to identify both coding and noncoding strand of the U3 region of HIV-1 LTR (Fig. 1). For each set, primer 1 was annealed to DNA and extended with Sequenase to generate molecules that have blunt ends on one side. A common linker (same as described by Mueller and Wold [27]) was ligated to the blunt ends, and PCR of the ligated fragments was performed by using the longer oligonucleotide of the linker (linker primer

* Corresponding author.



PRIMER SET A

1. 5'-TATTGAGGCTTAAGCAGTGGGTTCC-3'
2. 5'-CCCTAGTTAGCCAGAGAGCTCCCAG-3'
3. 5'-GCCAGAGAGCTCCCAGGCTCAGATCTG-3'

PRIMER SET C

1. 5'-AGGCTACTTCCCTGATTAGCAGAAC-3'
2. 5'-TAGCAGAACTACACACCAGGGCCAG-3'
3. 5'-TAGCAGAACTACACACCAGGGCCAGGG-3'

PRIMER SET B

1. 5'-AGCAAGCTCGATGTCAGCAGTTCTT-3'
2. 5'-GAAGTACTCCGGATGCAGCTCTCGG-3'
3. 5'-GGATGCAGCTCTCGGGCCACGTGATGA-3'

PRIMER SET D

1. 5'-AGTGTTAGAGTGGAGGTTTGACAGC-3'
2. 5'-GCCGCCTAGCATTTCATCACGTGGC-3'
3. 5'-CCGCCTAGCATTTCATCACGTGGCCCG-3'

FIG. 1. Schematic representation of the HIV-1 LTR and sequences of the oligonucleotide primers used for this study. (A) Representation of the LTR and of the nuclear proteins known to interact by in vitro experiments. References are indicated in parentheses. The scale indicates nucleotides upstream of the transcription start site, which is shown by an arrow. On the lower part, the locations and orientations of the four oligonucleotide sets used are shown; the numbers in parentheses refer to the positions of the 3' nucleotide of primer 3 of each set. (B) Sequences of the four oligonucleotide sets.

[27]) and primer 2 of each set. After 18 cycles, the amplification products were labeled by nine cycles of primer extension with ^{32}P-end-labeled primer 3 and visualized by sequencing gel electrophoresis. To overcome problems due to high G+C content, the nucleotide analog 7-deaza-2'-GTP was included in the primer extension and the PCR reaction steps (26).

The patterns of DMS methylation of the upper, coding strand (obtained by using primer sets A and B) and lower, noncoding strand (primer sets C and D) are shown in Fig. 2 and 3, respectively, and the results are summarized along with the LTR sequence in Fig. 4. The guanines with altered DMS sensitivity indicated in the figures are those which have been reproducibly detected in at least three independent experiments. In Fig. 2 to 4, the regions of the LTR recognized as binding sites for nuclear proteins by in vitro experiments are also shown.

In HIV-$1_{HTLVIIIB}$-infected H9 cells, supporting a highly productive viral infection, major footprints appear at the basal promoter and enhancer elements; in particular, proteins appear to be located over the TATA box, the three Sp1 sites, and the two repeats of the enhancer region (primer sets A and D). Most of the bases protected or hypersensitive to DMS methylation in this region correspond to those detected by in vitro methylation interference studies (18, 20). Interestingly, no sites in close proximity to or downstream of transcription start site appear to contact specific proteins by this technique. In the NRE, protections appear over the USF/MLTF site (15) on both strands (primer sets A and C) and over the binding site for NFAT-1 (33), mainly on the coding strand (primer set B); in the same region, the two adenines at positions −247 and −249 appear clearly hypersensitive. Although the piperidine treatment used mainly favors cleavage at alkylated guanines, it may also reveal, with reduced sensitivity, adenine residues that are particularly reactive with DMS (27). Few protections are also visible in the region from nucleotides −285 to −315 (primer sets B and C), where the binding sites for several proteins, including the AP-1 factor (12) and a factor named A1 (16), are located and where purified c-*myb* protein was shown to interact in vitro (9). Finally, two protected guanines are evident at positions −331 and −338 on the noncoding strand (primer set C), in a region containing a palindromic sequence with homology to the steroid/thyroid hormone response element (29).

Very interestingly, two novel sites, located between nucleotides −260 and −275 and nucleotides −204 and −216, respectively, appear to be involved in protein-DNA interactions in vivo (primer sets B and C). Both sites are purine rich and share the common sequence 5'-GGAGAGA-3' (positions −269 and −216, respectively, on the coding strand), with the first two guanines protected from DMS methylation. Adenines at position −274, −262, and −260, surrounding the consensus sequence of the upstream site, are hypersensitive.

Finally, the guanine at position −132 on the coding strand (primer set A) is clearly hypersensitive, possibly as a consequence of interactions of factors bound on the LTR both upstream and downstream of this location.

Together, these data suggest that in a highly productive cell line the enhancer and Sp1 sites on the LTRs are fully

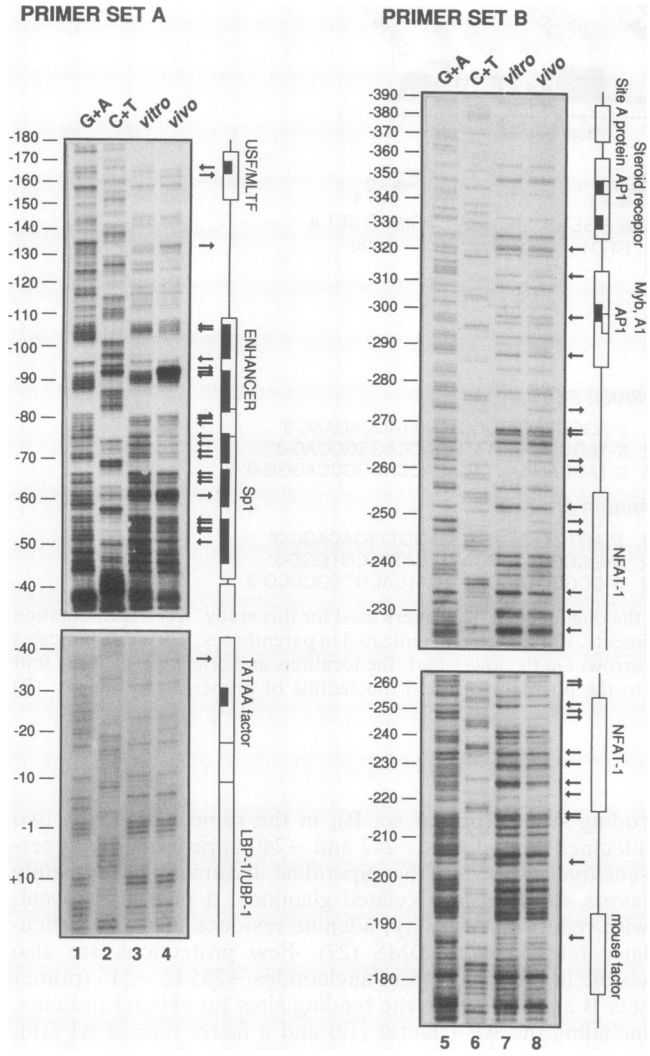


FIG. 2. In vivo DMS footprinting of the coding strand of the HIV-1 LTR. Lanes were obtained by amplification with primer sets A and B, as indicated. Lanes 1, 2, 5, and 6 are sequencing controls (G+A, T+C). Naked DNA controls (in vitro-treated DNA from HIV-infected H9 cells) are in lanes 3 and 7. In vivo DMS-treated samples from the same cells are in lanes 4 and 8. The locations of the bands with respect to the transcription start site are indicated at the left. Protected guanines are shown by arrows pointing leftward; hypersensitive purines are indicated by arrows pointing rightward. To allow easy location of the purines with altered sensitivity, on the right sides of the autoradiograms the regions protected by in vitro footprinting experiments reported in the literature are indicated by open boxes; black boxes represent the locations of the consensus sequences of the factor binding sites.

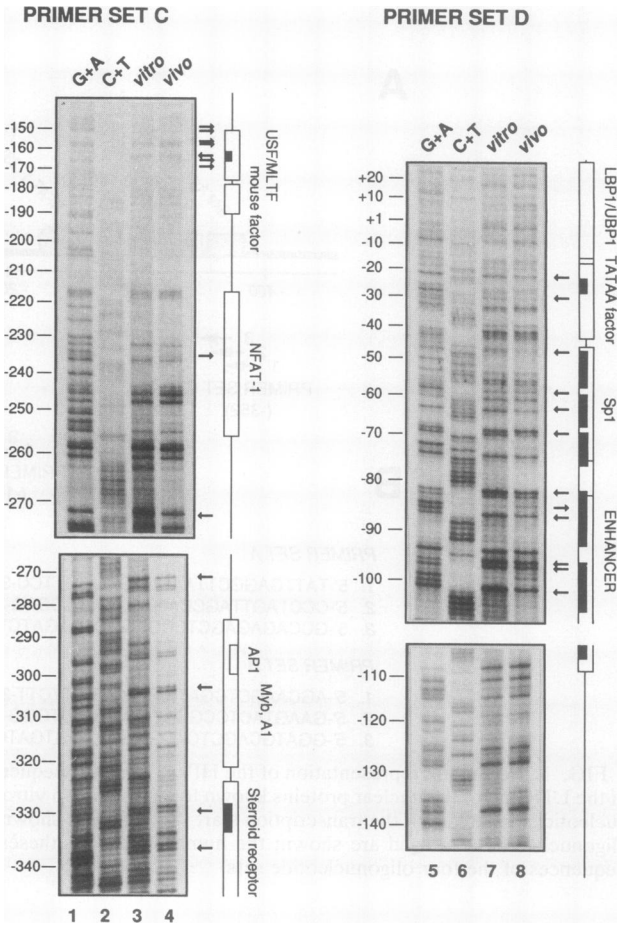


FIG. 3. In vivo DMS footprinting of the noncoding strand of the HIV-1 LTR. Lanes were obtained by amplification with primer sets C and D, as indicated. Lanes 1, 2, 5, and 6 are sequencing controls (G+A, T+C). Naked DNA controls (in vitro-treated DNA from HIV-infected H9 cells) are in lanes 3 and 7. In vivo DMS-treated samples from the same cells are in lanes 4 and 8. Symbols are as in Fig. 2.

engaged in protein-DNA interactions, as expected from the well-demonstrated role of these regions in the promotion of transcription (6, 17). Somewhat surprisingly, however, clear interactions are also present in the NRE for the USF/MLTF and NFAT-1 sites, which were shown by two independent groups to behave as downregulators of transcription (15, 24). One possible explanation is that USF/MLTF and NFAT-1, despite their residing on DNA, might not work as repressors in chronically infected H9 cells because they are inactive or because they do not interact with ancillary factors required for function. Alternatively, the NFAT-1 and USF/MLTF sites could be occupied by different proteins that mask the repressor binding sites and prevent negative regulation.

Altered DMS sensitivity at positions −260 to −275 and −204 to −216 suggests that these sites also are occupied by proteins in vivo. Both of these sites contain a purine-rich tract very similar to a sequence of a protein binding site present in the fibronectin promoter (from −414 to −429), in a region possibly involved in downregulation of transcription (3). The biological significance of these interactions is currently under investigation.

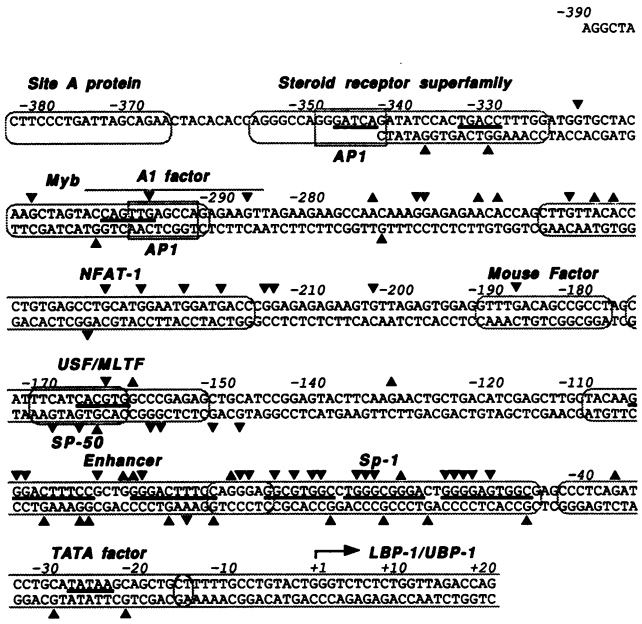


FIG. 4. Summary of the in vivo footprinting data. Protection of specific guanines on the upper or lower strand is indicated by arrows pointing inward. Hyperreactive sites are marked by arrows pointing outward. Open boxes indicate locations of the regions protected by in vitro footprinting experiments reported in the literature (see Fig. 1 for references); consensus sequences are underlined. The region bound by the A1 factor (16), for which no precise boundaries were defined, is indicated by a straight line. The portion of the LTR sequence reported corresponds only to that analyzed in this study (from nucleotides −390 to +20 on the coding strand and from nucleotides −343 to +20 on the noncoding strand).

The results described here represent a reliable, static picture of the contact points of nuclear factors residing on DNA in chronically HIV-$1_{HTLVIIIB}$-infected H9 cells. Studies are under way to extend this type of analysis to other HIV-1-infected cell lines suitable for induction and repression by chemical and biochemical agents. By this means, we should also be able to reconstruct a dynamic picture of the traffic of factors on the LTR and ultimately to contribute to unravelling the mechanism of viral latency and reactivation.

This work was partially supported by grant 6206-042 of the AIDS program of the Istituto Superiore di Sanità (Rome).

REFERENCES

1. **Baldwin, A. S. J., and P. A. Sharp.** 1988. Two transcriptions factors, NF-κB and H2TF1, interact with a single regulatory sequence in the class I major histocompatibility complex promoter. Proc. Natl. Acad. Sci. USA **85:**723–727.
2. **Becker, P. B., S. Ruppert, and G. Schutz.** 1987. Genomic footprinting reveals cell-type specific DNA binding of ubiquitous factors. Cell **51:**435–445.
3. **Bernath, V. A., A. F. Muro, A. D. Vitullo, M. A. Bley, J. L. Baranao, and A. R. Kornblihtt.** 1990. Cyclic AMP inhibits fibronectin gene expression in a newly developed granulosa cell line by a mechanism that suppresses c-AMP-responsive element-dependent transcriptional activation. J. Biol. Chem. **265:** 18219–18226.
4. **Calvert, I., Z. Peng, H. Kung, and Raziuddin.** 1991. Cloning and characterization of a novel sequence-specific DNA-binding protein recognizing the negative regulatory element (NRE) region of the HIV-1 long terminal repeat. Gene **101:**171–176.
5. **Cann, A. J., and J. Karn.** 1989. Molecular biology of HIV: new insights into the virus life-cycle. AIDS **3**(Suppl. 1)**:**S19–S34.
6. **Chinnadurai, G.** 1991. Modulation of HIV-enhancer activity by heterologous agents: a minireview. Gene **101:**165–170.
7. **Clark, L., J. R. Matthews, and R. T. Hay.** 1990. Interaction of enhancer-binding protein EBP1 (NF-κB) with the human immunodeficiency virus type 1 enhancer. J. Virol. **64:**1335–1344.
8. **Cullen, B. R., and W. C. Greene.** 1989. Regulatory pathways governing HIV-1 replication. Cell **58:**423–426.
9. **Dasgupta, P., P. Saikumar, C. D. Reddy, and E. P. Reddy.** 1990. Myb protein binds to human immunodeficiency virus 1 long terminal repeat (LTR) sequences and transactivates LTR-mediated transcription. Proc. Natl. Acad. Sci. USA **87:**8090–8094.
10. **Ephrussi, A., G. M. Church, S. Tonegawa, and W. Gilbert.** 1985. B lineage-specific interactions of an immunoglobulin enhancer with cellular factors in vivo. Science **227:**134–140.
11. **Franza, B. R., S. F. Josephs, M. Z. Gilman, W. Ryan, and B. Clarkson.** 1987. Characterization of cellular proteins recognizing the HIV enhancer using a microscale DNA-affinity precipitation assay. Nature (London) **330:**391–395.
12. **Franza, B. R., Jr., F. J. Rauscher III, S. F. Josephs, and T. Curran.** 1988. The fos complex and fos-related antigens recognize sequence elements that contain AP1 binding sites. Science **239:**1150–1153.
13. **Garcia, J. A., D. Harrich, E. Soultanakis, F. Wu, R. Mitsuyasu, and R. B. Gaynor.** 1989. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. EMBO J. **8:**765–778.
14. **Garcia, J. A., F. K. Wu, R. Mitsuyasu, and R. B. Gaynor.** 1987. Interactions of cellular proteins involved in the transcriptional regulation of the human immunodeficiency virus. EMBO J. **6:**3761–3770.
15. **Giacca, M., M. I. Gutierrez, S. Menzo, F. d'Adda di Fagagna, and A. Falaschi.** 1992. A human binding site for transcription factor USF/MLTF mimics the negative regulatory element of human immunodeficiency virus type 1. Virology **186:**133–147.
16. **Guy, B., R. B. Acres, M. P. Kieny, and J. P. Lecocq.** 1990. DNA binding factors that bind to the negative regulatory element of the human immunodeficiency virus-1; regulation by *nef*. J. Acquired Immune Defic. Syndr. **3:**797–809.
17. **Jones, K., J. Kadonaga, P. Luciw, and R. Tjian.** 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor SP1. Science **232:**755–759.
18. **Jones, K., P. Luciw, and N. Duchange.** 1988. Structural arrangement of transcription control domain within the 5′-untranslated leader regions of HIV-1 and HIV-2 promoters. Genes Dev. **2:**1101–1114.
19. **Kato, H., M. Horikoshi, and R. G. Roeder.** 1991. Repression of HIV-1 transcription by a cellular protein. Science **251:**1476–1479.
20. **Kawakami, K., C. Scheidereit, and R. G. Roeder.** 1988. Identification and purification of a human immunoglobulin-enhancer-binding protein (NF-κB) that activates transcription from a human immunodeficiency virus type 1 promoter *in vitro*. Proc. Natl. Acad. Sci. USA **85:**4700–4704.
21. **Korner, M., A. H. Bellan, A. T. Brini, and W. L. Farrar.** 1990. Characterization of the human immunodeficiency virus type 1 enhancer-binding proteins from the human T-cell line Jurkat. Biochem. J. **265:**547–554.
22. **Lenardo, M. J., A. Kuang, A. Giffird, and D. Baltimore.** 1988. NF-κB protein purification from bovine spleen: nucleotide stimulation and binding site specificity. Proc. Natl. Acad. Sci. USA **85:**8825–8829.
23. **Leonard, J., C. Parrot, A. J. Buckler-White, W. Turner, E. K. Ross, M. A. Martin, and A. B. Rabson.** 1989. The NF-κB binding sites in the human immunodeficiency virus type 1 long terminal repeat are not required for virus infectivity. J. Virol. **63:**4919–4924.
24. **Lu, Y., N. Touzjian, M. Stenzel, T. Dorfman, J. G. Sodroski, and W. A. Haseltine.** 1990. Identification of *cis*-acting repressive sequences within the negative regulatory element of human immunodeficiency virus type 1. J. Virol. **64:**5226–5229.
25. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.

26. **McConlogue, L., M. A. D. Brow, and A. Innis.** 1988. Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine. Nucleic Acids Res. **16**:9869.
27. **Mueller, P. R., and B. Wold.** 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. Science **246**:780–786.
28. **Nabel, G., and D. Baltimore.** 1987. An inducible factor activates expression of human immunodeficiency virus in T cells. Nature (London) **326**:711–713.
29. **Orchard, K., N. Perkins, C. Chapman, J. Harris, V. Emery, G. Goodwin, D. Latchman, and M. Collins.** 1990. A novel T-cell protein which recognizes a palindromic sequence in the immunodeficiency virus long terminal repeat. J. Virol. **64**:3234–3239.
30. **Pfeifer, G. P., and A. Riggs.** 1991. Chromatin differences between active and inactive X chromosome revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev. **5**:1102–1113.
31. **Pomerantz, R. J., D. Trono, M. B. Feinberg, and D. Baltimore.** 1990. Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: a molecular model for latency. Cell **61**:1271–1276.
32. **Rosen, C. A., J. G. Sodrosky, and W. A. Haseltine.** 1985. The location of *cis*-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell **41**:813–823.
33. **Shaw, J. P., P. J. Utz, D. B. Durand, J. J. Toole, E. A. Emmel, and G. R. Crabtree.** 1988. Identification of a putative regulator of early T cell activation genes. Science **241**:202–205.
34. **Smith, M. R., and W. C. Greene.** 1989. The same 50 kDa cellular protein binds to the negative regulatory elements of the interleukin-2 receptor α-chain gene and the HIV-1 LTR. Proc. Natl. Acad. Sci. USA **86**:8526–8530.
35. **Wu, F., J. Garcia, R. Mitsuyasu, and R. Gaynor.** 1988. Alterations in binding characteristics of human immunodeficiency virus enhancer factor. J. Virol. **62**:218–225.
36. **Wu, F. K., J. A. Garcia, D. Harrich, and R. B. Gaynor.** 1988. Purification of the human immunodeficiency virus type 1 enhancer and TAR binding proteins EBP-1 and UBP-1. EMBO J. **7**:2117–2129.
37. **Yano, O., J. Kanellopoulos, M. Kieran, O. Le Bail, A. Israël, and P. Kourilsky.** 1987. Purification of KBF1, a common factor binding to both H-2 and β2-microglobulin enhancers. EMBO J. **6**:3317–3324.