JOURNAL OF VIROLOGY, May 1995, p. 2765–2775 Vol. 69, No. 5
0022-538X/95/$04.00+0

Molecular and Functional Interactions of Transcription Factor USF with the Long Terminal Repeat of Human Immunodeficiency Virus Type 1

FABRIZIO D'ADDA DI FAGAGNA, GIUSEPPE MARZIO, MARIA INES GUTIERREZ, LI YA KANG,† ARTURO FALASCHI, AND MAURO GIACCA*

International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

Received 12 October 1994/Accepted 24 January 1995

The human transcription factor USF, purified from HeLa cells, and its recombinant 43-kDa component bind to the long terminal repeat (LTR) of human immunodeficiency virus type 1. The proteins footprint over nucleotides from position −173 to −157 upstream of the transcription start site, generating strong DNase I hypersensitivity sites at the 3′ sides on both strands. As detected by methylation protection studies, the factor forms symmetric contacts with the guanines of the palindromic CACGTG core of the recognized sequence. Its binding ability is abolished by the mutation of this core sequence and is strongly reduced by the cytosine methylation of the central CpG dinucleotide. Upon binding, both recombinant and purified USFs bend the LTR DNA template. The role of USF in the control of transcription initiation from the LTR was tested by in vitro transcription assays. Upon addition of the protein, transcription from constructs containing an intact binding site is increased, while the responsiveness in constructs with a mutated sequence is abolished. Furthermore, the addition of a decoy plasmid which contains multiple repeats of the target sequence results in downregulation of transcription from the LTR. These results suggest that USF is a positive regulator of LTR-mediated transcriptional activation.

Like that of all retroviruses, the rate of transcription of the integrated proviral form of human immunodeficiency virus type 1 (HIV-1) is one of the major determinants of the levels of viral gene expression and replication. A single viral transcript originates from the transcription start site at the long terminal repeat (LTR) at the 5′ end of the provirus and finishes at the polyadenylation site at the 3′ LTR. From this genome-length transcript, which corresponds to the viral RNA eventually to be encapsidated, different gene products arise through the extensive use of differential processing events, translation frameshifts, and proteolytic cleavage of precursor polypeptides. Several groups have shown that although viral replication can be detected in all clinical stages of the infection (40, 46, 50), the levels of viral expression correlate with disease progression (2, 57). For this reason, it appears that the study of the molecular mechanisms controlling the rate of transcription initiation could offer useful insights into the mechanisms of disease development.

The regulation of HIV-1 transcription initiation is achieved through the recognition of the 5′ LTR by human nuclear factors and through their interactions with the basal transcriptional machinery. In this respect, the regulation of transcription of the integrated provirus does not differ from that of any other cellular gene. Only one viral protein (the product of the *tat* gene) cooperates with this machinery to increase the rates of transcription initiation and elongation.

A number of reports indicate that the U3 and R regions of the LTRs (~550 bp) contain the target sites for several nuclear factors and that many of these factors can bind to the LTR in vitro (for a review, see reference 44). With the use of an in vivo footprinting technique with infected cell lines, we have shown that most of the LTR regions are indeed engaged in protein-DNA interactions also in vivo, although the pattern of these interactions is different in different cell lines (15, 16). All of the protein binding sites of the LTR are recognized by nuclear proteins physiologically controlling the expression of a variety of cellular genes, indicating that the LTR represents a very interesting biological example of evolutionary tinkering. The reasons for the need for such a complex regulation in the HIV-1 life cycle still remain to be fully understood.

We have previously reported that a sequence of the LTR centered at position −164 upstream of the transcription start site contains the binding site for a constitutive binding activity of lymphoid and nonlymphoid cells (25). This region, which is located at the 3′ end of the so-called negative regulatory element of the LTR (55), contains the hexanucleotide CAC(A/G)TG (E box), which is the consensus target sequence of proteins of the B class of the basic–helix-loop-helix–leucine zipper (b-HLH-Zip) family (14). Members of this family are the c-Myc (43), Max (8), Mad (1), Mxi1 (72), USF (27, 65), TFE3 (3), and TFEB (11) proteins. All of these factors are characterized by the presence of a specific DNA binding domain (basic region) and of two motifs involved in the formation of homo- and heteromultimers (HLH and leucine zipper domains) (20, 43). Binding sites for b-HLH-Zip proteins and cognate binding activities are very conserved through evolution (24, 25), and they control the expression of several unrelated genes. For example, for the human genome, there have been descriptions of binding sites in the globin locus control region (10); in the promoter regions of the L-type pyruvate kinase gene (68), of the heme oxygenase 1 gene (58), of the class I alcohol dehydrogenase gene (18), of the type 1 plasminogen activator inhibitor gene (54), of the human growth hormone gene (48), of the N-ras gene (67), and of the topoisomerase I gene (29); in the enhancer of the insulin gene (53); and in a region which, as we have recently demonstrated, contains an origin of DNA

* Corresponding author. Mailing address: ICGEB, Padriciano, 99, 34012 Trieste, Italy. Phone: 39-40-3757.324. Fax: 39-40-226555. Electronic mail address: giacca@icgeb.trieste.it.

† Present address: Walther Oncology Center, Indiana University School of Medicine, Indianapolis, Ind.

replication (6, 13, 26). Additionally, it recently has been shown that the E box of the LTR of HIV-1 is also essential for the transcriptional control of the negative strand of the provirus exerted by the LTR at the 3′ end (41).

Among the members of the b-HLH-Zip family, transcription factor USF (also known as MLTF [12] or UEF [42]) was originally identified as a human cellular factor which binds to the upstream sequence of the adenovirus major late promoter (MLP) at position −63 to −52 upstream of the transcription start site and stimulates transcription in vitro (60). USF is the major binding activity detected by in vitro assays with crude nuclear extracts from several cell types and species, including T cells and monocytes (16); after extensive purification, it appears to be composed of two polypeptides with apparent molecular masses of 43 and 44 kDa (13, 61), both of which are capable of binding independently to the DNA recognition site, even if they preferentially bind as heterodimers (20). The 43-kDa form of USF (USF^{43}) was cloned and shown to belong to the b-HLH-Zip family (27). More recently, the full-length RNA of the 44-kDa form was also isolated, and it showed a striking similarity to USF^{43} (65).

In this paper, we report the results of our studies on the molecular features of the in vitro interactions of both purified human USF and recombinant USF^{43} ($rUSF^{43}$) with the E box of the HIV-1 LTR, as well as the effects of this interaction on the regulation of transcription.

MATERIALS AND METHODS

Purification of USF from HeLa cells. The procedure for the purification of USF from HeLa cells has already been reported (13). Briefly, a nuclear extract was precipitated by 35% ammonium sulfate; the precipitated protein fraction was subsequently purified through BioRex 70, HiLoad S-Sepharose, and Mono Q columns. The active fractions from the Mono Q column were loaded on a specific DNA-Sepharose affinity column containing ligated concatemers of the E box site. The purified binding activity consists of a protein doublet with molecular masses of 43 and 44 kDa. The binding activity was monitored throughout the purification by gel retardation and Southwestern (DNA-protein) assays.

Cloning, expression, and purification of $rUSF^{43}$. The plasmid pGST-USF^{43} was constructed by cloning the coding region of USF^{43} in the commercial vector pGEX2T (Pharmacia, Uppsala, Sweden). This plasmid expresses the glutathione-S-transferase (GST) gene in *Escherichia coli*, under the control of a promoter inducible by isopropyl-galactopyranoside (IPTG). The coding sequence of USF^{43} was rescued by PCR amplification from the plasmid pBSA1MLTF (a kind gift of Giuseppe Biamonti), which contains the coding sequence of USF^{43} under the control of the human A1 gene promoter (unpublished). The primers for amplification (RUSF1 [5′-CGAGGATCCAAGGGGCAGCAGAAAACA-3′] and RUSF2 [5′-GCTGAATTCCTTAGTTGCTGTCATTCTTGATGAG-3′]) were designed in order to generate a DNA fragment containing the restriction sites for *Eco*RI and *Bam*HI at the two extremities, allowing oriented cloning in the vector. Amplification was carried out in 50 μl of a solution containing 10 mM Tris (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM each deoxynucleoside triphosphate (dNTP), 1 μM the two primers, 1 ng of template plasmid DNA, and 2.5 U of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, Conn.), with 35 repetitions of a cycle of 45 s at 94°C, 45 s at 52°C, and 45 s at 72°C in a Perkin-Elmer 480 Thermal Cycler. The sequence of the plasmid insert was verified by DNA sequencing.

A colony of the SF8 strain of *E. coli* transformed by pGST-USF^{43} was grown overnight in 100 ml of Terrific Broth containing 50 μg of ampicillin per ml at 30°C. The culture was diluted by the addition of 900 ml of fresh medium and was allowed to grow until an optical density of 0.6 to 0.8 at 600 nm was reached; protein expression was then induced by the addition of IPTG (Sigma, St. Louis, Mo.; 1 mM final concentration). After an additional 3 to 5 h of incubation, the culture was centrifuged at 5,000 × *g* at 4°C, and the bacterial pellet was resuspended in 10 ml of cold phosphate-buffered saline (PBS) containing 4 mM dithiothreitol (DTT). Cells were sonicated in ice by three pulses of 20 s each. After centrifugation of the lysate, the supernatant was mixed with 1 ml of a 50% (vol/vol) slurry of glutathione-cross-linked agarose beads (Sigma). The fusion protein was allowed to bind to the beads at 4°C on a rotating wheel for 1 h. The suspension was then loaded on an empty plastic column (Bio-Rad, Richmond, Calif.), letting the unbound proteins pass through, and the beads were subsequently washed with 50 ml of PBS containing 4 mM DTT. Finally, $rUSF^{43}$ was eluted in 1 ml of 100 mM Tris containing 4 mM DTT and 20 mM free glutathione (Sigma). With this procedure, 2 to 3 mg of protein per liter of medium was usually obtained. The purity and integrity of the protein were checked by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining.

Western blotting (immunoblotting). Fifteen nanograms of recombinant protein was resolved on an SDS–12% polyacrylamide gel and transferred to a nitrocellulose filter by electroblotting. The filter was then incubated in 10% milk–TBS buffer (10% [wt/vol] nonfat dried milk in 125 mM NaCl–10 mM Tris, pH 7.4) for 1 h at 37°C. Incubation with antibodies against full-length $rUSF^{43}$ (kindly donated by R. G. Roeder) was performed in 5% milk–TBS for 1 h at room temperature. The filter was then washed in TBS–0.1% Tween 20 and incubated in 5% milk–TBS with alkaline phosphatase-conjugated swine anti-rabbit immunoglobulins for 1 h at room temperature. After several washes, bound antibodies were revealed with BCIP-NBT (5-bromo-4-chloro-3-indolylphosphate toluidinium–nitroblue tetrazolium) color development solution (Bio-Rad).

Gel retardation and competition assays. All of the oligonucleotides used in this work were synthesized by the International Centre for Genetic Engineering and Biotechnology Oligonucleotide Synthesis Service on an Applied Biosystems 380B synthesizer by using phosphoramidite chemistry. The sequences of the double-stranded oligonucleotides oligo B48BS, oligo AdMLP, and oligo HIV (25) and oligo GAL2, oligo Met-I, and oligo Met-III (24) have already been described. Oligo Sp1 contains the sequence of the HIV-1 LTR from nucleotide −69 to −46 upstream of the transcription start site, which contains two binding sites for transcription factor Sp1. Oligo MUT is a derivative of oligo HIV in which the core E box sequence (CACGTG) was mutated to CATATG.

The synthesized oligonucleotides were resolved by PAGE, eluted from the gels, purified, annealed with the complementary strand, and end labelled with [γ-^{32}P]ATP (Amersham, Little Chalfont, United Kingdom; 3,000 Ci/mmol, 10 mCi/ml) and T4 polynucleotide kinase.

Gel retardation assays were carried out by the incubation of end-labelled DNA probes (10^4 cpm) with 1 to 5 ng of purified $rUSF^{43}$ or 2 μl of purified HeLa USF (Mono Q fraction [13]) and 0.25 μg of poly[d(A-T)] · poly[d(A-T)] or 1.5 μg of poly[d(I-C)] · poly[d(I-C)], respectively, in binding buffer (20 mM HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid] [pH 7.3], 50 mM NaCl, 4 mM DTT, 0.2 mM EDTA, 5% glycerol) in a 10-μl final volume. The addition of 1 μl of heat-treated (95°C for 5 min) normal human serum to the binding reaction mixture was used to stabilize the protein-DNA complex. After 20 min of incubation at 30°C, samples were resolved by 5% PAGE in 0.5× Tris-borate-EDTA; the gel was then dried and exposed. Under these conditions, 1 ng of $rUSF^{43}$ is able to generate a retarded complex which can be detected after an overnight exposure, with an estimated molar DNA/protein ratio of 1:5.6.

Competition experiments were carried out by mixing a 6- to 150-fold excess of cold oligonucleotides with the probe before the addition of the protein. Supershifting of the retarded complex was obtained by the addition to the reaction mixture of 1 μl of a 1:10 dilution of immune serum before the addition of the protein.

Circular permutation assay. A set of six probes (B1 to B6; see Fig. 6A) for the circular permutation assay was generated by PCR amplification with the pLTRCAT plasmid (25) as a template. The locations on the LTR of the oligonucleotides utilized for the amplifications are indicated in Fig. 6A. Amplifications were carried out in 50 μl of a solution containing 10 mM Tris (pH 8.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM each dNTP, 0.1 μM both primers (one of which was previously end labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase), 1 ng of template plasmid DNA, and 2.5 U of *Taq* DNA polymerase, with 35 repetitions of a cycle of 45 s at 94°C, 45 s at 60°C, and 45 s at 72°C. PCR products were resolved on a 5% polyacrylamide native gel and eluted for a few hours at 65°C in 10 mM Tris (pH 8.0)–0.1 mM EDTA. Curve fitting was performed with the Curve Fit program on a Macintosh computer by the application of a cosine function.

DNase I footprinting. The probes for DNase I footprinting were generated by PCR amplification of the HIV-1 LTR from the plasmid pLTRCAT. For the experiments with purified USF, the primers for amplifications were FOOTP (5′-GCAAGCTTGAAGAGGCCAAT-3′) and USF1 (5′-AGCAAGCTCGAT GTCAGCAGTTCTT-3′); for the experiments with $rUSF^{43}$, the primers used were from position −256 to −220 and from position −46 to −70, respectively, relative to the LTR sequence numbering of the transcription start site. One of the two primers was end labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase before PCR amplification in order to generate an asymmetrically labelled DNA fragment.

Plasmid pLTRΔUSF was obtained by PCR amplification from plasmid pLTRCAT with primers pLTR-GlessI (5′-GCGAATTCTAGGGCGAATTGGG TACC-3′) and ΔUSF (5′-GCTCTCGGGGCGAATTCATGAAATGCTAGG CGGC-3′), the latter bearing the CACGTG sequence mutated into an *Eco*RI (GAATTC) restriction site. The amplification product was purified, cut with *Ava*I and *Kpn*I, and ligated between the *Ava*I and *Kpn*I sites of pLTRCAT, thereby substituting for the wild-type sequence.

In the experiments with USF purified from HeLa cells, about 10^5 cpm of the probe was incubated with 20 μl of purified protein (Mono Q fraction), with 2 μg of poly[d(I-C)] · poly[d(I-C)], in the same buffer as was used for the gel retardation assays. After 20 min of incubation at room temperature, the sample was mixed with an equal volume of a solution containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$. DNase I (Boehringer, Mannheim, Germany) was added to a final concentration of 6 ng/μl. After 1 min, DNase I activity was stopped by the addition

of SDS and EDTA (1% and 25 mM final concentrations, respectively). Proteins were removed by phenol extraction, and the DNA fragments were precipitated, dissolved, denatured, and loaded on an 8% polyacrylamide sequencing gel. As a control, the probe was incubated under the same conditions as described above without the addition of the protein. In this case, DNase I was used at a concentration of 2 ng/μl.

When the recombinant protein was tested, 10 ng of $rUSF^{43}$ or GST (in the control reaction) was mixed with 5×10^4 cpm of the probe under the same conditions as described above. DNase I was allowed to digest for 30 s at a final concentration of 0.5 ng/μl.

A G+A ladder of the probe, obtained by the Maxam-Gilbert chemical cleavage method (38), was loaded alongside to align the DNase I digestion products.

Methylation protection assay. Methylation protection assays with dimethyl sulfate were carried out essentially as described previously (47). Briefly, fivefold-scaled-up gel retardation assays were assembled; after electrophoresis, the gel was immersed in a 0.2% (vol/vol) dimethyl sulfate solution for 4 min, and the methylation reaction was stopped by soaking the gel in a 0.5 M β-mercaptoethanol solution. After autoradiography, the free and retarded bands were excised separately, and DNA was eluted and cleaved by piperidine treatment (38). The products were resolved on a sequencing gel alongside a G+A ladder.

In vitro transcription assays. The LTR-containing G-less plasmids for in vitro transcription assays are derivatives of the pUGL400 plasmid, which contains an *Eco*RI-*Xho*I fragment from plasmid Syn-O-TG (62), carrying a G-free cassette of ~380 bp (59), cloned between the *Eco*RI and *Sal*I sites of pUC19. The HIV-1 promoter element to be cloned upstream of the G-less cassette was obtained by PCR amplification from plasmid pLTRCAT (25), with primers pLTR-GlessI (containing an *Eco*RI restriction site at the 5′ end continuing with the HIV-1 sequence upstream of the LTR) and pLTR-GlessII (5′-CGGAGCTCAGGCA AAAAGCAGCTGCTTA-3′, containing a *Sac*I site at the 5′ end in the corresponding part of the LTR transcription start site); the amplification product obtained was cloned between the *Eco*RI and *Sac*I sites of pUGL400 to obtain plasmid pGLA. Plasmid pGLE was obtained by the same strategy, with the exception that the plasmid pLTRΔUSF was the template DNA for PCR amplification. As a consequence, the nucleotides at position −162 to −167 are GAATTC (containing an *Eco*RI site) instead of CACGTG.

Plasmid pFN2, containing the fibronectin promoter upstream of a ~200-nucleotide G-less cassette, was used as an internal control for the in vitro transcription experiments. It is a derivative of plasmid pUGL200, which was obtained by cloning a PCR amplification product between the *Eco*RI and *Hin*dIII sites of pUC19. The amplification product is derived from plasmid pUGL400 and was obtained by using the M13 universal primer and primer GL202 (5′-GGAAGCTTGGATCCCGGGATAAGATTG-3′). The latter contains at the 3′ end a region complementary to the sequence from position 187 to 108 of the ~380-nucleotide G-less cassette (59); PCR amplification was followed by digestion with *Eco*RI and *Hin*dIII. Plasmid pFN2 was obtained by cloning of a ~710-bp fragment obtained by partial digestion with *Aat*II and *Sac*I of plasmid p-220 (a kind gift of Alberto Kornblihtt), into the corresponding sites of pUGL200. The cloned fragment contains the human fibronectin promoter.

HeLa nuclear extracts for in vitro transcription assays were prepared according to the protocol of Hattori et al. (28) and modified as follows: cells were washed twice in PBS, washed once in RSB (10 mM HEPES, 10 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 0.74 mM spermidine, 1 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride, 2 μg of aprotinin [Sigma] per ml, 2 μg of leupeptin [Sigma] per ml, 5 mM bestatin [Sigma]), and resuspended in RSB; following a 20-min incubation on ice, the cells were lysed with a Dounce homogenizer.

Transcription reactions were carried out in a 25- to 30-μl final volume containing 250 ng (in the experiments monitoring the effects of $rUSF^{43}$) or 750 ng (in the competition experiments with the E box decoys) of DNA templates, 75 μg of nuclear extract, 500 μM (each) ATP and CTP, 25 μM UTP, 10 μCi of [α-^{32}P]UTP (Amersham); 3,000 Ci/mmol, 10 mCi/ml), 0.1 mM *o*-methylguanosine, 10 mM HEPES (pH 7.4), 7.5 mM $MgCl_2$, 5 U of RNase T_1 (Boehringer), 30 U of RNasin (Promega, Madison, Wis.), and 4 mM DTT. Nuclear extract was equilibrated for 15 min at 30°C with the addition of the appropriate amounts of protein or decoy template. DNA was then added and preincubated for 10 min at 30°C, and the reaction was initiated by the addition of the above-described reaction mixture. Transcription was allowed to run at 30°C for 45 min and was stopped by adding 175 μl of a solution containing 300 mM Tris (pH 7.4), 300 mM sodium acetate, 0.5% SDS, 2 mM EDTA, and 3 μg of tRNA per ml. Proteins were then extracted with 200 μl of phenol-chlorophorm-isoamilic alcohol (25:24:1), and RNA was precipitated with ethanol, dried, and run at 12 W on a denaturing 8% polyacrylamide gel in 1× Tris-borate-EDTA–0.1% SDS. Finally, the gel was dried and exposed for autoradiography.

Plasmid pUF128 (a gift of Fabio Cobianchi) is a pUC19 derivative containing 128 copies, arranged in tandem, of a 65-bp *Alu*I-*Fnu*D2 sequence from the insert of plasmid pL15 (25), containing the E box present in the human lamin B2 origin of DNA replication (7, 26).

RESULTS

Purified USF and recombinant USF^{43} bind to the LTRs of HIV-1. The presence of a human nuclear factor binding to the HIV-1 LTR region centered around nucleotide −164 upstream of the transcription start site was originally identified by DNase I footprinting with HeLa cell nuclear extracts (23). Subsequently, we have shown that the protein interacting with this sequence also binds to the upstream element of the adenovirus MLP and to several other sequences containing the hexanucleotide consensus sequence CAC(A/G)TG (25), which is the E box target of members of the B class of the b-HLH family of proteins (14).

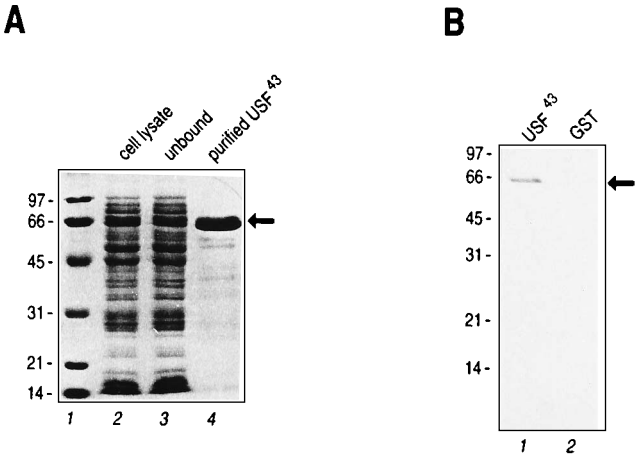

FIG. 1. Expression and purification of $rUSF^{43}$. (A) Single-step purification of USF^{43}. Protein samples were resolved on a 12% Laemmli gel and stained with Coomassie blue. Lane 1, protein molecular weight markers (in thousands); lane 2, 10 μl of the supernatant of a bacterial lysate of the SF8 *E. coli* strain transformed with pGST-USF^{43} after induction with IPTG; lane 3, 10 μl of the flowthrough fraction from the agarose-glutathione column; lane 4, 40 μl of a fraction obtained by elution of the column with free glutathione. The arrow indicates the position of the GST-USF^{43} fusion protein. (B) Western blot analysis. Lane 1, purified $rUSF^{43}$ (15 ng); lane 2, purified GST (15 ng). The arrow indicates the position of the GST-USF^{43} fusion protein. The positions of protein molecular weight markers (in thousands) are indicated.

By means of a combination of ion-exchange and sequence-specific affinity chromatography techniques (13), we have purified to homogeneity a protein complex binding to the E box. This complex is composed of two polypeptides of 42 and 44 kDa; its size, heat stability, and target DNA sequence suggest that it corresponds to the transcription factor USF, a member of the b-HLH-Zip family (27, 65). Furthermore, the 42-kDa polypeptide is recognized by antibodies raised against USF^{43} (13).

In order to study in further detail the interactions of this transcription factor with the HIV-1 LTR, we expressed and purified USF^{43} from bacteria as a recombinant protein fused to GST. On the basis of the available cDNA sequence (27), the coding region of USF^{43} was cloned in plasmid pGEX2T, at the 3′-end of the GST gene and maintaining the same open reading frame. The resulting plasmid, pGST-USF^{43}, expresses the USF^{43} protein as an extension of the C terminus of GST under the control of a promoter inducible by IPTG. In Fig. 1A the results of a single-step purification of $rUSF^{43}$ from bacteria are shown. The identity of the purified protein was further confirmed by Western blot analysis with antibodies raised against $rUSF^{43}$ (51) (Fig. 1B).

The interactions of USF purified from HeLa cells and of $rUSF^{43}$ with the E box of the HIV-1 LTR and other E boxes are shown in Fig. 2. A double-stranded oligonucleotide corresponding to the sequence from position −174 to −151 of the LTR (oligo HIV) specifically binds to purified USF in a gel retardation assay, resulting in the formation of a retarded

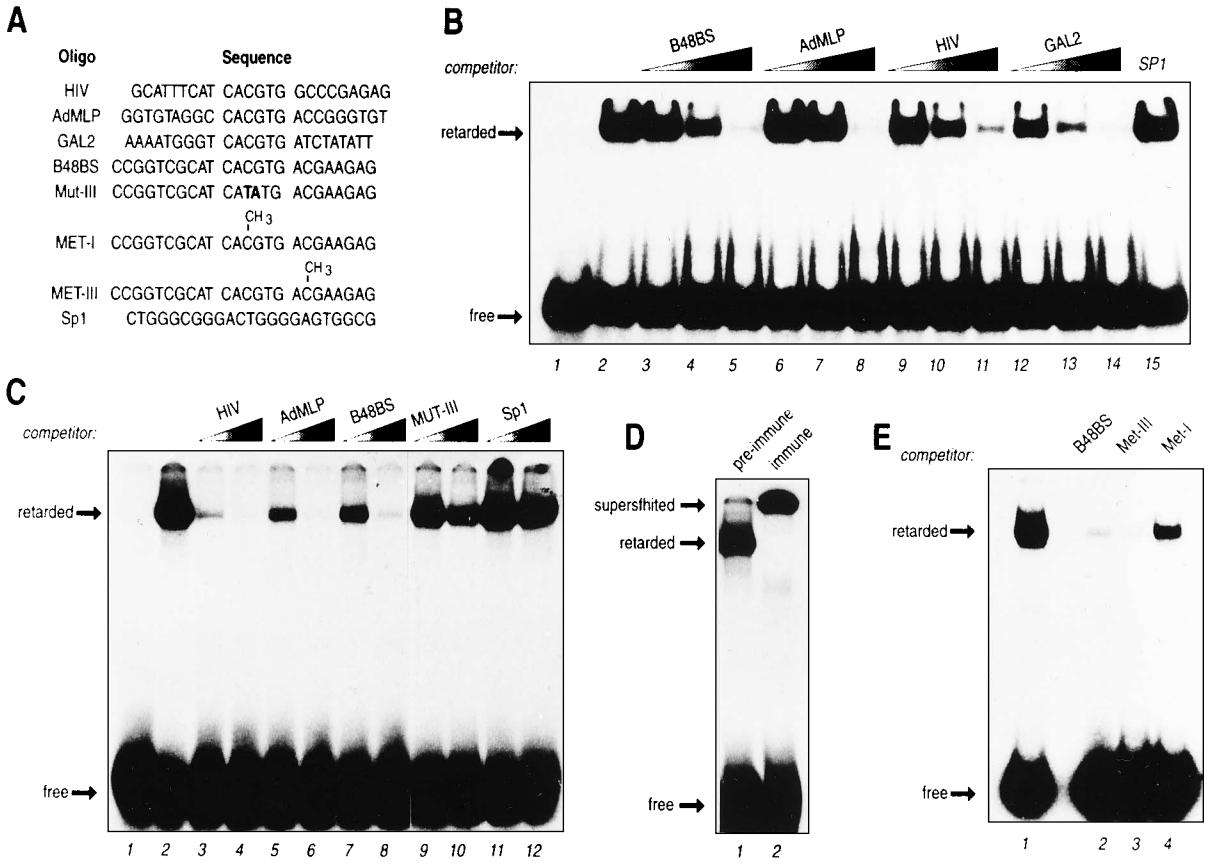

FIG. 2. Gel retardation and competition assays. (A) Sequences of the upper strands of the oligonucleotides (Oligo) used for the gel retardation and competition assays (see Materials and Methods for their locations). The E box sequence is boxed. Mutations of this sequence in oligo MUT-III are indicated by boldface. The locations of the methyl groups in oligos Met-I and Met-III are indicated. (B) Gel retardation and competition assays with purified USF and oligo HIV as a probe. For the competition experiments, increasing amounts of competitor (6-, 20-, and 60-fold molar excesses with respect to the probe [but 60-fold only for the Sp1 competitor]) were added to the binding reaction mixtures as indicated at the top. The arrows indicate the positions of the free and retarded bands. Lane 1, probe without protein. (C) Gel retardation and competition assays with rUSF^{43} and oligo HIV as a probe. Lane 1, probe without protein. Competitor molar excesses were 6- and 30-fold for each competitor. (D) Gel retardation-supershift analysis with anti-rUSF^{43} antiserum. The reaction mixture for binding between oligo HIV and rUSF^{43} was incubated with anti-rUSF^{43} antiserum (lane 2) or preimmune serum (lane 1). The arrows indicate the positions of the free, retarded, and supershifted bands. (E) Effect of cytosine methylation on rUSF^{43} binding to oligo HIV. Lanes 1 to 4, oligo HIV plus rUSF^{43}; lanes 2 to 4, competition with a 30-fold molar excess of the indicated oligonucleotides.

complex (Fig. 2B, lane 2). This complex can be inhibited by the addition of an excess of cold oligonucleotides with the same sequence (Fig. 2B, lanes 9 to 11) or corresponding to other E box sequences (oligo B48BS, contained in a human origin of DNA replication [26]; oligo AdMLP, the MLP upstream element; and oligo GAL2, a sequence upstream of the yeast *GAL2* gene [24]). An oligonucleotide encompassing the two downstream-positioned Sp1 sites of the LTR (oligo SP1; Fig. 2B, lane 15) is ineffective in the competition assay.

Like the purified factor, the rUSF^{43} protein also binds to oligo HIV in gel retardation assays (Fig. 2C, lane 2). The addition of anti-rUSF^{43} antibodies to the binding reaction mixture causes a supershift of the protein-DNA complex, in analogy with the effect described for the purified protein (13) (Fig. 2D). Again, the specificity of binding was challenged by the addition of a 6- to 30-fold excess of other cold oligonucleotides to the binding reaction mixtures. Competition was obtained with the same HIV oligonucleotide (Figure 2C, lanes 3 and 4) and with the related oligonucleotides oligo AdMLP and oligo B48BS (lanes 5 to 8) but not with the oligonucleotide containing the Sp1 sites (lanes 11 and 12).

An oligonucleotide carrying a TpA dinucleotide in the core position of the E box instead of CpG (CATATG, oligo MUT-III) is not able to compete for binding (Fig. 2C, lanes 9 and 10). This result indicates that these nucleotides are essential for the specificity of sequence recognition by the protein, as already suggested (5, 14). Since the CpG dinucleotide is the target for physiological cellular methylation, we investigated the role of cytosine methylation within the E box by competition experiments using methylated oligonucleotides with the same sequence of oligo B48BS (Fig. 2E). The results indicated that while methylation on both strands of a CpG outside the E box consensus sequence (oligo Met-III) has no effect on the competing ability of the oligonucleotide with respect to the unmethylated sequence (Fig. 2E; compare lanes 2 and 3), methylation of the core CpG dinucleotide of the E box (oligo Met-I) greatly affects competition (lane 4). These results obtained with the rUSF^{43} protein reflect those described for crude nuclear extracts from HeLa cells (24).

Interactions of USF with the E box of the HIV-1 LTR. Further details of the interactions between purified USF or rUSF^{43} and the HIV-1 LTR were obtained by a DNase I footprinting assay. As shown in Fig. 3A, the incubation of an LTR DNA probe labelled at the 5′ end of the noncoding strand with the purified factor (lane 3) prevents DNase I cleavage of a box centered around the CACGTG sequence (nucle-

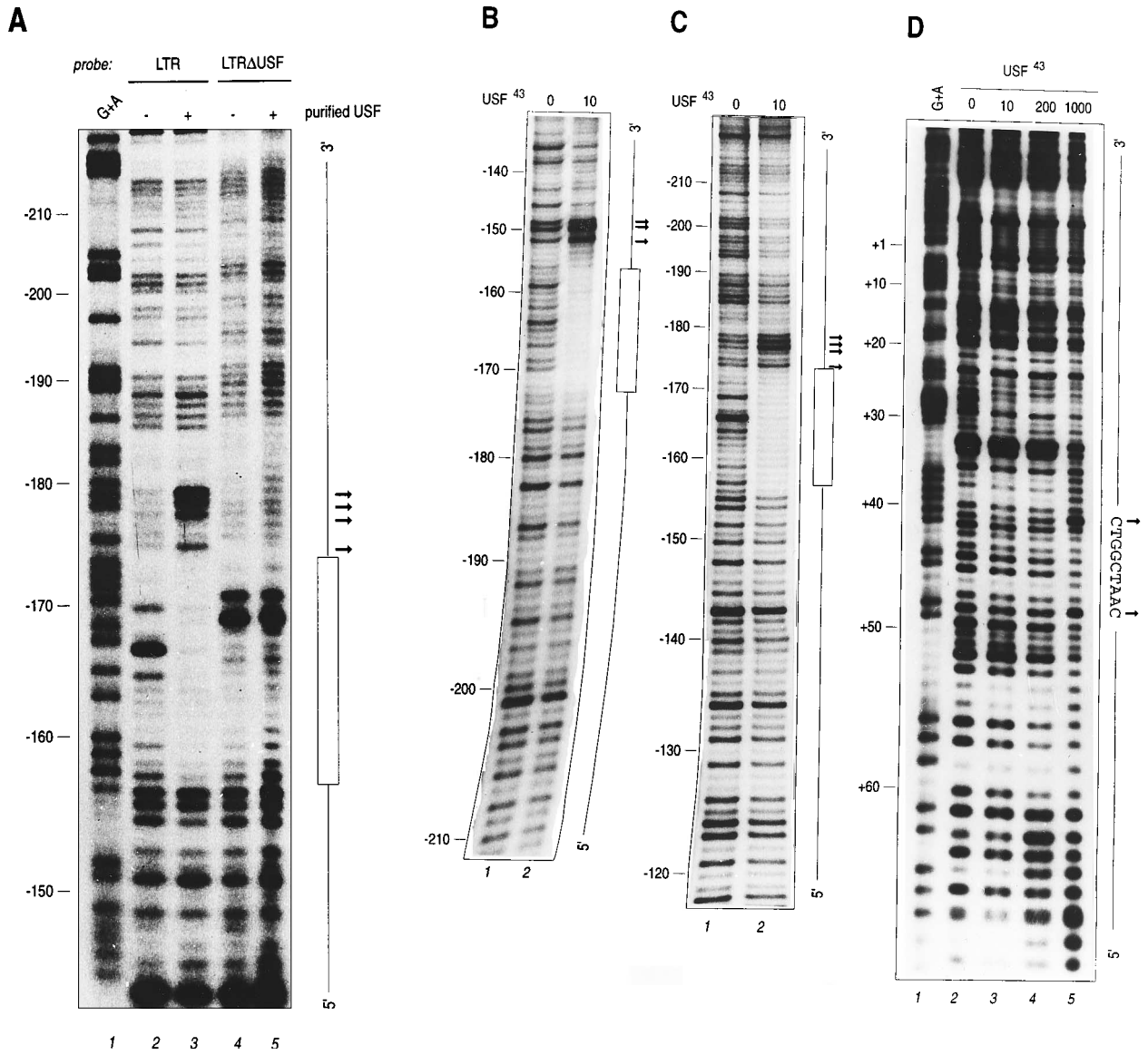

FIG. 3. DNase I footprinting assays. (A) DNase I footprinting with purified USF. Lane 1, G+A chemical cleavage ladder; lanes 2 and 3, DNA fragment pattern generated by DNase I treatment after incubation of a 5′-end-labelled probe in the absence (lane 2) or in the presence (lane 3) of USF; lanes 4 and 5, same as lanes 2 and 3 but with a DNA probe with a mutated E box. The box and the arrows on the right indicate the protected region and the hypersensitive nucleotides, respectively, of lane 3. (B) DNase I footprinting on the LTR coding strand with rUSF^{43}. Lane 1, 10 ng of GST added; lane 2, 10 ng of recombinant GST-USF^{43} fusion protein. The box and the arrows on the right indicate the protected region and the hypersensitive nucleotides, respectively, of lane 2. (C) DNase I footprinting on the LTR noncoding strand with rUSF^{43}. Lane 1, 10 ng of GST; lane 2, 10 ng of recombinant GST-USF^{43} fusion protein. The box and arrows are as in panel B. (D) DNase I footprinting on the region encompassing the transcription start site with rUSF^{43}. Lane 1, G+A chemical cleavage ladder; lane 2, 10 ng of GST; lanes 3 to 5, 10, 200, and 1,000 ng, respectively, of recombinant GST-USF^{43} fusion protein. A sequence with altered DNase I sensitivity appearing with 1,000 ng of protein, with the formation of two hypersensitive sites (indicated by arrows), is shown on the right.

otides from position −173 to −157 upstream of the transcription start site). Lanes 4 and 5 of Fig. 3A show the results of DNase I digestion of a probe derived from plasmid pLTRΔ USF, in which the CACGTG motif was mutated into an *Eco*RI site (GAATTC). In this case, no difference in the DNase I digestion pattern obtained in the presence or absence of the protein can be detected, again indicating that the E box is absolutely required for protein binding. Protein binding causes the appearance of four strong hypersensitive sites at the 3′ end of the recognized box, corresponding to a G at position −179, an A at −178, a T at −177, and a T at −174 (arrows in Fig. 3A). A similar pattern of footprinting is produced by the binding of rUSF^{43} (Fig. 3C). On the upper, coding strand, the recombinant protein protects the E box from DNase I digestion, footprinting over nucleotides −157 to −172, again generating three strong hypersensitive sites at the 3′ end (a C at −152, a G at −150, and a C at −149) (Fig. 3B). The generation of these DNase I hypersensitive sites is most likely due to a structural alteration of the double helix upon protein binding, resulting in a widened groove where nucleotides are more accessible to DNase I digestion.

It has been reported that rUSF^{43}, in addition to binding E box elements, can also bind to the initiator elements of the adenovirus MLP and of HIV-1 (17, 56). Therefore, we chal-

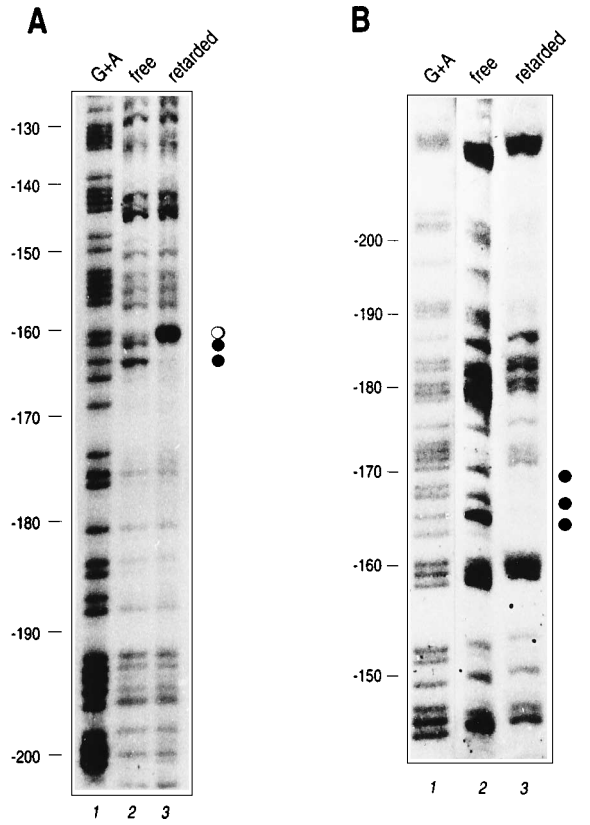

FIG. 4. Methylation protection assays. (A) Methylation protection on the LTR coding strand with rUSF[43]. Lane 1, G+A chemical cleavage ladder; lane 2, methylation pattern of the free band recovered from the gel shift assay; lane 3, methylation pattern of the retarded band. The filled and empty circles on the right indicate, respectively, protected and hypersensitive purines. (B) Methylation protection on the noncoding strand of the LTR with rUSF[43]. Lane 1, G+A chemical cleavage ladder; lane 2, methylation pattern of the free band recovered from the gel shift assay; lane 3, methylation pattern of the retarded band. The filled circles on the right indicate protected purines.

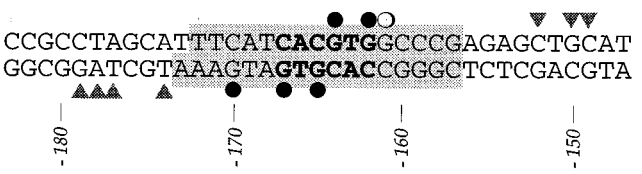

FIG. 5. Summary of the interactions of USF with the LTR E box obtained by DNase I footprinting and methylation protection experiments. The nucleotides of the E box motif are in boldface. ▨, region protected from DNase I digestion; ▲, nucleotide hypersensitive to DNase I digestion; ●, nucleotide protected from methylation; ○, nucleotide hypersensitive to methylation.

lenged a DNA probe encompassing the HIV-1 transcription start site in DNase I footprinting experiments with increasing amounts of highly purified (>95% homogeneity) rUSF[43] (from 10 to 1,000 ng) (Fig. 3D). No changes were observed with up to 200 ng of protein, in contrast to the clear protection of the E box, which is complete with even 10 ng (Fig. 3B and C). Only upon addition of 1 μg of protein were some changes in the footprinting pattern detected in the region from nucleotide +42 to +47 relative to transcription start site, with the formation of two hypersensitive sites corresponding to a C at position 40 and a C at position 48. However, it should be considered that this sequence does not fully match that of the previously reported Inr elements of the HIV-1 promoter (17) and that these changes could be nonspecific, because of the very large amount of protein used.

A further insight into the pattern of interaction of USF with the HIV-1 LTR was obtained by methylation protection experiments with dimethyl sulfate, a sensitive chemical method to probe DNA-protein contacts which allows the determination of the purines protected from methylation by DNA-bound proteins. The results of methylation protection experiments with rUSF[43] and the HIV-1 LTR as a probe are shown in Fig. 4A for the upper, coding strand and in Fig. 4B for the lower, noncoding strand. Guanines at positions −162 and −164 on the coding strand and at positions −165, −167, and −170 on the noncoding strand are clearly protected from methylation, while a guanine at position −161 on the coding strand is hypersensitive.

The results obtained from the DNase I and methylation protection experiments are summarized in Fig. 5. It is evident that the E box consensus sequence CACGTG is centered at the twofold rotational axis of symmetry of protein-DNA interaction, with major contacts occurring at the 3′ half of the DNA sequence on both strands. This pattern closely resembles the one produced by USF binding to the MLP upstream element (42, 60).

Template LTR bending upon USF binding. The generation of sites of increased sensitivity to DNase I symmetrically located on both strands and the presence of a site hypersensitive to methylation on the coding strand (Fig. 5) suggest that DNA undergoes a structural distortion upon protein binding.

The ability of USF to bend the LTR template upon binding was tested by the circular permutation assay, a method based on the position-dependent effect of DNA bends on the electrophoretic mobilities of DNA fragments (71). To exploit this technique, a set of six probes (B1 to B6) was obtained by PCR amplification with the LTR as a template. All of these probes have approximately the same length (from 244 to 246 bp) but differ because of the position of the USF binding site (Fig. 6A). They were tested in gel retardation assays both with rUSF[43] and with purified USF (Fig. 6B and C, respectively). As expected, all of the probes gave rise to a retarded complex upon incubation with both protein species. However, the relative mobilities of the retarded bands of the probes containing the binding site in the middle were lower than those of the probes containing the binding site at one extremity, suggesting that the template DNA becomes bent upon protein binding. The center of the flexure was mapped by plotting the mobility as a function of the distance of the binding site from the end of the probe (Fig. 6B and C, bottom) and was found to correspond to the position of the CACGTG box.

Estimation of the bending angle as described by Thompson and Landy (66) indicates values of 90° of bending for purified USF and 110° for rUSF[43], as evaluated with the average values obtained from three independent experiments.

As a negative control, the Ku protein, which binds to DNA without an apparent sequence specificity, giving rise to several retarded complexes due to multiple protein dimers bound to the probe (13), was also tested in gel retardation assays with the same probes. As shown in Fig. 6D, the retarded bands generated by all of the probes migrate with the same mobility.

In vitro transcription assays. The role of USF in the control of transcription from the HIV-1 LTR was determined by in vitro transcription assays. For this purpose, the LTR region upstream of nucleotide −1 was cloned upstream of a ~380-nucleotide G-less cassette. Each in vitro transcription experiment was simultaneously performed by the addition to the

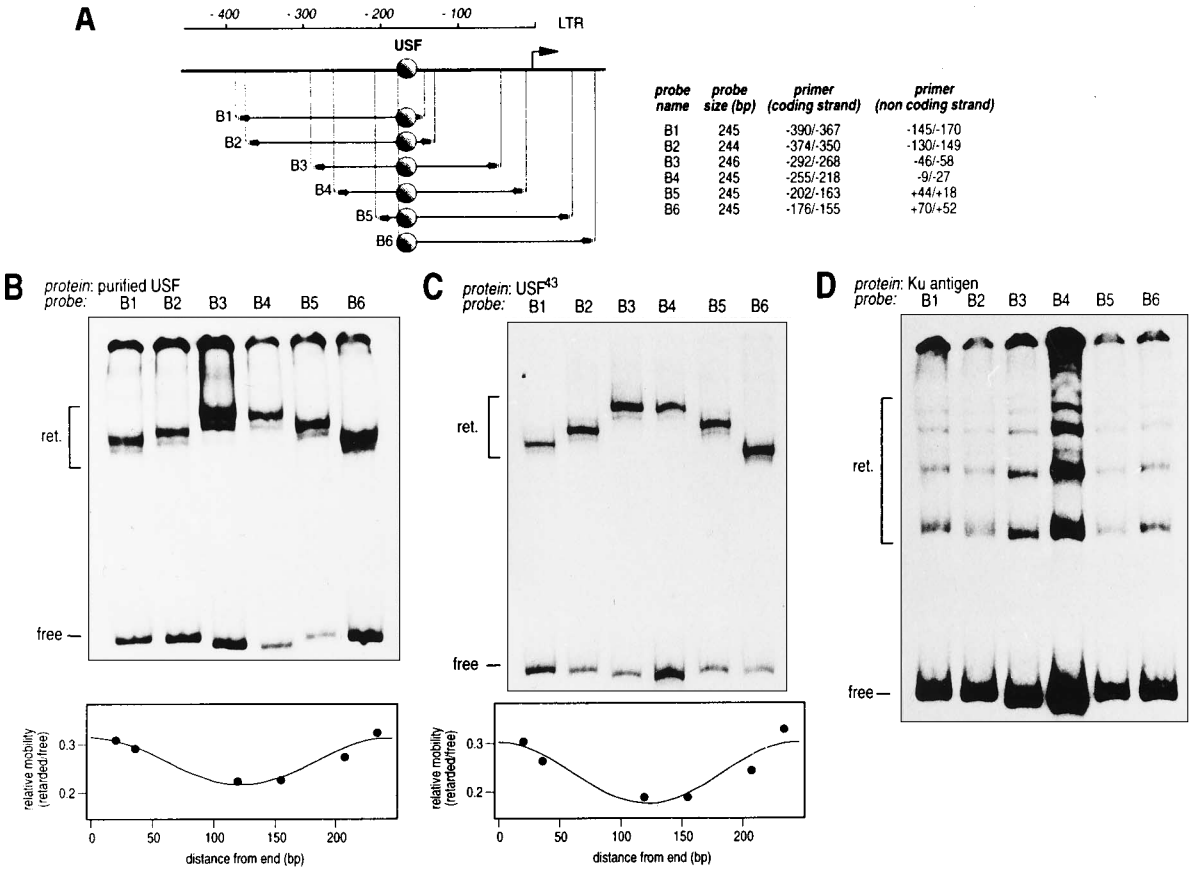

FIG. 6. Circular permutation assay. (A) A set of six probes (B1 to B6) containing the USF site in different positions was generated by PCR amplification with the primers localized as indicated in the table on the right (numbering refers to the transcription start site according to sequence hxb2cg of GenBank). The arrow at the top indicates the transcription start site. (B to D) Circular permutation assays with probes B1 to B6 and purified USF (B), rUSF^{43} (C) and the Ku protein (13) (D). The locations of the free and retarded (ret.) complexes are indicated. The graphs in the lower parts of panels B and C show the relative mobilities of the retarded complexes (distance from the well of the retarded band/distance of the free band) plotted against the distance of the CACGTG box from the end of the fragment.

same tube of a plasmid template containing a shorter G-less cassette (~200 nucleotides) under the control of the fibronectin promoter, to be used as an internal control.

The addition of increasing amounts of highly purified rUSF^{43} to HeLa cell nuclear extracts progressively upregulates transcription from the LTR in a concentration-dependent manner, while it does not affect the transcription driven by the fibronectin promoter (Fig. 7A, lanes 2 to 6). Quantification by scintillation counting and optical scanning of the autoradiograms indicated a reproducible three- to fourfold increase in the LTR-driven signal intensity after normalization to the results for fibronectin; the ratio of the intensities of the RNA bands for the two constructs for each experimental point is reported in the graph below the gel.

Mutation of the CACGTG hexanucleotide to an unrelated *Eco*RI restriction site abolishes the responsiveness of the LTR construct to protein addition (Fig. 7B, lanes 2 to 6). Furthermore, the transcription efficiency of this mutated promoter is invariably lower than that of the wild-type LTR (compare lanes 1 in Fig. 7A and C with lanes 1 in Fig. 7B and D).

The HeLa cell extract used in the in vitro transcription experiments contains, among other E box-binding factors, an endogenous amount of USF (estimated to be in the range of 0.15 to 1.5 ng/μg of extract [51]), which presumably contributes to the basal level of transcriptional activation from the LTR template. In order to saturate the endogenous E box-binding proteins, increasing amounts of a plasmid (pUF128) carrying 128 copies of a human E box-containing sequence were added to the transcription reaction mixtures (Fig. 7C, lanes 2 to 5). The addition of this decoy progressively decreases the level of transcription from the LTR compared with that from the fibronectin promoter. The same molar amounts of vector pUC19, used as a control, are ineffective (Fig. 7C, lanes 6 to 9). Again, mutation of the E box of the LTR abolishes the responsiveness of the LTR construct to the specific decoy addition (Fig. 7D, lanes 2 to 5).

Altogether, these results show that the E box in the context of the U3 LTR region is a positive *cis*-acting element in the control of transcription and that USF is a *trans*-acting factor contributing to this function.

DISCUSSION

The regulation of the rate of transcription of the HIV-1 provirus is achieved by the interaction of several human transcription factors with the LTR DNA and of the viral Tat transactivator with the 5′ end of the nascent RNA. Although the LTR appears as a highly inducible promoter responsive to a number of stimuli which trigger cellular activation of proliferation, it is interesting that most of the human factors binding to the LTR are constitutively present in most cell types and tissues. The reasons for this evolutionary outcome have still to

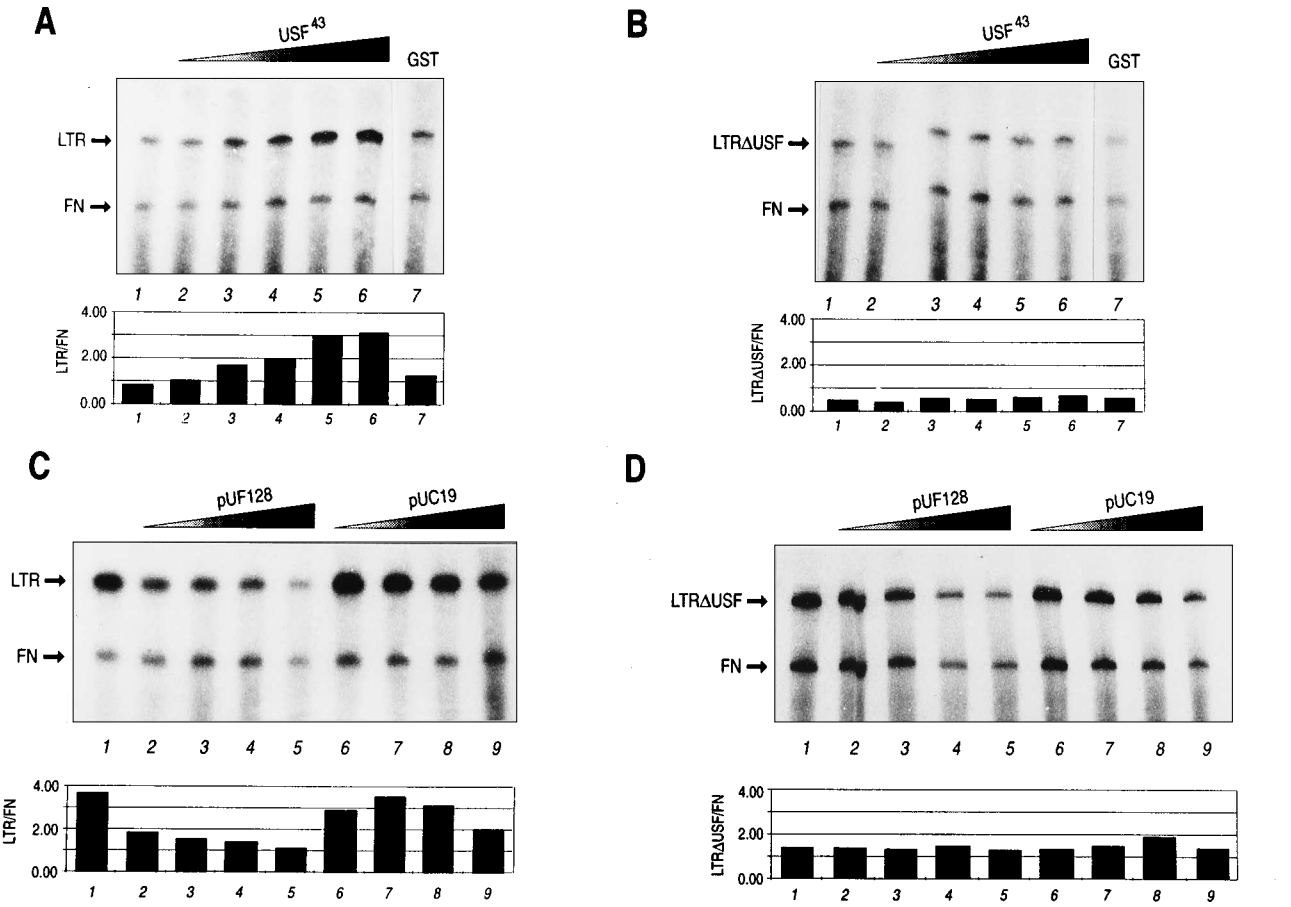

FIG. 7. In vitro transcription assays. (A) Effect of the addition of rUSF[43] on the LTR template. In vitro transcription assays were performed with plasmids pGLA (LTR upstream of a 380-nucleotide G-less cassette) and pFN2 (fibronectin promoter upstream of a 200-nucleotide G-less cassette) and HeLa nuclear extracts. Lane 1, plasmid templates plus nuclear extract; lanes 2 to 6, plasmid templates plus increasing amounts of recombinant GST-USF[43] (50, 100, 200, 500, and 1,000 ng, respectively); lane 7, plasmid templates plus 1,000 ng of GST. The RNA bands corresponding to the two transcripts are indicated. The graph in the lower part shows the ratio of the amounts of the two transcripts for each experimental point. (B) Effect of the addition of rUSF[43] on the LTRΔUSF template. In vitro transcription assays were performed with plasmids pGLE (LTRΔUSF/G-less cassette) and pFN2 and HeLa nuclear extracts. The amounts of USF[43] and GST are the same as in panel A. (C) Effect of the addition of E box decoys on the LTR template. In vitro transcription assays were performed with plasmids pGLA and pFN2 and HeLa nuclear extracts. Lane 1, plasmid templates plus nuclear extract; lanes 2 to 5, plasmid templates plus increasing amounts of plasmid pUF128, containing 128 copies of a CACGTG box (250, 500, 750, and 1,000 ng, respectively); lanes 6 to 9, plasmid templates plus the same molar excesses of vector pUC19. (D) Effect of the addition of E box decoys on the LTRΔUSF template. In vitro transcription assays were performed with plasmids pGLE and pFN2 and HeLa nuclear extracts. The amounts of pUF128 and pUC19 are the same as in panel C.

be understood. It is conceivable that this situation reflects the need either for responsiveness to a variety of cellular environments or for fine tuning of viral transcription in response to a variety of cellular events.

One of the LTR regions which are recognized by a constitutive binding activity in human nuclear extracts is the E box centered at position −164 upstream of the transcription start site. This site is actively engaged in protein-DNA interactions in an actively producing T-cell line in vivo (15) and appears to be strictly conserved, despite the overall HIV-1 sequence variability, in DNA amplified directly from the peripheral blood lymphocytes of patients with AIDS over several years (39). Furthermore, this same site appears to play an active role in the control of transcription by the recently described negative-strand LTR promoter (41).

These observations, as well as the evolutionary conservation of the E box motif and the cognate binding proteins in the upstream sequences of several cellular genes (24, 25), suggest that this sequence has an important biological role in the context of the control of HIV-1 expression.

Interactions of USF with the LTR E box target site. Several nuclear proteins are capable of specifically binding to the CACPuTG sequence. In Southwestern experiments, at least three species (of ~43/44, 70, and 110 kDa) are detectable (13, 25), and in the last few years, the cDNA sequences of several proteins of the b-HLH-Zip family interacting with this E box were cloned.

The major DNA-binding activity detectable by gel retardation assays in nuclear extracts corresponds to transcription factor USF, which is purified as a heterodimer composed of two polypeptides of 43 and 44 kDa encoded by two different genes (27, 65). We have shown, with the experiments reported in this work, that purified human USF specifically binds to the E box at position −162 to −167 of the HIV-1 LTR. Indeed, the 39-kDa HIV-TF1 factor binding to this sequence, which was purified by Maekawa et al., is likely to correspond to USF (36).

The 43- and 44-kDa polypeptides preferentially bind as a heterodimer to the target site (20), even if each one is individually able to bind to DNA (51, 65), as demonstrated also by Southwestern analysis (13). We exploited the available cDNA

sequence for the 43-kDa form of USF to produce a recombinant fusion USF^{43} protein in bacteria. Both the purified ($USF^{43/44}$) and the recombinant ($rUSF^{43}$) proteins show the same binding specificity for the HIV-1 LTR and other targets containing an E box sequence. As expected, $rUSF^{43}$ binding is affected in the same way by the mutations that abolish DNA binding of the eukaryotic protein.

Analogous to the case of the interactions occurring at the MLP upstream element (42, 60), the G residues which exhibit decreased methylation in the presence of USF are symmetrically situated in the two halves of the region of dyad symmetry. This symmetry is consistent with the proposed binding of the factor as a dimer to the target sequence, with each monomer contacting the DNA sequence on the 3′ site of the dyad axis of the E box element (20). It is likely that the protein dimer binds at the center of the palindromic site with the basic domains extending symmetrically into the major grooves of each half site, as suggested previously (22).

Analogous to the sensitivity of the native factor (24, 69), binding of $rUSF^{43}$ to the LTR is abolished by methylation of the core CpG dinucleotide of the E box. This observation suggests that a possible mechanism of control for the function of USF (which is constitutively and ubiquitously expressed [64]) could be through epigenetic modification of the target site, namely, by altering the methylation state of DNA. Since it has been shown that methylation of the LTR represses HIV-1 transcription (4), it is likely that the E box could be one of the target sites mediating this effect. Finally, it is interesting that the mechanism for preventing binding by target site methylation is not restricted to USF, since binding by other members of the b-HLH-Zip family is also sensitive to methylation of the core CpG sequence of the E box (52).

Template LTR bending upon USF binding. The circular permutation analyses performed indicate that both USF purified from HeLa cells and $rUSF^{43}$ bend DNA upon binding to the target site. It is likely that the sites of hypersensitivity to DNA methylation and DNase I digestion detected in the methylation protection and footprinting experiments are generated by an increase in the groove width, typical of DNA bending upon protein binding. Although the circular permutation assay alone does not allow the distinction between DNA bending and increased DNA flexibility, the extent of bending detected (90 to 110°) cannot be due only to a mere increase in flexibility. DNA bending by a number of b-HLH-Zip proteins (including USF and Max) has been reported (22, 70). However, crystal analysis of the complex of DNA with Max (21) and with the USF b-HLH-Zip regions (20) could not show any net bend in the double helical axis. The reasons for this discrepancy are likely to be related to the crystal packing mode of the two structures.

In our experiments, the calculated bending angles induced by HeLa-purified USF and $rUSF^{43}$ were 90 and 110°, respectively. A trivial explanation for this difference could be that $rUSF^{43}$, bearing the GST extension, has a shape different from and a size greater than those of HeLa-purified USF. However, it was demonstrated that there is no significant correlation between the molecular weight of the protein and the extent of the induced DNA bending (33). Therefore, our preferred interpretation is that the difference in the bending abilities is due to the different molecular compositions of the dimers interacting with DNA: an obligate USF^{43} homodimer in the experiments with the recombinant protein and a preferred $USF^{43/44}$ heterodimer in the experiments with the purified factor. In the latter case, the bending angle results from the vectorial sum of the bending angles induced by the two different monomers. In this respect, there are also other examples of differential bending by monomeric or heteromeric forms, such as Myc/Max and Fos/Jun heterodimers (33, 70).

In the context of HIV-1 transcriptional regulation, the studies on the functional role of LTR DNA bending by transcriptional factors probably deserve further scrutiny. In fact, it should be considered that, in addition to USF, other transcription factors that also bind to the LTR (Jun/Fos [33], NF-κB [63], TBP [30], YY1 [37, 45], and Sp1 [31]) are able to bend the target DNA sequence. As a consequence, the traditional picture of the LTR as a linear structure should be replaced by a more realistic three-dimensional view of a highly structured promoter in which nonadjacent proteins also, by DNA bending and looping, can interact among themselves and with the basal transcriptional machinery. In accordance with this view, it has been recently reported that Sp1 interacts with NF-κB (49), that both Sp1 and Tat bind to TBP (19, 32), that YY1 physically interacts with Sp1 (34), and that Fos is able to bind to the 44-kDa form of USF (9). In this respect, it is also interesting that the b-HLH-Zip domain of USF was reported to exist as a bivalent tetramer, potentially able to bind simultaneously to two independent sites, with a possible role in DNA looping (20). However, despite reported evidence showing interactions between USF and the initiator element of the HIV-1 LTR (17, 56), we were not able to obtain a clear footprint on this region, even with the addition of an amount of protein 100-fold greater than that needed to footprint over the E box. Accordingly, the analysis of the USF binding sites by the random oligonucleotide selection procedure showed that the E box motif is almost absolutely required for binding (5).

Functional significance of USF binding to the LTR. In vitro transcription experiments, performed either by the addition of $rUSF^{43}$ or by the subtraction of endogenous E box-binding proteins by E box decoys, indicate that USF acts as a positive regulator of transcription driven by the LTR promoter. This effect is strictly and solely dependent upon the presence of an intact E box, since mutation of the E box impairs the responsiveness of the promoter in both types of experiments and weakens its basal strength.

We have previously reported (25) that a human binding site for b-HLH proteins, contained in a human origin of DNA replication (6, 26), acts as a downregulator of transcription when cloned upstream of the enhancer region of the LTR, therefore replacing the whole negative regulatory element, in agreement with the postulated negative function exerted by the site in the context of the infectious virus (35). Several explanations could be proposed to solve this apparent contradiction between the in vitro and the in vivo experiments. Since the E box is the potential target of different factors, the possibility that USF is not the major protein binding to the LTR site within the cell cannot be ruled out. Alternatively, it could be argued that the major E box-binding factor interacting with the human origin sequence is different from the one binding to the HIV-1 LTR, since the sequences flanking the E box also are important in determining the affinities of different b-HLH-Zip proteins (5). Finally, we have recently shown that other proteins also can modulate binding of USF to its target site (73); similarly, the in vivo function of the USF site could necessitate the presence of a complex combination of other DNA-binding factors, a situation which is hardly reproducible in the in vitro transcription assays. Further experiments are obviously needed to clarify these points.

ACKNOWLEDGMENTS

The work described in this paper was supported by grant 9204-50 of the AIDS program of the Istituto Superiore di Sanita', Rome.

F.D.D.F. and G.M. are supported by SISSA (Trieste) predoctoral fellowships.

We are very grateful to P. Norio for his help in some experiments, to G. Biamonti for the gift of plasmid pBSA1MLTF, to R. Roeder for the anti-rUSF antiserum, to F. Cobianchi for the gift of plasmid pUF128, and to Ann Crum for careful reading of the manuscript.

REFERENCES

1. **Ayer, D. E., L. Kretzner, and R. N. Eisenman.** 1993. Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. Cell **72:**211–222.
2. **Bagnarelli, P., S. Menzo, A. Valenza, A. Manzin, M. Giacca, F. Ancarani, G. Scalise, P. E. Varaldo, and M. Clementi.** 1992. Molecular profile of human immunodeficiency virus infection in asymptomatic subjects and AIDS patients. J. Virol. **66:**7328–7335.
3. **Beckmann, H., L.-K. Su, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4:**167–179.
4. **Bednarik, D. P., J. A. Cook, and P. M. Pitha.** 1990. Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. EMBO J. **9:**1157–1164.
5. **Bendall, A. J., and P. L. Molloy.** 1994. Base preferences for DNA binding by the bHLH-Zip protein USF: effects of $MgCl_2$ on specificity and comparison with binding of Myc family members. Nucleic Acids Res. **22:**2801–2810.
6. **Biamonti, G., M. Giacca, G. Perini, G. Contreas, L. Zentilin, F. Weighardt, M. Guerra, G. Della Valle, S. Saccone, S. Riva, and A. Falaschi.** 1992. The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S phase. Mol. Cell. Biol. **12:**3499–3506.
7. **Biamonti, G., G. Perini, F. Weighardt, S. Riva, M. Giacca, P. Norio, L. Zentilin, S. Diviacco, D. Dimitrova, and A. Falaschi.** 1992. A human DNA replication origin: localization and transcriptional characterization. Chromosoma **102:**S24–S31.
8. **Blackwood, E. M., and R. N. Eisenman.** 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA binding complex with Myc. Science **251:**1211–1217.
9. **Blanar, M. A., and W. J. Rutter.** 1992. Interaction cloning—identification of a helix-loop-helix zipper protein that interacts with c-Fos. Science **256:**1014–1018.
10. **Bresnick, E. H., and G. Felsenfeld.** 1993. Evidence that the transcription factor USF is a component of the human beta-globin locus control region heteromeric protein complex. J. Biol. Chem. **268:**18824–18834.
11. **Carr, C. S., and P. A. Sharp.** 1990. A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol. Cell. Biol. **10:**4384–4388.
12. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43:**439–448.
13. **Csordás Tóth, E., L. Marusic, A. Ochem, A. Patthy, S. Pongor, M. Giacca, and A. Falaschi.** 1993. Interactions of USF and Ku antigen with a human DNA region containing a replication origin. Nucleic Acids Res. **21:**3257–3263.
14. **Dang, C. V., C. Dolde, M. L. Gillison, and G. J. Kato.** 1992. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. Proc. Natl. Acad. Sci. USA **89:**599–602.
15. **Demarchi, F., P. Dagaro, A. Falaschi, and M. Giacca.** 1992. Probing protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1 by in vivo footprinting. J. Virol. **66:**2514–2518.
16. **Demarchi, F., P. Dagaro, A. Falaschi, and M. Giacca.** 1993. In vivo footprinting analysis of constitutive and inducible protein-DNA interactions at the long terminal repeat of human immunodeficiency virus type 1. J. Virol. **67:**7450–7460.
17. **Du, H., A. L. Roy, and R. G. Roeder.** 1993. Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. EMBO J. **12:**501–511.
18. **Edenberg, H. J., and C. J. Brown.** 1992. Regulation of human alcohol dehydrogenase genes. Pharmacogenetics **2:**185–196.
19. **Emili, A., J. Greenblatt, and C. J. Ingles.** 1994. Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Mol. Cell. Biol. **14:**1582–1593.
20. **Ferré-D'Amaré, A. R., P. Pognonec, R. G. Roeder, and S. K. Burley.** 1994. Structure and function of the b/HLH/Z domain of USF. EMBO J. **13:**180–189.
21. **Ferré-D'Amaré, A. R., G. C. Prendergast, E. B. Ziff, and S. K. Burley.** 1993. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature (London) **363:**38–45.
22. **Fisher, D. E., L. A. Parent, and P. A. Sharp.** 1992. Myc/Max and other helix-loop-helix/leucine zipper proteins bend DNA toward the minor groove. Proc. Natl. Acad. Sci. USA **89:**11779–11783.
23. **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.
24. **Giacca, M., M. I. Gutierrez, S. Diviacco, F. De Marchi, G. Biamonti, S. Riva, and A. Falaschi.** 1989. A protein target site in an early replicated human DNA sequence: a highly conserved binding motif. Biochem. Biophys. Res. Commun. **165:**956–965.
25. **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.
26. **Giacca, M., L. Zentilin, P. Norio, S. Diviacco, D. Dimitrova, G. Contreas, G. Biamonti, G. Perini, F. Weighardt, S. Riva, and A. Falaschi.** 1994. Fine mapping of a replication origin of human DNA. Proc. Natl. Acad. Sci. USA **91:**7119–7123.
27. **Gregor, P. D., M. Sawadogo, and R. G. Roeder.** 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. **4:**1730–1740.
28. **Hattori, M., A. Tugores, L. Veloz, M. Karin, and D. A. Brenner.** 1990. A simplified method for the preparation of transcriptionally active liver nuclear extracts. DNA Cell Biol. **9:**777–781.
29. **Heiland, S., R. Knippers, and N. Kunze.** 1993. The promoter region of the human type-I-DNA-topoisomerase gene—protein-binding and sequences involved in transcriptional regulation. Eur. J. Biochem. **217:**813–822.
30. **Horikoshi, M., M. F. Carey, H. Kakidani, and R. G. Roeder.** 1988. Mechanism of activation of a yeast activator: direct effect of GAL4 derivatives on mammalian TFIID-promoter interactions. Cell **54:**665–669.
31. **Ikeda, K., K. Nagano, and K. Kawakami.** 1993. Possible implications of Sp1-induced bending of DNA on synergistic activation of transcription. Gene **136:**341–343.
32. **Kashanchi, F., G. Piras, M. F. Radonovich, J. F. Duvall, A. Fattaey, C. M. Chiang, R. G. Roeder, and J. N. Brady.** 1994. Direct interaction of human TFIID with the HIV-1 transactivator Tat. Nature (London) **367:**295–299.
33. **Kerppola, T. K., and T. Curran.** 1991. Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: implications for transcription factor cooperativity. Cell **66:**317–326.
34. **Lee, J. S., K. M. Galvin, and Y. Shi.** 1993. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc. Natl. Acad. Sci. USA **90:**6145–6149.
35. **Lu, Y., N. Touzjan, 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.
36. **Maekawa, T., T. Sudo, M. Kurimoto, and S. Ishii.** 1991. USF-related transcription factor, HIV-TF1, stimulates transcription of human immunodeficiency virus-1. Nucleic Acids Res. **19:**4689–4694.
37. **Margolis, D. M., M. Somasundaran, and M. R. Green.** 1994. Human transcription factor YY1 represses human immunodeficiency virus type 1 transcription and virion production. J. Virol. **68:**905–910.
38. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
39. **Michael, N. L., L. D'Arcy, P. K. Ehrenberg, and R. R. Redfield.** 1994. Naturally occurring genotypes of the human immunodeficiency virus type 1 long terminal repeat display a wide range of basal and Tat-induced transcriptional activities. J. Virol. **68:**3163–3174.
40. **Michael, N. L., M. Vahey, D. S. Burke, and R. R. Redfield.** 1992. Viral DNA and mRNA expression correlate with the stage of human immunodeficiency virus (HIV) type 1 infection in humans: evidence for viral replication in all stages of HIV disease. J. Virol. **66:**310–316.
41. **Michael, N. L., M. T. Vahey, L. d'Arcy, P. K. Ehrenberg, J. D. Mosca, J. Rappaport, and R. R. Redfield.** 1994. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. J. Virol. **68:**979–987.
42. **Miyamoto, N. G., V. Moncollin, J. M. Egly, and P. Chambon.** 1985. Specific interaction between a transcription factor and the upstream element of the adenovirus-2 major late promoter. EMBO J. **4:**3563–3570.
43. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58:**537–544.
44. **Nabel, G. J.** 1993. The role of cellular transcription factors in the regulation of human immunodeficiency virus gene expression, p. 49–73. *In* B. R. Cullen (ed.), Human retroviruses. IRL Press, Oxford.
45. **Natesan, S., and M. Z. Gilman.** 1993. DNA bending and orientation-dependent function of YY1 in the c-fos promoter. Genes Dev. **7:**2497–2509.
46. **Pantaleo, G., C. Graziosi, J. F. Demarest, L. Butini, M. Montroni, C. H. Fox, J. M. Orenstein, D. P. Kotler, and A. S. Fauci.** 1993. HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. Nature (London) **362:**355–358.
47. **Papavassiliou, A. G.** 1993. Localization of DNA-protein contact points by DMS resistance of complexes resolved in gel retardation assays. Nucleic Acids Res. **21:**757–758.
48. **Peritz, L. N., E. J. B. Fodor, D. W. Silversides, P. A. Cattini, J. D. Baxter, and N. L. Eberhardt.** 1988. The human growth hormone gene contains both

positive and negative control elements. J. Biol. Chem. **263:**5005–5007.
49. **Perkins, N. D., N. L. Edwards, C. S. Duckett, A. B. Agranoff, R. M. Schmid, and G. J. Nabel.** 1993. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. EMBO J. **12:**3551–3558.
50. **Piatak, M., Jr., M. S. Saag, L. C. Yang, S. J. Clark, J. C. Kappes, K.-C. Luk, B. H. Hahn, G. M. Shaw, and J. D. Lifson.** 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science **259:**1749–1754.
51. **Pognonec, P., and R. G. Roeder.** 1991. Recombinant 43-kDa USF binds to DNA and activates transcription in a manner indistinguishable from that of natural 43/44-kDa USF. Mol. Cell. Biol. **11:**5125–5136.
52. **Prendergast, G. C., and E. B. Ziff.** 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science **251:**186–189.
53. **Read, M. L., A. R. Clark, and K. Docherty.** 1993. The helix-loop-helix transcription factor USF (upstream stimulating factor) binds to a regulatory sequence of the human insulin gene enhancer. Biochem. J. **295:**233–237.
54. **Riccio, A., P. V. Pedone, L. R. Lund, T. Olesen, H. S. Olsen, and P. A. Andreasen.** 1992. Transforming growth factor beta 1-responsive element: closely associated binding sites for USF and CCAAT-binding transcription factor-nuclear factor I in the type 1 plasminogen activator inhibitor gene. Mol. Cell. Biol. **12:**1846–1855.
55. **Rosen, C. A., J. G. Sodroski, 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.
56. **Roy, A. L., M. Meisterernst, P. Pognonec, and R. G. Roeder.** 1991. Cooperative interaction of an initiator-binding transcription initiation factor and the helix-loop-helix activator USF. Nature (London) **354:**245–248.
57. **Saksela, K., C. Stevens, P. Rubinstein, and D. Baltimore.** 1994. Human immunodeficiency virus type 1 mRNA expression in peripheral blood cells predicts disease progression independently of the numbers of CD4+ lymphocytes. Proc. Natl. Acad. Sci. USA **91:**1104–1108.
58. **Sato, M., S. Ishizawa, T. Yoshida, and S. Shibahara.** 1990. Interaction of upstream factor with the human heme oxygenase gene promoter. Eur. J. Biochem. **188:**231–237.
59. **Sawadogo, M., and R. G. Roeder.** 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. Proc. Natl. Acad. Sci. USA **82:**4394–4398.
60. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43:**165–175.
61. **Sawadogo, M., M. W. Van Dyke, P. D. Gregor, and R. G. Roeder.** 1988. Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. J. Biol. Chem. **263:**11985–11993.
62. **Schorpp, M., W. Kugler, U. Wagner, and G. U. Ryffel.** 1988. Hepatocyte-specific promoter element HP1 of the *Xenopus* albumin gene interacts with transcriptional factors of mammalian hepatocytes. J. Mol. Biol. **202:**307–320.
63. **Schreck, R., H. Zorbas, E. L. Winnacker, and P. A. Baeuerle.** 1990. The NF-κB transcription factor induces DNA bending which is modulated by its 65-kD subunit. Nucleic Acids Res. **18:**6497–6502.
64. **Sirito, M., Q. Lin, T. Maity, and M. Sawadogo.** 1994. Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. Nucleic Acids Res. **22:**427–433.
65. **Sirito, M., S. Walker, Q. Lin, M. T. Kozlowsi, W. H. Klein, and M. Sawadogo.** 1992. Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. Gene Exp. **2:**231–240.
66. **Thompson, J. F., and A. Landy.** 1988. Empirical estimation of protein-induced DNA bending angles: application to site-specific recombination complexes. Nucleic Acids Res. **16:**9687–9705.
67. **Thorn, J. T., A. V. Todd, D. Warrilow, F. Watt, P. L. Molloy, and H. J. Iland.** 1991. Characterization of the human N-ras promoter region. Oncogene **6:**1843–1850.
68. **Vaulont, S., and A. Kahn.** 1994. Transcriptional control of metabolic regulation genes by carbohydrates. FASEB J. **8:**28–35.
69. **Watt, F., and P. L. Molloy.** 1988. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. **2:**1136–1143.
70. **Wechsler, D. S., and C. V. Dang.** 1992. Opposite orientations of DNA bending by c-Myc and Max. Proc. Natl. Acad. Sci. USA **89:**7635–7639.
71. **Wu, H. M., and D. M. Crothers.** 1984. The locus of sequence-directed and protein-induced DNA bending. Nature (London) **308:**509–513.
72. **Zervos, A. S., J. Gyuris, and R. Brent.** 1993. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. Cell **72:**223–232.
73. **Zijderveld, D., F. d'Adda di Fagagna, M. Giacca, H. T. M. Timmers, and P. C. van der Vliet.** 1994. Stimulation of the adenovirus major late promoter in vitro by transcription factor USF is enhanced by the adenovirus DNA-binding protein. J. Virol. **68:**8288–8295.