JOURNAL OF VIROLOGY, June 1995, p. 3584–3596 Vol. 69, No. 6
0022-538X/95/$04.00+0

Cloning and Characterization of a Novel Cellular Protein, TDP-43, That Binds to Human Immunodeficiency Virus Type 1 TAR DNA Sequence Motifs

S.-H. IGNATIUS OU, FOON WU, DAVID HARRICH, LEÓN F. GARCÍA-MARTÍNEZ, AND RICHARD B. GAYNOR*

Division of Molecular Virology, Departments of Medicine and Microbiology, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235

Received 28 December 1994/Accepted 21 February 1995

Human immunodeficiency virus type 1 (HIV-1) gene expression is modulated by both viral and cellular factors. A regulatory element in the HIV-1 long terminal repeat known as TAR, which extends from nucleotides −18 to +80, is critical for the activation of gene expression by the transactivator protein, Tat. RNA transcribed from TAR forms a stable stem-loop structure which serves as the binding site for both Tat and cellular factors. Although TAR RNA is critical for Tat activation, the role that TAR DNA plays in regulating HIV-1 gene expression is not clear. Several studies have demonstrated that TAR DNA can bind cellular proteins, such as UBP-1/LBP-1, which repress HIV-1 gene expression and other factors which are involved in the generation of short, nonprocessive transcripts. In an attempt to characterize additional cellular factors that bind to TAR DNA, a λgt11 expression cloning strategy involving the use of a portion of TAR DNA extending from −18 to +28 to probe a HeLa cDNA library was used. We identified a cDNA, designated TAR DNA-binding protein (TDP-43), which encodes a cellular factor of 43 kDa that binds specifically to pyrimidine-rich motifs in TAR. Antibody to TDP-43 was used in gel retardation assays to demonstrate that endogenous TDP-43, present in HeLa nuclear extract, also bound to TAR DNA. Although TDP-43 bound strongly to double-stranded TAR DNA via its ribonucleoprotein protein-binding motifs, it did not bind to TAR RNA extending from +1 to +80. To determine the function of TDP-43 in regulating HIV-1 gene expression, in vitro transcription analysis was performed. TDP-43 repressed in vitro transcription from the HIV-1 long terminal repeat in both the presence and absence of Tat, but it did not repress transcription from other promoters such as the adenovirus major late promoter. In addition, transfection of a vector which expressed TDP-43 resulted in the repression of gene expression from an HIV-1 provirus. These results indicate that TDP-43 is capable of modulating both in vitro and in vivo HIV-1 gene expression by either altering or blocking the assembly of transcription complexes that are capable of responding to Tat.

Multiple regulatory elements in the human immunodeficiency virus type 1 (HIV-1) long terminal repeat (LTR) are critical in controlling the level of gene expression (24). These elements, including NF-κB (51), Sp1 (27, 30), TATA (5, 23, 31, 43, 53, 54), and TAR (23, 31, 34, 62, 71, 72, 74), serve as binding sites for a variety of cellular transcription factors which may either directly or indirectly influence the degree of transactivation by the Tat protein. The exact mechanism by which Tat stimulates gene expression is not known, although effects on both transcriptional initiation and elongation seem likely (20, 33, 34, 42, 45, 55). Identifying cellular transcription factors that modulate Tat function is therefore critical in understanding the mechanisms regulating HIV-1 gene expression.

One regulatory element in the HIV-1 LTR known as TAR is critical for Tat activation (57). The TAR element, which extends from −18 to +80 in the HIV-1 LTR, is transcribed into a stable stem-loop RNA structure between +1 and +60 which is critical for Tat activation (4, 21, 22, 28, 60). At least three regions within TAR RNA, including the stem structure, the bulge, and the loop sequences, are each required for high levels of Tat activation (6, 21, 22, 59). The bulge region in TAR RNA serves as the binding site for Tat (12, 17, 69), while the loop sequences serve as the binding site for the cellular factor TRP-185 (62, 71). The use of heterologous promoter constructs has demonstrated that the presence of TAR is sufficient to confer responsiveness to Tat (5, 55). In addition, recent studies indicate that HIV-1 strains containing TAR element mutations exhibit markedly decreased growth properties (27).

In contrast to the established role of TAR RNA on modulating HIV-1 gene expression, the function of TAR DNA remains less clear. DNase I footprinting with both HeLa and T-lymphocyte nuclear extracts gave extensive protection over TAR DNA extending from −10 to +52 (22, 23, 31, 44). Chromatographic fractionation of HeLa nuclear extract resulted in the purification of a 64-kDa protein designated UBP-1 or LBP-1 (72), which binds to multiple sites present in both TAR DNA and sequence motifs flanking the TATA box (31, 34, 72, 74). The binding of purified UBP-1/LBP-1 (34, 74) to these latter sites represses HIV-1 gene expression in in vitro transcription assays, probably by preventing the assembly of functional transcription initiation complex. The cloning and subsequent analysis of several cDNAs encoding four members of the LBP-1 family of transcription factors have further demonstrated the ability of these proteins to bind to TAR DNA and repress basal gene expression from the HIV-1 LTR (74). Other proteins such as CTF/NF1 have also been shown to bind to

* Corresponding author. Mailing address: Division of Molecular Virology, Department of Internal Medicine, University of Texas Southwestern Medical School, 5323 Harry Hines Blvd., Dallas, TX 75235-8594. Phone: (214) 648-7570. Fax: (214) 648-8862.

downstream regions of TAR DNA, although their roles in HIV-1 gene regulation remain to be determined (31).

Another TAR DNA element has also been implicated in the generation of short, nonprocessive transcripts that terminate at the base of the TAR RNA stem-loop structure in the absence of Tat (55, 61). The regulatory element responsible for the generation of these transcripts, designated the inducer of short transcripts, has been mapped to a DNA element in TAR located between nucleotides −4 and +25 (61). When TAR is placed downstream of a variety of heterologous promoters, the inducer of short transcripts possesses enhancer activity which generates high levels of short transcripts that terminate approximately at nucleotide +60 (55). Although the generation of these short transcripts is not required for Tat activation, such transcripts may play an important role in other facets of the HIV-1 life cycle such as viral latency. In short, TAR DNA may possess distinct regulatory elements that play an important role in modulating HIV-1 gene expression. Therefore, it is important to identify and characterize additional cellular proteins that bind to the TAR DNA region and determine their potential roles on HIV-1 gene expression.

In this study, we used UV cross-linking of a labeled TAR DNA probe, in the presence of different HeLa nuclear extract fractions, to identify three prominent cellular DNA-binding proteins with molecular masses of approximately 43, 65, and 110 kDa. In an attempt to identify genes encoding these proteins, we used an expression cloning method (63) with oligonucleotides corresponding to a portion of the HIV-1 LTR extending from −18 to +28 to probe a λgt11 library containing HeLa cDNA. We identified a cDNA, designated the TAR DNA-binding protein of 43 kDa (TDP-43), which has homology with other cellular proteins that contain the ribonucleoprotein-binding domain (RNP) (for reviews, see references 3, 17, 35, and 46). Gel retardation analysis indicated that recombinant TDP-43 bound specifically to pyrimidine-rich sequence motifs in TAR DNA but did not bind to TAR RNA. TDP-43 repressed gene expression from the HIV-1 LTR in both the absence and the presence of Tat in in vitro transcription assays and in vivo when cotransfected with an HIV-1 proviral construct in HeLa cells. These results describe the characterization of a novel transcription factor which binds to TAR DNA and is involved in both the in vitro and in vivo regulation of HIV-1 gene expression.

MATERIALS AND METHODS

UV cross-linking of endogenous HeLa nuclear proteins binding to the HIV-1 LTR from −18 to +28. HeLa nuclear extract was fractionated on a phosphocellulose column and step-eluted with KCl into four fractions: A (0.1 M KCl eluate), B (0.3 M KCl eluate), C (0.5 M KCl eluate), and D (1.0 M KCl eluate) (16). About 1 μg of each of these phosphocellulose column fractions was incubated at room temperature for 10 min in gel retardation buffer (50 mM KCl, 10 mM Tris [pH 7.9], 0.1 mM EDTA, 10% glycerol, 0.5 mM dithiothreitol, 5 mM $MgCl_2$) with 3 μg of nonspecific competitor poly(dG-dC)-poly(dG-dC) in a total volume of 40 μl. Then 150,000 cpm of a ^{32}P-end-labeled double-stranded HIV-1 LTR oligonucleotide extending from nucleotide −18 to +28 was added to the samples, and the mixtures were incubated for an additional 20 min at room temperature. The samples were put on ice and irradiated with UV light for 30 min, using a UV transilluminator (UVP, Inc.) placed about 4 cm above the samples (72). The 1.5-ml microcentrifuge tubes containing the samples were uncapped during the period of UV irradiation. An equal volume of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer was added, and the samples were heated at 95°C for 5 min and then analyzed by autoradiography following electrophoresis on an SDS-PAGE gel (10% polyacrylamide).

For oligonucleotide competition experiments, only the phosphocellulose column A fraction was used. The conditions were essentially the same as above, except that only 100 ng of nonspecific competitor poly(dG-dC)-poly(dG-dC) was used. The fraction A sample was incubated with 10 ng of double-stranded oligonucleotides as specific competitors at room temperature for 10 min prior to the addition of the ^{32}P-end-labeled probe. The sequences of the oligonucleotides used as specific competitors were as follows: wild-type HIV-1 LTR from −18 to +28 (sequences described elsewhere); −18 to +28 with mutations of pyrimidine-rich sequences, 5′-CTGACGTACGTACGTACGTACATGCGTACGTTACGGACCATCT-3′; and a region of the interleukin-2 promoter enhancer from −295 to −275, 5′-AATTGGAGGAAAAACTGTTTCATACAGAAGGCGT-3′.

Expression cloning of a cDNA encoding TDP-43. A random-primed λgt11 HeLa cell spinner cDNA library was screened with double-stranded ^{32}P-labeled multimerized oligonucleotides that correspond to nucleotides which extend from −18 to +28 in the HIV-1 LTR (63). The sequence of one strand of these oligonucleotides used was 5′-CTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAG-3′. Multiple overlapping clones were identified, and the longest cDNA clone (2.8 kb) contained the complete coding sequence of TDP-43. Different *Eco*RI fragments generated from the cDNA were used in a Northern (RNA) blot analysis with HeLa poly(A)-selected mRNA to ensure that all the *Eco*RI fragments were derived from one contiguous cDNA.

Northern analysis of TDP-43. Northern blot analysis was performed with 10 μg of HeLa or 5 μg of Jurkat poly(A)-selected mRNA following electrophoresis through a 1% formaldehyde–agarose gel. A human tissue Northern blot (Clonetech) containing 2 μg of poly(A)-selected mRNA from various human tissues was also analyzed. An *Eco*RI-*Nde*I DNA fragment of the TDP-43 cDNA corresponding to amino acids 59 to 165 was random-primed labeled with a room temperature labeling kit (Pharmacia). This probe was hybridized to the Northern blots by a rapid-hybridization protocol (Amersham) followed by autoradiography for 4 h as recommended by the manufacturer.

Generation of rabbit polyclonal antiserum against TDP-43. A portion of TDP-43 corresponding to amino acids 92 to 315 was expressed as a gene 10 fusion protein, mixed with complete Freund's adjuvant, and injected into rabbits (54). Fifteen days after the first injection, another injection of this fraction in incomplete Freund's adjuvant was given. Bleeding was performed 10 days after the second injection, and the serum obtained was diluted 1:1 with 1.5 M glycine–3 M NaCl (pH 8.9). It was then loaded onto a protein A-Sepharose column, equilibrated with the same buffer, and washed extensively with 20 column volumes, and the bound immunoglobulin G was eluted with 0.1 M citric acid (pH 3.0) and neutralized with 1/10 volume of 3 M Tris (pH 8.9) at 4°C. The peak immunoglobulin G fraction, as determined by the Bradford method, was dialyzed into 1× phosphate-buffered saline (PBS) at 4°C overnight. The purified TDP-43 antiserum was subsequently used in Western immunoblot experiments.

Western blot analysis of endogenous TDP-43. Equal amounts of protein (~100 μg) from each of the phosphocellulose fractions (0.1, 0.3, 0.5, and 1 M KCl) were subjected to SDS-PAGE (10% polyacrylamide) and transferred to nitrocellulose paper. The immunoreactive endogenous TDP-43 band was visualized by the chemiluminescence protocol (Amersham) with a 1:1,000 dilution of TDP-43 rabbit polyclonal antibody and a 1:10,000 dilution of secondary antibody.

"Shift-Western" assay of endogenous TDP-43. The "Shift-Western" assay is a modification of the published procedure (15). For gel retardation analysis, HeLa nuclear extract phosphocellulose fraction A (100 μg) was incubated with oligonucleotides corresponding to a region in TAR DNA between −18 and +28. The binding was performed at 30°C for 30 min in buffer containing 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.9), 25 mM KCl, 2 mM spermidine, 0.1 mM EDTA, 0.5 mM dithiothreitol, bovine serum albumin at 100 μg/ml, 10% glycerol, 0.025% Nonidet P-40, 5 μg of poly(dG-dC)-poly(dG-dC), and 5 mM $MgCl_2$. DNA-protein complexes were subjected to electrophoresis at 4°C with a constant voltage of 120 V on a 5% native polyacrylamide gel containing 0.5× Tris-borate-EDTA (TBE), 1 mM EDTA, and 0.05% Nonidet P-40. Binding assays were carried out in identical triplicate sets, and the products were resolved on the same gel. After electrophoresis was completed, one set of binding-assay mixtures was dried and subjected to autoradiography while the other two sets were subjected to a standard Western blot protocol as described above. The nitrocellulose blot was then divided in half: one half was probed with the preimmune rabbit serum, and the other half was probed with rabbit polyclonal antiserum against TDP-43. The resultant immunoreactive TDP-43 band was then aligned with the gel-retarded DNA-protein complexes seen following autoradiography.

Construction of TDP-43 carboxy-terminal truncation mutants. TDP-43 (amino acids 7 to 414) was cloned downstream of glutathione *S*-transferase (GST) (64) by a three-way ligation with a *Bst*EII (end-filled)-*Bam*HI-cut TDP-43 cDNA fragment, a 967-bp *Bam*HI-*Pst*I-cut pGEX-2T fragment, and a 3,986-bp *Sma*I-*Pst*I-cut pGEX-2T fragment, so that the resultant vector had duplicated *Bam*HI and *Sma*I sites in front of the *Bam*HI site in the polylinker of pGEX-2T. Carboxy-terminal truncations of TDP-43 fused to GST were also constructed. TDP-43 (amino acids 7 to 333) was generated in a three-way ligation with a *Bam*HI-*Nco*I fragment from the TDP-43 (amino acids 7 to 414) clone and an *Nco*I-*Eco*RI fragment from TDP-43 cDNA cloned into a *Bam*HI-*Eco*RI-cut pGEX-2T vector. A GST–TDP-43 (amino acids 7 to 240) construct was generated by cloning a *Bam*HI-*Fsp*I fragment from the TDP-43 (amino acids 7 to 414) clone into a *Bam*HI-*Sma*I-cut pGEX-2T vector. A GST–TDP-43 (amino acids 7 to 198) construct was generated in a three-way ligation with a *Bam*HI-*Nde*I fragment. The TDP-43 (amino acids 7 to 166) construct was generated by ligating a *Bam*HI-*Nde*I (end-filled) fragment from the TDP-43 (amino acids 7 to 414) clone into a *Bam*HI-*Sma*I-cut pGEX-2T vector. Finally, a GST–TDP-43 (amino acids 7 to 125) construct was generated by cloning an *Nde*I-*Rsa*I fragment from this vector with a *Bam*HI-*Rsa*I fragment from TDP-43 (amino acids 7 to 414) into a *Bam*HI-*Sma*I-cut pGEX-2T vector. All the TDP-43 fusion proteins with GST

were purified as described (64) by the manufacturer (Pharmacia) and visualized by SDS-PAGE (10% polyacrylamide) with Coomassie blue staining.

Gel retardation analysis with wild-type and truncated TDP-43 proteins. The wild-type and truncated TDP-43 cDNAs expressed as GST fusion proteins were purified and used in gel retardation analysis with a ^{32}P-end-labeled double-stranded oligonucleotide corresponding to −18 to +28 in the HIV-1 LTR. Either 1 μg of GST–TDP-43 or GST protein alone was incubated at room temperature for 10 min in gel retardation buffer consisting of 50 mM KCl, 10 mM Tris (pH 7.9), 5 mM $MgCl_2$, 10% glycerol, 0.1 mM EDTA, 0.5 mM dithiothreitol, and 1 μg of poly(dI-dC)-poly(dI-dC). For competition, 10, 20, or 50 ng of unlabeled double-stranded oligonucleotides was added as either specific or nonspecific competitor followed by 100,000 cpm of labeled probe incubated at room temperature for 20 min. The resultant protein-DNA complex was resolved by electrophoresis at 120 V on a 6% native polyacrylamide gel containing 0.5× TBE at 4°C followed by autoradiography. For gel retardation analysis with wild-type TDP-43, oligonucleotide competitors included those from the adenovirus major late promoter and the HIV-1 LTR indicated by nucleotides are shown below: adenovirus major late promoter TATA, 5′-AAGGGGGGCTATAAAAGGGG GTGGGGG-3′; −45/+1, 5′-GCGTGCCTCAGATGCTGCATATAAGCAGC AGCTGCTTTTTGCCTGTACT-3′; −46/−13, 5′-GCGTGCCTCAGATGCTG CATATAAGCAGCTGCTTT-3′; and +10/+50, 5′-GGTTAGACCAGATCTG AGCCTGGGAGCTCTCTGGCTAACTAG-3′. In another set of experiments, we used a region of the HIV-1 LTR extending from −18 to +28 which contained either wild-type or mutated sequences in two pyrimidine-rich sequences (Δ1, Δ2, Δ1+Δ2), as shown: wild type, 5′-CTGCTTTTTGCCTGTACTGGGTCTCTCT GGTTAGACCAGATCT-3′; Δ1, 5′-CTGACGTACGTACGTACTGGGTCTC TCTGGTTAGACCAGATCT-3′; Δ2, 5′-CTGCTTTTTGCCTGTACGTACATT GCGTACGTTAGACCAGATCT-3′; and Δ1+Δ2, 5′-CTGACGTACGTACG TACGTACATGCGTACGTTACGGACCATCT-3′.

Phenanthroline-copper footprinting. A 126-bp HIV-1 LTR fragment extending from −46 to +80 was cloned into pUC19 cut with *Sma*I-*Hin*d3, and probes were prepared by end labeling at either the *Hin*dIII or the *Eco*RI site. Gel retardation was performed with 100,000 cpm of each of these probes and 2 μg of thrombin-cleaved GST–TDP-43 protein. The wet retardation gel was exposed to X-ray film for 30 min at room temperature and then subjected to autoradiography for 30 min at room temperature until the retarded bands and free probes were visible. The bands were excised with a razor and immersed in 100 μl of 50 mM Tris (pH 8.0); this was followed by the addition of phenanthroline-copper for 10 min at room temperature (9, 41). The reaction was quenched, and the eluted DNA was subject to precipitation before it was loaded on an 8% DNA sequencing gel.

Cyclic amplification and selection of targets procedure. The target 78-mer oligonucleotide for cyclic amplification and selection of targets contained specific sites for PCR amplification in addition to restriction enzyme sites flanking a central core of 35 degenerate nucleotides (7, 70). A 25-μg portion of the single-stranded oligonucleotide was converted to double-stranded DNA by annealing a twofold excess of the 3′ primer and performing the primer extension reaction with 25 U of *Taq* DNA polymerase, 0.2 mM deoxynucleoside triphosphates (dNTPs), and 1.5 mM $MgCl_2$ at 94°C for 30 s, 55°C for 1 min, and 72°C for 20 min.

To coat the magnetic beads with TDP-43 antibody, 1 ml of TDP-43 polyclonal rabbit serum was mixed with 200 μl of magnetic beads (Dynan) overnight at 4°C. The beads were collected by magnetic pull, washed with 500 μl of 1× PBS–0.1% bovine serum albumin–0.1% Nonidet P-40, and then resuspended in 200 μl of washing solution and kept at 4°C. The binding reactions were done at room temperature for 30 min in a total volume of 20 μl containing one-fifth of the double-stranded oligonucleotides and 100 ng of GST–TDP-43 (amino acids 7 to 414) protein in the same buffer as used in the gel retardation analysis. Either TDP-43-coated antibody beads (5 μl) or beads coated with preimmune serum were added and mixed gently at 4°C for 30 min; they were then collected by magnetic pull, washed four times with buffer, and added to a PCR containing 500 pmol each of F (forward) and R (reverse) primers, 0.2 mM dNTPs, and 1.5 mM $MgCl_2$. After 10, 12, and 14 PCR cycles, 30 μl of the reaction mixture was withdrawn and 10 μl was used to visualize the product on a 1.2% agarose gel while another 10 μl that contained the earliest detectable DNA product was used for the next round of binding and PCR selection. After five rounds, the DNA product from the reaction mixture was cloned directly into a pCRII vector (Invitrogen), the clones were subject to DNA sequence analysis, and the DNA sequences from the clones were aligned and analyzed.

Generation of a full-length recombinant TDP-43. To generate a full-length recombinant TDP-43, an *Nco*I site was engineered at the beginning of TDP-43 cDNA as the initiating methionine by PCR. The PCR product was then sequenced and cloned as a partial *Nco*I-*Xho*I fragment into pGEX-KG. Full-length cleaved TDP-43 was then produced by thrombin cleavage as described previously (71). This thrombin-cleaved, bacterially produced full-length TDP-43 was used in UV cross-linking and in vitro transcription experiments.

UV cross-linking with cleaved, full-length recombinant TDP-43. The thrombin-cleaved recombinant TDP-43 was used in UV cross-linking experiments. The experimental conditions were essentially the same as described for UV cross-linking with HeLa nuclear extract, except that 1 μg of poly(dI-dC)-poly(dI-dC) was used as nonspecific competitor. In the experiment with unlabeled oligonucleotide as the competitor, 800 ng of TDP-43 was cross-linked to labeled HIV-1 LTR oligonucleotide from −18 to +28 in the presence of either 50 ng of unlabeled −18 to +28 oligonucleotide or −18 to +28 scrambled oligonucleotide (described above). In another set of experiments, both wild-type Δ1 and Δ2 oligonucleotides were annealed, labeled, and UV cross-linked to 160 and 800 ng of TDP-43. The cross-linked products were resolved by SDS-PAGE (10% polyacrylamide) and visualized by autoradiography.

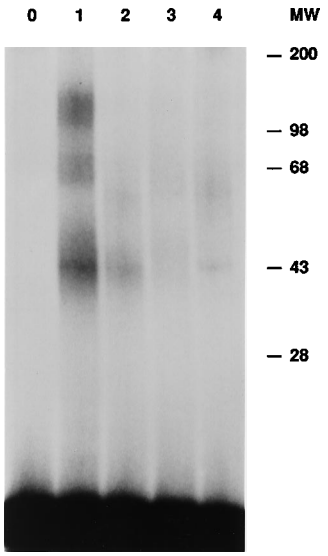

FIG. 1. UV cross-linking of HeLa nuclear extract fractions eluted from a phosphocellulose column. Oligonucleotides corresponding to a region of the HIV-1 LTR extending from −18 to +28 were ^{32}P end labeled and subjected to UV cross-linking in the absence of extract (lane 0) or with 3 μg of HeLa nuclear extract eluted from a phosphocellulose column with either 0.1 M (lane 1), 0.3 M (lane 2), 0.5 M (lane 3), or 1.0 M (lane 4) KCl followed by PAGE and autoradiography. MW, molecular weight (in thousands).

In vitro transcription assays with TDP-43. The HIV-1 template is a 5′ proviral construct cloned into pUC19 and was linearized at the *Spe*I restriction site in the *gag* gene. The *Spe*I-linearized template gives a runoff transcript of about 1,100 bp. Transcription reactions were carried out at 30°C for 1 h in 25 to 50 μl with a final concentration of 10 mM HEPES (pH 7.9), 50 mM KCl, 0.25 mM dithiothreitol, 3 mM $MgCl_2$, 10% glycerol, 0.1 mM EDTA, 40 U of RNasin (Promega), 0.4 mM each ATP, CTP, and UTP, 16 μM GTP, 10 μCi of [α-^{32}P]GTP (added 30 min into the reaction), and 200 ng of linearized template. HeLa nuclear extract was purchased from Promega and used as specified by the manufacturer. Tat protein used in the transcription reaction was cleaved and purified from GST-Tat by a modified procedure as described previously (71) and shown to be active in in vitro transcription with the HIV-1 LTR (54). After the completion of reactions, the transcription was stopped by addition of 450 μl of stop solution containing 7.0 M urea, 0.1 M Tris HCl (pH 7.4), 0.5 M NaCl, 0.5% SDS, and 10 mM EDTA. The transcription product was extracted with phenol-chloroform, precipitated with 1 μg of oyster glycogen and ethanol, and resolved by electrophoresis at 60 W on a 5% denaturing polyacrylamide gel (7 M urea, 1× TBE).

In the first set of reactions, the linearized HIV template was incubated at room temperature for 20 min in the presence or absence of 800 ng of purified TDP-43. After the addition of HeLa nuclear extract and Tat, the reaction mixtures were incubated at 30°C for 30 min, [α-^{32}P]GTP was added, and the reaction was continued for another 30 min at 30°C. In the second set of reactions, 800 ng of TDP-43 was added either 20 min prior to the addition of HeLa nuclear extract and Tat, simultaneously with nuclear extract and Tat, or 30 min after the addition of HeLa nuclear extract and Tat. In all reactions, [α-^{32}P]GTP was added 30 min after addition of HeLa nuclear and Tat. In the third set of reactions, a linearized adenovirus major late promoter template was used as an internal control. Either 0, 160, or 800 ng of TDP-43 was incubated with the HIV-1 and adenovirus major late promoter templates at room temperature for 20 min prior to the addition of HeLa nuclear extract. The reaction mixtures were then incubated at 30°C for 30 min, [α-^{32}P]GTP was added, and the reactions were continued at 30°C for 30 min and then terminated.

Analysis of TDP-43 effects on HIV-1 gene expression. HeLa cells were maintained in complete Iscove's medium supplemented with 5% newborn calf serum, 2.5% fetal bovine serum, and 1% penicillin and streptomycin. One day before transfection, 60-mm plates were seeded with 3×10^5 HeLa cells so that the plates were 50 to 70% confluent at the time of transfection. The lipofectamine transfection method (Bethesda Research Laboratories) was used to cotransfect 1 μg of a wild-type HIV-1 proviral construct (26) with 1 μg of either a Rous sarcoma virus (RSV) based eucaryotic vector expressing full-length TDP-43, a deletion construct comprising amino acids 1 to 95 of TDP-43, exon 1 of the rabbit

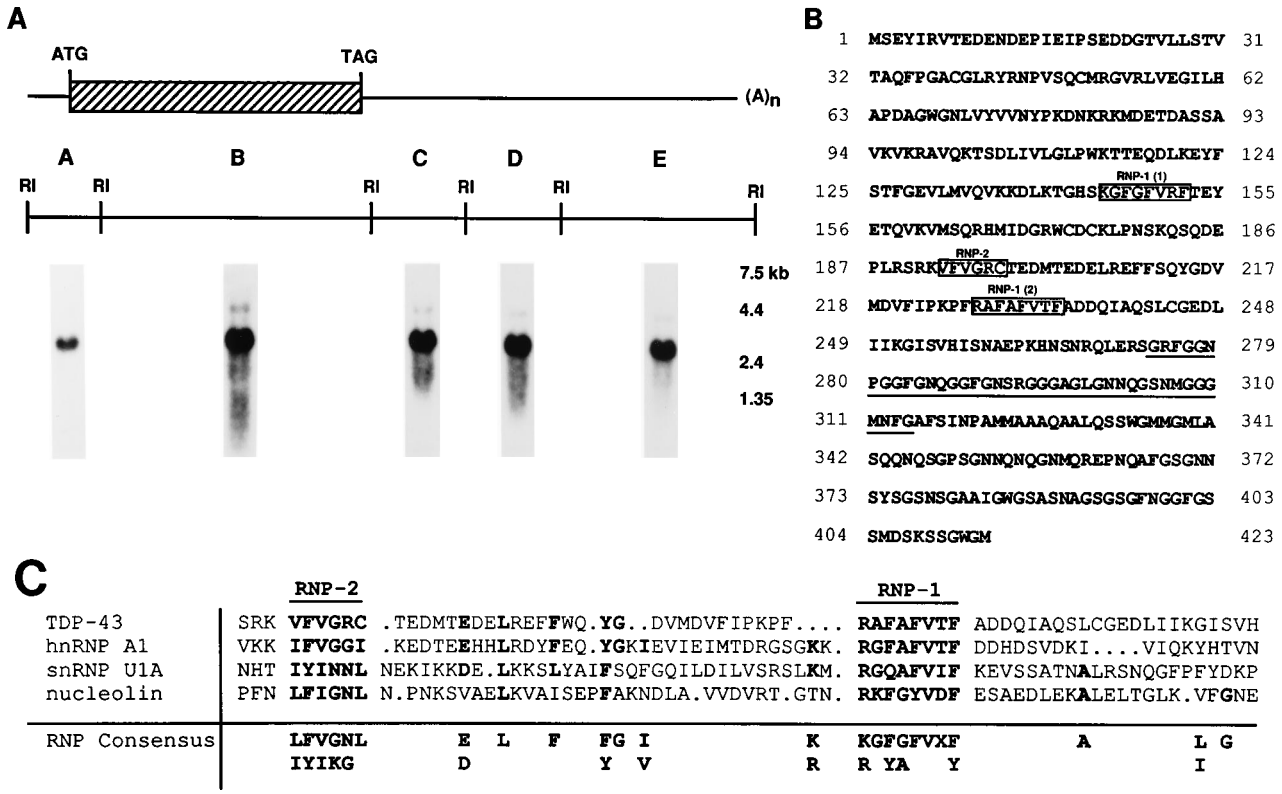

FIG. 2. Structure of TDP-43 DNA. (A) Schematic of the 2.8-kb TDP-43 cDNA which was isolated from a HeLa λgt11 cDNA library. The coding sequence is indicated by hatched lines The five *Eco*RI fragments within the 2.8-kb cDNA are indicated (RI), as are the results of Northern analysis with HeLa RNA with each of these fragments as probes. (B) Amino acid sequence of the TDP-43 protein. The positions of the two RNP-1 motifs and the RNP-2 motif are indicated by boxes, and the glycine-rich region is underlined. (C) Portion of the RNP-binding domain within the TDP-43 protein. Its homology with other proteins which have been demonstrated to contain RNP-binding motifs, including hnRNP A1, small nuclear RNP U1A, and nucleolin, are shown.

β-globin gene, or the parental RSV vector. All transfection mixtures included 1 μg of an RSV–β-galactosidase expression vector to monitor transfection efficiencies. All transfections were performed three times. On day 3 and 6, 0.5 ml of culture supernatant was removed from each transfection and assayed for p24 antigen by the Abbott HIV AG enzyme-linked immunosorbent assay (ELISA) system (26). The cells were fixed and stained for β-galactosidase activity, and the total number of positive cells was determined for each transfection. The p24 antigen levels were normalized to reflect the transfection efficiency.

Nucleotide sequence accession number. The GenBank accession number of the TDP-43 nucleotide sequence is U23731.

RESULTS

Multiple cellular factors bind to TAR DNA. To determine the molecular weight of the different cellular proteins that can potentially bind to TAR DNA, UV cross-linking was performed with labeled oligonucleotides corresponding to a region of the HIV-1 LTR extending from −18 to +28. HeLa nuclear extract was fractionated on a phosphocellulose column and eluted with either 0.1, 0.3, 0.5, or 1.0 M KCl (48). Equal amounts of protein from each of these phosphocellulose fractions were then subjected to UV cross-linking, resolved by SDS-PAGE, and analyzed by autoradiography (Fig. 1). Three species with molecular masses of approximately 110, 65, and 43 kDa were found predominantly in the phosphocellulose A fraction (Fig. 1, lane 1). No detectable UV-cross-linked species were seen when probe alone was used (lane 0). Other UV-cross-linked species were found in lesser quantities in the phosphocellulose B, C, and D fractions (lanes 2 to 4, respectively). These results indicated that multiple cellular factors present in HeLa nuclear extract are capable of binding to TAR DNA.

Cloning of the gene encoding TDP-43. Since several cellular factors were demonstrated to be capable of binding to TAR DNA by UV cross-linking, we wished to isolate cDNAs that encoded one or more of these factors. We used expression cloning with ^{32}P-labeled oligonucleotides corresponding to a portion of the HIV-1 LTR extending from −18 to +28 and probed a λgt11 HeLa cDNA library (63). This technique has been used successfully to identify a number of DNA-binding proteins, including members of the activating transcription factor/cyclic AMP-responsive element-binding protein (ATF/CREB) and the octamer transcription factors (65, 75). One million plaques from this amplified λgt11 library were screened with the labeled oligonucleotide probe and yielded six positive plaques, which were confirmed by secondary and tertiary screening. Use of both wild-type and unrelated oligonucleotides corresponding to different regions in the HIV-1 LTR demonstrated that these phages encoded fusion proteins which bound specifically only to the region of the HIV-1 LTR extending from −18 to +28 (data not shown).

DNA sequence analysis of these six cDNAs indicated that they contained identical sequences in the portion of the cDNA between amino acids 115 and 250, although they contained cDNA inserts of different sizes. The largest cDNA spans 2.8 kb and contains five *Eco*RI fragments, each of which hybridized the same size (2.8 kb) to poly(A)-selected HeLa mRNA in Northern analysis. This result indicated that each of these cDNA fragments was derived essentially from the same cDNA (Fig. 2A). This cDNA contained in-frame stop codons in both its amino and carboxy termini and encoded a protein of 414

amino acids that was designated as TDP-43 (Fig. 2B). A search of GenBank indicated that a region extending from amino acids 270 to 384 in the TDP-43 cDNA had 96% nucleic acid homology with a partial cDNA of unknown function isolated from a random cDNA isolation procedure (2). Also, nucleic acid homology of 100% was noted in a 100-bp region of TDP-43 extending from amino acids 380 to 414 with a gene translocation involved in acute myelogeous leukemia (52). The significance of this latter homology is presently unknown.

A search of Swiss and PIR databases indicated that TDP-43 had 36% amino acid homology with a recently cloned human heterogeneous nuclear ribonucleoprotein (hnRNP) C-like protein (66) and approximately 33% amino acid homology with both a human type A/B hnRNP protein (37) and a murine transcriptional repressor (32). In particular, the highest degree of homology TDP-43 shared with these proteins (32, 37, 66) was in a region of TDP-43 between amino acids 144 and 235 which was related to the previously described RNP-binding domains (1, 3, 18, 36, 49). TDP-43 also shared a high degree of homology in the RNP-binding domain region with similar regions in the hnRNP core protein A1 and nucleolin (8, 10) (Fig. 2C). The RNP domain is found in a variety of RNA-binding proteins and has been demonstrated to be responsible for the RNA-binding properties of these proteins. The RNP motif in these previously described proteins spans approximately 90 amino acids and contains consensus sequences designated RNP-1 and RNP-2 located about 30 amino acids apart (3, 18). The RNP-1 motif is an octapeptide, which is the most highly conserved region within RNP-1 motif, while the RNP-2 motif is a less well conserved hexapeptide sequence. Several other amino acids in the RNP motif are also highly conserved (Fig. 2C).

Within the RNP homology region of TDP-43, there is an unusual arrangement of the RNP-1 and RNP-2 motifs, with two RNP-1 motifs flanking a single RNP-2 motif. Another region of TDP-43, located between amino acids 274 and 314, is composed of a cluster of amino acids of which approximately 40% are glycine residues (Fig. 2B). This glycine-rich domain is similar to other glycine-rich domains which are found in a variety of RNA-binding proteins and may function in nucleic acid strand separation (25) or in mediating protein-protein interactions (13). An amino acid sequence, Arg-Gly-Gly (RGG), which has been implicated in the RNA-binding properties of other proteins is present once within the glycine-rich domain of TDP-43 (38).

Northern analysis was performed with a probe consisting of portion of the TDP-43 cDNA on HeLa and Jurkat poly(A)-selected mRNA as well as mRNA extracted from a variety of human tissues (Fig. 3). TDP-43 hybridized to one major mRNA species of approximately 2.8 kb in both HeLa and Jurkat mRNA (Fig. 3A, lanes 1 and 2) and also hybridized to the same-sized mRNA species prepared from a variety of human tissues (Fig. 3B). The increased amounts of TDP-43 RNA seen in muscle and kidney were due to increased amounts of RNA loaded in these lanes (data not shown). We have also transcribed and translated TDP-43 RNA in vitro in rabbit reticulocyte lysate in the presence of [^{35}S]methionine. Autoradiography revealed that TDP-43 migrated as a 43-kDa protein on SDS-PAGE, consistent with its predicted molecular mass (data not shown). Thus, TDP-43 is a ubiquitously expressed 43-kDa protein whose amino acid structure is similar to that found in a variety of RNA-binding proteins.

Endogenous TDP-43 binds to TAR DNA in HeLa nuclear extract. UV cross-linking indicated that a protein of approximately 43 kDa, present in the 0.1 M KCl fraction of HeLa nuclear extract eluted from a phosphocellulose column, bound

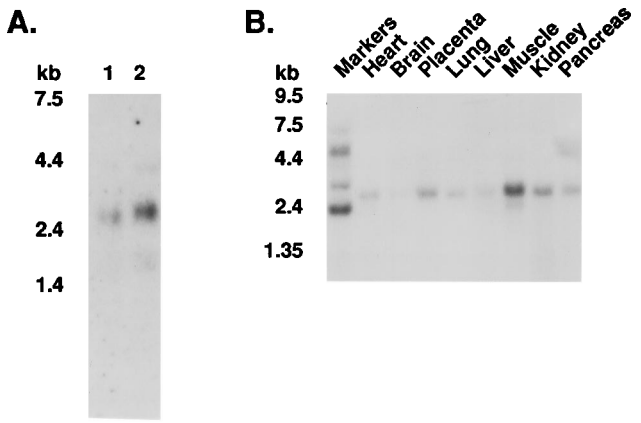

FIG. 3. Expression of TDP-43 mRNA. (A) Northern analysis of TDP-43 was performed with 5 μg of poly(A)-selected Jurkat mRNA (lane 1) or 10 μg of poly(A)-selected HeLa mRNA (lane 2). (B) Northern analysis of TDP-43 was performed with 2 μg of poly(A)-selected RNA isolated from human tissue including heart, brain, placenta, lung, liver, muscle, kidney, and pancreas.

specifically to TAR DNA. To determine if TDP-43 was present in this phosphocellulose column fraction, Western blot analysis was performed with rabbit polyclonal antiserum which was raised against recombinant bacterium-produced TDP-43. The TDP-43 antiserum detected a 43-kDa species present predominantly in the 0.1 M KCl fraction (Fig. 4A, lane 1), with lesser amounts detected in the 0.3 M KCl fraction eluted from the phosphocellulose column (lane 2). The TDP-43 antibody did not detect HeLa nuclear proteins present in the 0.5 and 1.0 M KCl fractions eluted from the phosphocellulose column (lanes 3 and 4). Preimmune rabbit serum did not react with proteins present in any of the phosphocellulose column fractions (Figure 4B).

Next, we wished to determine whether the endogenous TDP-43 protein present in HeLa nuclear extract bound to TAR DNA. The TDP-43 antibody did not result in a supershift of recombinant TDP-43 in gel retardation assays (data not shown), so we used the TDP-43 antibody in a Shift-Western assay to determine whether TDP-43 present in HeLa nuclear extract bound to TAR DNA. This procedure involves gel retardation followed by Western blot analysis to determine the immunologic characteristics of different gel-retarded species (15). Gel retardation analysis was performed with an oligonucleotide probe extending from −18 to +28 in the HIV-1 LTR and HeLa nuclear extract eluted from a phosphocellulose column with 0.1 M KCl, since endogenous TDP-43 was present predominantly in this fraction as determined by Western blot analysis. Three sets of identical gel retardation reactions were carried out. Following electrophoresis, a portion of the polyacrylamide gel containing one of the gel retardation assay mixtures was subjected to autoradiography while the other gel portions containing the remaining two reactions were transferred to nitrocellulose paper. Western blot analysis was performed with rabbit polyclonal antibody directed against TDP-43 on one reaction and preimmune serum on the remaining reaction. The autoradiogram and Western blots were then carefully aligned, as shown in Fig. 4C. Two major gel-retarded species were seen in the autoradiogram of the gel retardation assay (Fig. 4C, lane 1). One of the gel-retarded species comigrated exactly and can in fact be superimposed on the position of a species which was identified by Western blot analysis with TDP-43 antibody after gel retardation (lane 2). This gel-retarded species was not detected when preimmune serum was

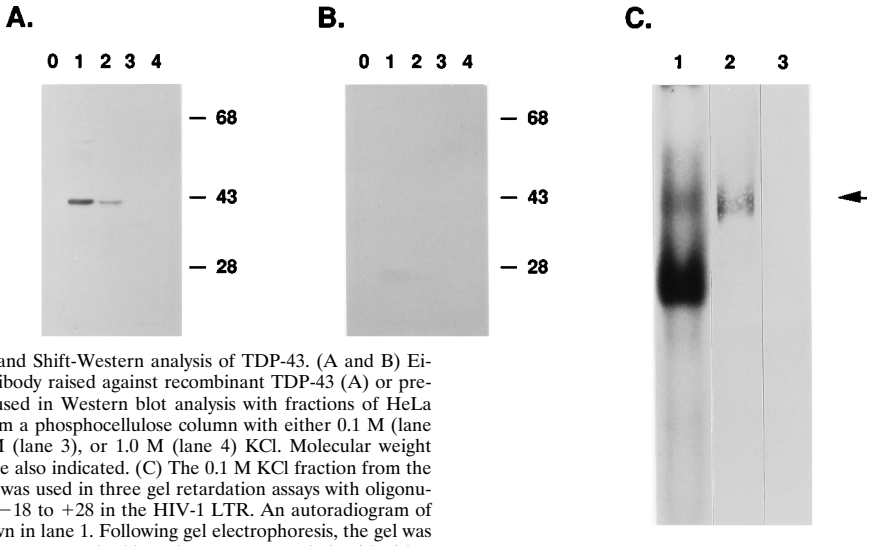

FIG. 4. Western blot and Shift-Western analysis of TDP-43. (A and B) Either rabbit polyclonal antibody raised against recombinant TDP-43 (A) or preimmune serum (B) was used in Western blot analysis with fractions of HeLa nuclear extract eluted from a phosphocellulose column with either 0.1 M (lane 1), 0.3 M (lane 2), 0.5 M (lane 3), or 1.0 M (lane 4) KCl. Molecular weight markers (in thousands) are also indicated. (C) The 0.1 M KCl fraction from the phosphocellulose column was used in three gel retardation assays with oligonucleotides extending from −18 to +28 in the HIV-1 LTR. An autoradiogram of the gel retardation is shown in lane 1. Following gel electrophoresis, the gel was transferred to nitrocellulose paper and subjected to Western analysis with either rabbit polyclonal antiserum directed against TDP-43 (lane 2) or preimmune serum (lane 3). The autoradiogram of the gel retardation assay (lane 1) and the Western blots (lanes 2 and 3) were then carefully aligned.

used in the Western blot analysis (lane 3). These results indicate that endogenous TDP-43 present in HeLa nuclear extract was capable of binding to TAR DNA and that TDP-43 was a likely candidate for the 43-kDa species detected in UV cross-linking analysis with a portion of TAR DNA.

Bacterially produced TDP-43 binds to TAR DNA. A portion of the TDP-43 cDNA encoding amino acids 7 to 414 was cloned downstream of a portion of the GST protein, and the fusion protein was expressed in bacteria and purified by glutathione-agarose affinity chromatography (64). This fusion protein was then used in gel retardation analysis with oligonucleotides corresponding to a region of the HIV-1 LTR extending from −18 to +28 in an attempt to determine the sequence motifs in TAR DNA which bound TDP-43. As shown in Fig. 5A, TDP-43 bound specifically to TAR DNA sequences (Fig. 5A, lane 2). This binding was inhibited by the addition of increasing amounts of unlabeled wild-type oligonucleotides (lanes 3 to 5) but not by a similar excess of unlabeled oligonucleotides corresponding to the adenovirus major late promoter TATA element (lanes 6 to 8). Since a variety of proteins with RNP-binding domains can bind nonspecifically to single-stranded DNA (13, 14, 29, 40, 56), we also tested whether TDP-43 was capable of binding to single-stranded oligonucleotides extending from −18 to +28. TDP-43 was shown to exhibit reduced binding affinity and specificity to these single-stranded oligonucleotides compared with double stranded oligonucleotides (data not shown). Since the structure of TDP-43 was similar to that of other RNA-binding proteins, we also tested whether it could bind specifically to wild-type TAR RNA. No binding to TAR RNA from +1 to +80 was detected by RNA gel retardation analysis under conditions conducive for Tat and TRP-185 binding to TAR RNA, even after prolonged exposure of the gel to film (data not shown). These results were consistent with the fact that the binding of TDP-43 was specific for double-stranded TAR DNA.

Finally, we used recombinant TDP-43 that was cleaved from GST–TDP-43 with thrombin to perform phenanthroline-copper footprinting (Fig. 5B and C) (9, 41). This procedure allows one to directly cleave the gel-retarded complex in the gel and then determine the region of the DNA bound by the protein of interest. A 5′-end-labeled HIV-1 LTR fragment extending from −46 to +80 was labeled on either the coding (Fig. 5B) or the noncoding (Fig. 5C) strand and used in gel retardation with TDP-43. Following the gel retardation, autoradiography was performed on the wet gel and both the free probe and gel-retarded species were excised and treated with phenanthroline-copper. The DNA was then extracted and subjected to electrophoresis on a DNA sequencing gel. TDP-43 bound to both the coding and noncoding strands of the HIV-1 LTR extending over regions from +1 to +36 and −6 and +36, respectively (Fig. 5B and C). These results agree with results of gel retardation and competition analysis, which indicate specific binding of TDP-43 to a region of the HIV-1 LTR extending from −18 to +28.

Determination of the sequences required for TDP-43 binding to TAR DNA. Given the extensive region of TAR DNA that was bound by TDP-43, the nucleotide sequences bound by TDP-43 were determined by the previously described site selection procedure. This technique has been used successfully to determine the binding sites for a variety of proteins including MyoD and myogenin (7, 70). Oligonucleotides containing defined sequences on both their 5′ and 3′ ends and random 35-mer internal sequences were incubated with recombinant TDP-43 under gel retardation conditions. This was followed by five cycles of repeated binding, immunoprecipitation with specific TDP-43 antibody, and PCR amplification of the DNA coimmunoprecipitated with TDP-43 antibody. Twenty-one of these enriched PCR products were subjected to DNA sequencing to help determine optimal consensus binding sites for TDP-43 (Fig. 6). As shown in Fig. 6, the common feature of these PCR products was the presence of at least eight contiguous pyrimidine residues containing alternating groups of cytidine and thymidine. A comparison of these sequences with that of TAR DNA revealed homology with two sets of pyrimidine-rich sequences in the HIV-1 LTR between −15 and −5 and between +4 and +11 (Fig. 7).

Next we determined whether mutations of either or both of these two groups of pyrimidine residues in the HIV-1 LTR designated Δ1 and Δ2 which extend between −15 and −5 and +4 and between +11, respectively, could eliminate the binding of TDP-43 to TAR DNA (Fig. 7). Gel retardation analysis indicated that TDP-43 was inhibited by unlabeled oligonucleotides extending from −18 to +28 in the HIV-1 LTR (Fig. 8A,

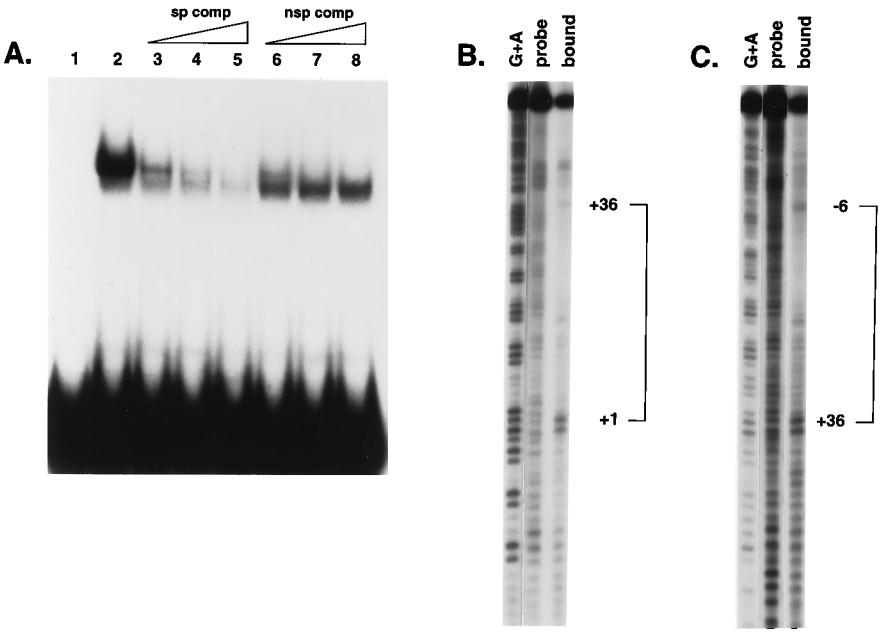

FIG. 5. Gel retardation analysis with recombinant TDP-43. (A) Gel retardation analysis was performed with labeled oligonucleotides corresponding to a region of the HIV-1 LTR extending from −18 to +28 in the presence of no added protein (lane 1) or 1 μg of GST–TDP-43 (lane 2). Competition experiments with either a 5-, 20-, or 50-fold molar excess of wild-type oligonucleotides (lanes 3 to 5) or unrelated oligonucleotides corresponding to the adenovirus major late promoter TATA element (lanes 6 to 8) were performed. sp comp, specific competitor; nsp comp, nonspecific competitor. (B and C) Gel retardation analysis was performed on a region of the HIV-1 LTR extending from −46 (*Eco*RI) to +80 (*Hin*dIII). Either the coding strand (B) or the noncoding strand (C) of the HIV-1 LTR was end labeled, and gel retardation was performed with 2 μg of recombinant TDP-43. Both free and bound probes were subject to cleavage with phenanthroline-copper, and the DNA was subjected to electrophoresis on an 8% sequencing gel next to a G+A sequencing lane of the HIV-1 LTR.

lanes 3 to 5) but not by the same oligonucleotides which contain mutations between −15 and −5 (Δ1) (lanes 6 to 8), +4 and +11 (Δ2) (lanes 9 to 11), or both of these elements (Δ1 + Δ2) (lanes 12 to 14). This indicated that mutation of either of the two pyrimidine-rich sequence motifs between −18 and +28 in the HIV-1 LTR disrupted the binding of TDP-43.

To confirm these findings, oligonucleotides extending from −18 to +28 in the HIV-1 LTR or the same oligonucleotides which contain mutations in either of the two pyrimidine-rich nucleotide sequences were analyzed by UV cross-linking. Wild-type and mutant oligonucleotides were ^{32}P end labeled and incubated with increasing amounts of recombinant full-length TDP-43 which was generated by thrombin cleavage of the GST moiety from the GST–TDP-43 fusion protein (Fig. 8B). The cleaved TDP-43 was approximately 95% pure, and no appreciable degradation products were detected (Fig. 8B, lane 2). As shown in Fig. 8C, wild-type oligonucleotides generated both 43- and 86-kDa UV-cross-linked species whereas oligonucleotides containing mutations in either of two pyrimidine-rich sequences in TAR DNA, Δ1 or Δ2, were unable to generate either the 43- or 86-kDa species (Fig. 8C, lanes 4 to 9). The presence of both these UV-cross-linked species suggested that TDP-43 may dimerize or that two molecules of TDP-43 may simultaneously bind to the pyrimidine-rich sequences between −18 and +28 in the HIV-1 LTR.

UV cross-linking of TDP-43 to TAR DNA. It was important to compare the binding specificity of recombinant TDP-43 and the endogenous approximately 43-kDa protein present in the phosphocellulose A fraction of HeLa nuclear extract detected by UV cross-linking. UV cross-linking with a ^{32}P-end-labeled TAR DNA probe extending from −18 to +28 to recombinant full-length TDP-43 again indicated the binding of two species of approximately 43 and 86 kDa (Fig. 9A, lane 2). Competition with a 50-fold molar excess of unlabeled wild-type competitor

1.	ccgaat	**TCCCTCCA**	ctctta
2.	gtttag	**CCCCTTCC**	aatacc
3.	acatca	**CCCCCTCT**	agttac
4.	cccctg	**CTTTTCTC**	tagaaa
5.	tttttg	**CTTTTCTT**	cttgta
6.	cgaatt	**CCTCTTCC**	tcgtct
7.	actaag	**CTCTTCTA**	gaaagc
8.	tccgaa	**TTCTCCCC**	tacacc
9.	caaggg	**CTCCTTTT**	ttgaca
10.	atgcaa	**TTTCATCT**	tctata
11.	aggata	**TCTTCTTC**	ttccta
12.	accact	**CTCTTTCA**	gctcaa
13.	aaaccg	**CTCCCGCC**	tccccg
14.	atccga	**TTCTTACT**	ccgcac
15.	ctagaa	**TCCCCCTT**	ccacct
16.	gaatag	**CCCCATTT**	tctgga
17.	catagc	**TCTCTTTC**	tgactt
18.	ctggag	**CCTTTCTC**	gaggta
19.	cactgg	**TTTCCCTC**	ccagcc
20.	tgcgaa	**TCCTCTTC**	tagaag
21.	catggg	**TCTTCCTC**	gctagg

FIG. 6. Site selection of oligonucleotides with recombinant TDP-43. Random oligonucleotides with unique 5′ and 3′ termini and 35 random internal nucleotides were incubated with recombinant TDP-43 and immunoprecipitated with rabbit polyclonal antibody directed against TDP-43. PCR analysis was performed with oligonucleotides complementary to the 5′ and 3′ termini followed by five additional cycles of binding, immunoprecipitation, and PCR amplification. The amplified oligonucleotides were cloned into an AT cloning vector (Invitrogen), 21 unique clones were analyzed by DNA sequence analysis, and similar sequences were aligned.

FIG. 7. TAR DNA elements involved in the binding of TDP-43. Oligonucleotides extending from −18 to +28 in the HIV-1 LTR corresponding to either the wild type or mutations of pyrimidine-rich motifs between −15 and −5 (Δ1), between +4 and +11 (Δ2), or in both regions (Δ1 + Δ2) were used in gel retardation analysis with TDP-43.

oligonucleotides eliminated both the 43- and 86-kDa UV-cross-linked species (lane 3), while oligonucleotides mutated in the two pyrimidine sequences (Δ1+Δ2) between −18 and +28 did not compete for the binding of either of these species (lane 4).

We also determined whether the binding specificity of the 43-kDa species detected in HeLa nuclear extract by UV cross-linking was similar to that of recombinant TDP-43. UV cross-linking was again performed with the phosphocellulose A fraction of HeLa nuclear extract and a labeled oligonucleotide probe extending from −18 to +28 in the HIV-1 LTR. Wild-type oligonucleotides competed for each of the three UV-cross-linked species in the HIV-1 LTR phosphocellulose A fraction (Fig. 9B, lane 2) while the same oligonucleotides containing mutations of both the pyrimidine-rich sequences (lane 3) or unrelated oligonucleotides corresponding to either the adenovirus major late promoter TATA element (lane 4) or interleukin-2 promoter sequences (lane 5) did not compete for the binding of the 43-kDa UV-cross-linked species. These results indicated that a protein species of approximately 43 kDa present in the phosphocellulose column A function of HeLa nuclear extract possesses similar DNA-binding activity in UV cross-linking assays as the recombinant full-length TDP-43 protein does.

RNP motifs are important for TDP-43 binding. Next, we wished to determine which domains in the TDP-43 protein were responsible for its binding to the HIV-1 LTR. In particular, we investigated whether either the glycine-rich domain which contains an RGG motif or the RNP domains were critical for TDP-43 binding to DNA. A series of carboxy-terminal truncations of the TDP-43 protein fused to GST were produced in bacteria and purified with glutathione-agarose affinity beads (Fig. 10). A Coomassie blue-stained polyacrylamide gel of these purified fusion proteins is shown along with a schematic of their structure (Fig. 10A and B). These fusion proteins which progressively deleted portions of TDP-43, including the glycine-rich domain, RNP-1 [2], RNP-2, and RNP-1 [1], were used in gel retardation analysis with a labeled oligonucleotide probe extending from −18 to +28 in the HIV-1 LTR. Truncated proteins with truncations which deleted the glycine-rich domain bound like wild-type TDP-43 (Fig. 10C, lanes 1 to 3). Truncated proteins in which most of the RNP-binding domain including RNP-1 [2] and RNP-2 was deleted also bound to the oligonucleotide probe (lanes 4 and 5). It is important to note that deletion of the last two RNP motifs decreased the affinity of TDP-43 binding to TAR DNA (data not shown). Deletions which removed RNP-1 [1] in addition to other amino-terminal residues eliminated the ability of TDP-43 to bind to TAR DNA (lane 6). Furthermore, during our initial λgt11 screening, we isolated a phage which encoded amino acids 110 to 255 of

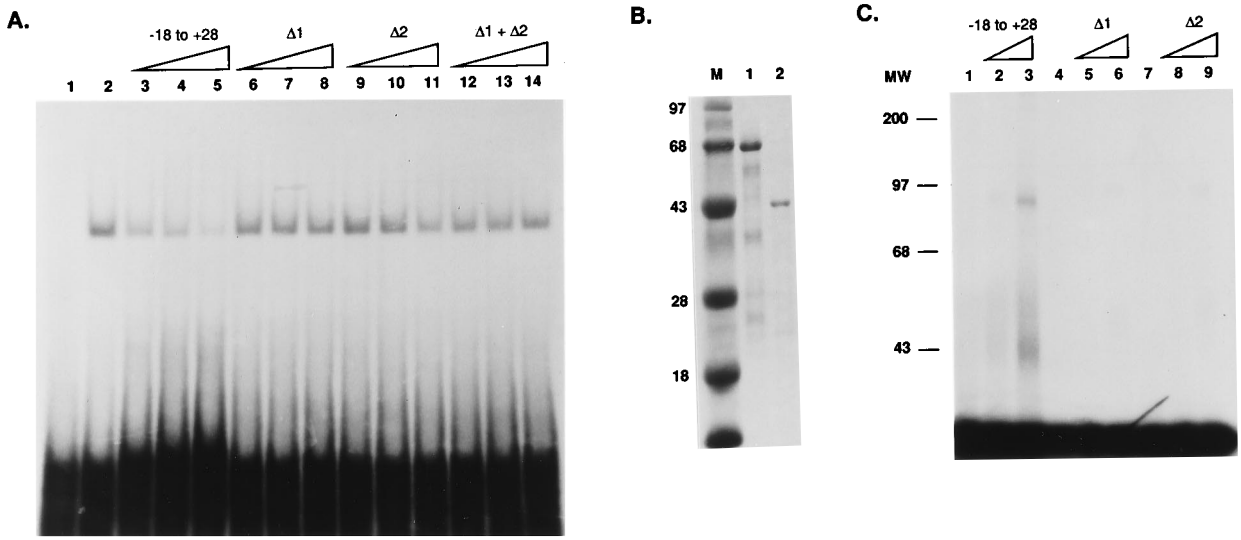

FIG. 8. Gel retardation and competition analysis defines TDP-43-binding sequences. Gel retardation analysis was performed with oligonucleotides corresponding to HIV-1 LTR sequences extending from −18 to +28 in the absence of extract (lane 1), in the presence of GST–TDP-43 alone (lane 2), in the presence of this protein with either a 5-, 20-, or 50-fold molar excess of unlabeled wild-type oligonucleotides (lanes 3 to 5), or in the presence of the same oligonucleotides containing mutations in either pyrimidine-rich sequences Δ1 (lanes 6 to 8), Δ2 (lanes 9 to 11), or both Δ1 and Δ2 (lanes 12 to 14). (B) A GST–TDP-43 fusion protein (lane 1) was produced in bacteria and bound to glutathione-agarose beads, and the native, full-length TDP-43 protein was generated by cleavage of GST–TDP-43 with thrombin (lane 2). Molecular weight markers (lane M) are also indicated (in thousands). (C) UV cross-linking was performed with no added protein (lanes 1, 4, and 7) or either 160 ng (lanes 2, 5, and 8) or 800 ng (lanes 3, 6, and 9) of TDP-43. A ^{32}P-labeled wild-type oligonucleotide probe extending from −18 to +28 in the HIV-1 LTR (lanes 1 to 3) or the same region containing mutations in either pyrimidine stretch Δ1 (lanes 4 to 7) or Δ2 (lanes 7 to 9) was used as the probe in the UV cross-linking. MW, molecular weight (in thousands).

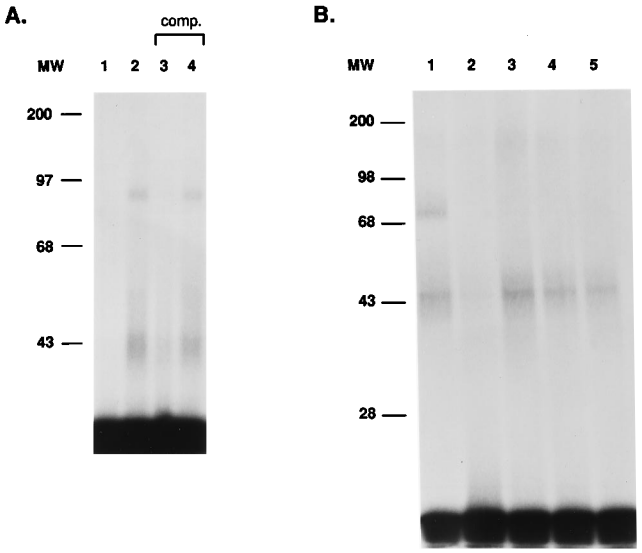

FIG. 9. UV cross-linking of recombinant and endogenous TDP-43. (A) UV cross-linking was performed with a ^{32}P-labeled oligonucleotide extending from −18 to +28 in the HIV-1 LTR in the absence of protein (lane 1), in the presence of 800 ng of TDP-43 alone (lane 2), in the presence of the same protein with a 50-fold molecular excess of unlabeled wild-type oligonucleotide competitor (comp.) (lane 3), or in the presence of the same oligonucleotides but containing mutations in both pyrimidine-rich sequences (lane 4). (B) The 0.1 M KCl phosphocellulose fraction of HeLa nuclear extract was incubated with the oligonucleotide probe used in panel A (lane 1), and competition was performed with 10 ng of either unlabeled wild-type TAR oligonucleotides (lane 2), the same oligonucleotides containing mutations of both pyrimidine-rich sequences (lane 3), major late promoter TATA element oligonucleotides (lane 4), or oligonucleotides corresponding to a region of the interleukin-2 promoter extending from −295 to −275 (lane 5). MW, molecular weight (in thousands).

TDP-43 that bound specifically to TAR DNA (data not shown). These results suggest that a region encompassing the RNP domains is important for TDP-43 binding to TAR DNA, although other amino acid residues may also play a role in its binding.

TDP-43 represses basal and Tat-induced gene expression. Next we determined the role of TDP-43 on HIV-1 gene expression in both the presence and absence of the transactivator protein, Tat. To directly assay the effect of the recombinant TDP-43 protein, in vitro transcription assays were performed. A thrombin-cleaved, full-length TDP-43 protein generated from GST–TDP-43 (Fig. 8B) was added to in vitro transcription assays with a linearized HIV-1 LTR template and HeLa nuclear extract in both the absence and presence of bacterial produced Tat (35, 46). The amount of TDP-43 protein added to the in vitro transcription assays was that used in UV cross-linking of TDP-43 to the HIV-1 LTR (Fig. 8 and 9). The HIV-1 LTR, when linearized with *Spe*I, generated an 1,100-bp runoff product which was inhibited by α-amanitin (data not shown). Tat strongly activated gene expression from the HIV-1 LTR under previously described assay conditions (46) (Fig. 11A, lanes 1 and 2). However, when TDP-43 was preincubated with the HIV-1 LTR template 20 min prior to the addition of Tat and nuclear extract, a marked inhibition of both basal and Tat-induced gene expression was noted (lanes 3 and 4).

Experiments were then performed to determine the mechanism by which TDP-43 repressed in vitro transcription from the HIV-1 LTR. In these experiments, the ^{32}P-labeled ribonucleotide [α-^{32}P]GTP was always added 30 min after the addition of Tat and HeLa nuclear extract and the reaction was stopped 30 min later so that each reaction mixture contained the labeled ribonucleotide for the same time. Similar to the results in Fig. 11A, TDP-43 added 20 min prior to the addition of Tat resulted in little activation by Tat (Fig. 11B, lanes 1 and 2). When TDP-43 was added simultaneously with Tat and nuclear extract, there was a moderate increase in Tat transactivation (lanes 3 and 4). Finally, when TDP-43 was added 30 min after the addition of Tat and nuclear extract, there was a much larger increase in Tat activation (lanes 5 and 6). To demonstrate that the effects of TDP-43 were specific for the HIV-1 promoter, increasing amounts of TDP-43 were incubated in in vitro transcription assays with both HIV-1 and adenovirus major late template (Fig. 11C, lanes 1 to 3). In-

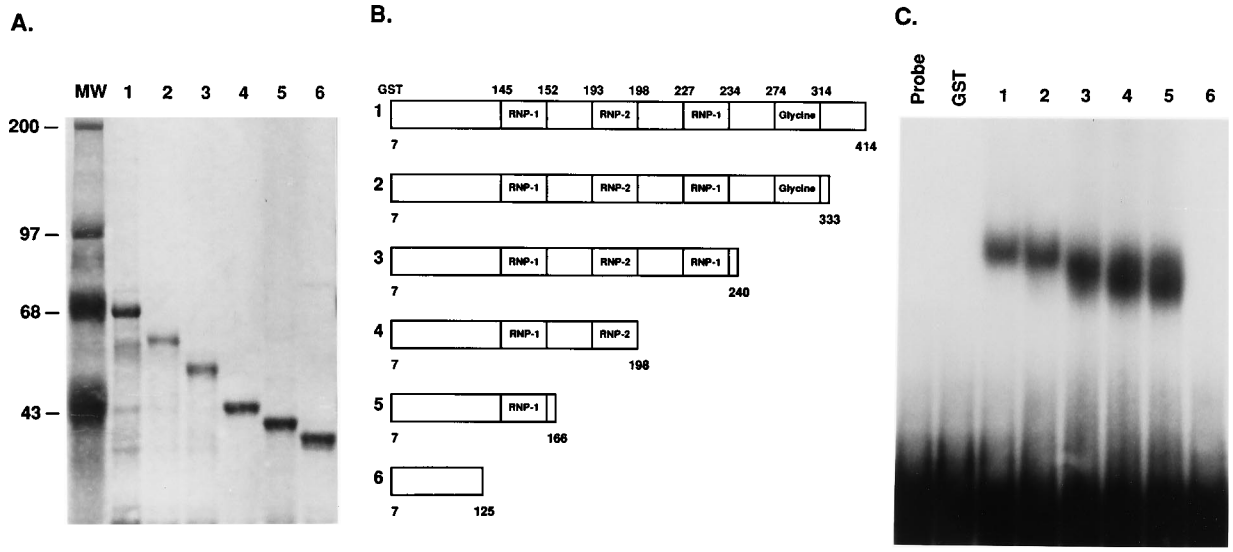

FIG. 10. Analysis of domains in TDP-43 required for its DNA-binding properties. (A) A variety of carboxy-terminal truncations of TDP-43 fused to the GST protein were subjected to PAGE and Coomassie staining (lanes 1 to 6). Lane MW, molecular weight markers (in thousands). (B) Schematic of the carboxy-terminal truncations of TDP-43 shown in panel A. (C) Gel retardation analysis was performed with oligonucleotides corresponding to a portion of the HIV-1 LTR extending from −18 to +28 in the absence of protein (lane Probe), with GST protein (lane GST), or with the GST/TDP-43 fusion proteins used in panels A and B (lanes 1 to 6).

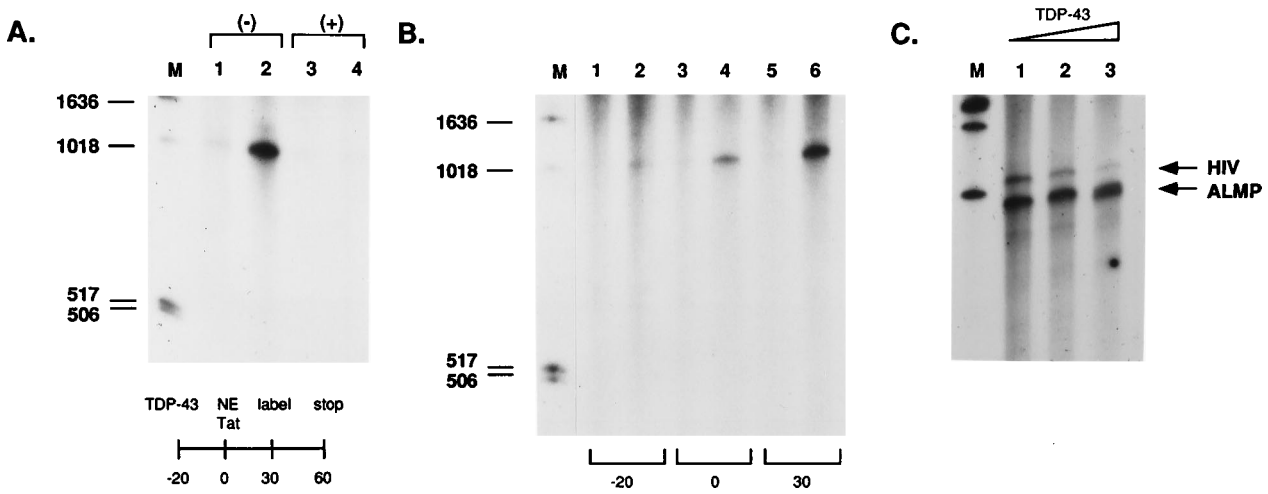

FIG. 11. TDP-43 represses in vitro transcription from the HIV-1 LTR. (A) In vitro transcription of the HIV-1 LTR was performed with a proviral construct linearized with *Spe*I to generate a 1,100-bp runoff product. Transcription from the HIV-1 LTR template in the presence of nuclear extract (NE) alone (lanes 1 and 3) or in the presence of both HeLa nuclear extract and Tat (lanes 2 and 4) is shown. Recombinant TDP-43 protein (800 ng) was added to the HIV-1 template 20 min prior to the addition of either nuclear extract (lane 3) or nuclear extract and Tat (lane 4). [α-^{32}P]GTP was added after 30 min, and the reaction was continued for an additional 30 min. (B) A similar experiment to that in panel A was performed, except that TDP-43 was added to the template at 20 min before the addition of HeLa nuclear extract alone (lane 1) or both HeLa nuclear extract and Tat (lane 2), at the same time as HeLa nuclear extract alone (lane 3) or both HeLa nuclear extract and Tat (lane 4), or 30 min after the addition of nuclear extract alone (lane 5) or both nuclear extract and Tat (lanes 6), [α-^{32}P]GTP was added after 30 min of reaction, and the reaction was continued for an additional 30 min. (C) The HIV-1 template (lane 1) and an adenovirus major late promoter (ALMP) template were used in in vitro transcription analysis either alone (lane 1) or in the presence of increasing amounts of TDP-43 (160 and 800 ng) (lanes 2 and 3), which were added 20 min before the addition of HeLa nuclear extract in the absence of Tat. M indicates the position of the DNA molecular weight markers (in thousands).

creasing the amount of TDP-43 progressively decreased the basal level of transcription from the HIV-1 LTR but not from the adenovirus major late promoter, indicating the specificity of this effect.

TDP-43 represses HIV-1 gene expression in vivo. To determine if TDP-43 could repress HIV-1 gene expression in vivo, TDP-43 was inserted downstream of the RSV promoter in a eucaryotic expression vector. As controls, either a construct that deleted the RNP-binding motifs in TDP-43 (RSV–ΔTDP-43) or sequences encoding exon 1 of the rabbit β-globin gene were cloned into the same vector. HeLa cells were transfected with 1 μg of RSV–TDP-43, RSV–ΔTDP-43, RSV–β-globin, or the parental RSV vector, along with 1 μg of an HIV-1 proviral molecular isolate which has been demonstrated to cause high-level HIV-1 gene expression following transfection (26, 27). Each transfection included RSV–β-galactosidase to control for variations in transfection efficiencies. At 3 and 6 days post-transfection, culture supernatant was removed from each transfection and assayed by ELISA for soluble HIV-1 p24 antigen. After 3 days, RSV–TDP-43 repressed HIV-1 gene expression 16- to 20-fold compared with transfections with RSV–ΔTDP-43, RSV–β-globin, or the RSV vector (Fig. 12). By day 6, RSV–TDP-43 inhibited HIV-1 gene expression approximately fourfold compared with RSV–ΔTDP-43, RSV–β-globin, or the RSV vector (Fig. 11). The β-galactosidase activity indicated that the efficiencies of all transfections were nearly identical. Furthermore, we noted that the transfection of RSV expression vectors encoding a variety of other transcription factors which do not bind to the HIV-1 LTR had no effects on gene expression from the HIV-1 provirus (data not shown). These results indicate that TDP-43 is a potent inhibitor of HIV-1 gene expression.

DISCUSSION

A number of elements in the HIV-1 LTR, including the NF-κB, Sp1, TATA, and TAR elements, are critical for both basal and Tat-induced transcriptional regulation (reviewed in reference 16). Mutagenesis studies have demonstrated the importance of TAR RNA, including the preservation of the stem structure, the loop sequences, and the bulge element on activation by Tat (6, 21, 22, 26, 59). TAR RNA is the binding site for both the Tat protein (11, 17, 68) and a cellular loop-binding protein, TRP-185 (62, 71). Although TAR RNA is critical for modulating Tat activation, the role of TAR DNA on regulating basal and Tat activation remains less clear.

TAR DNA has been demonstrated to bind a number of cellular factors whose function remains to be determined (31, 34, 72, 74). One factor known as UBP-1 or LBP-1 has been purified (34, 72), and cDNAs encoding this family of proteins (LBP-1 a to d) have been cloned (74). UBP-1/LBP-1 is a 64/62-kDa protein which has homology to a group of *Drosophila* transcription factors known as NTF1 (74). In addition to binding to TAR DNA between nucleotides +2 and +6, UBP-1/LBP-1 binds to a lower-affinity site adjacent to the HIV-1 TATA element and prevents the binding of the TFIID complex to the TATA box (34). Both purified and recombinant UBP-1/LBP-1 proteins have been shown to repress basal HIV-1 gene expression in in vitro transcription assays by preventing the binding of TFIID (34). However, the role of these proteins in modulating the level of Tat activation has not been elucidated. Nonetheless, UBP-1/LBP-1 is likely to be involved in negative regulation of HIV-1 gene expression. It should be noted, however, that the UBP-1/LBP-1 family of proteins also contain transcriptional activation domains which are capable of activating gene expression when these proteins are bound to sites located upstream of the TATA box, suggesting that under certain circumstances they are capable of increasing HIV-1 gene expression (74). Whether UBP-1/LBP-1 is capable of activating HIV-1 gene expression when bound exclusively to TAR DNA remains to be determined.

TAR DNA rather than TAR RNA has also been implicated in the generation of short, nonprocessive HIV-1 transcripts

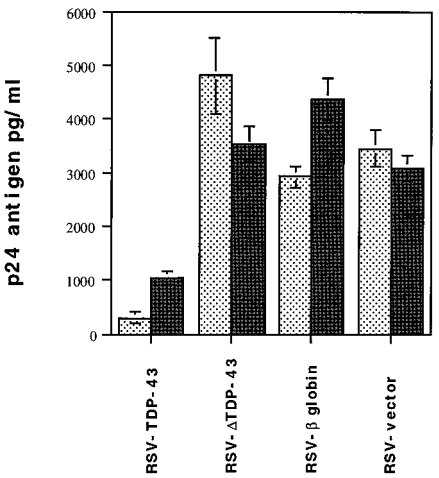

FIG. 12. TDP-43 represses HIV-1 proviral gene expression in vivo. One day prior to transfection, 3×10^5 HeLa cells were seeded onto 60-mm plates. When the plates were 50 to 70% confluent, they were cotransfected with a wild-type HIV-1 proviral molecular clone and an RSV-based eucaryotic vector expressing either TDP-43, ΔTDP-43 that expresses only the first 95 amino acids of TDP-43, intron 1 of rabbit β-globin gene, or the parental expression vector. Each transfection included an RSV–β-galactosidase vector to control for transfection efficiency. On days 3 and 6, 0.5 ml of culture supernatant was assayed for p24 antigen by ELISA, and the results were normalized to reflect the actual transfection efficiency as determined by cell staining for β-galactosidase on day 6. Each transfection was performed three times, and the results include the standard deviation for the three experiments. Symbols: ▨, day 3; ▩, day 6.

(55, 61). This element, known as the inducer of short transcripts, can act as a transcriptional enhancer when it is placed downstream of a variety of heterologous promoter elements and increases the number of short transcripts generated from the promoter (55). Mutagenesis of TAR demonstrated that the region responsible for generation of the short transcripts was not required for Tat activation of HIV-1 gene expression (61). The region in TAR that was responsible for the generation of short transcripts has been localized to a DNA element extending between −4 and +25. The cellular factors that are responsible for this enhancer activity have not yet been determined, although we have demonstrated that TAR DNA mutants that prevent the functioning of the inducer of short transcripts also do not bind TDP-43.

Given the importance of TAR in the regulation of HIV-1 gene expression, it was important to identify additional cellular factors that may modulate its function. UV cross-linking and gel retardation analyses indicated that multiple protein species were capable of binding to TAR DNA. In an attempt to identify these factors, we used a λgt11 expression cloning strategy (63) and identified a cellular factor, TDP-43, that bound specifically to TAR DNA via its RNP-binding motifs and repressed both basal and Tat-induced gene expression. The binding of TDP-43 to TAR DNA was demonstrated by using both recombinant and endogenous proteins present in HeLa nuclear extract. Thus, as with the previously described UBP-1/LBP-1 proteins, we have identified a second cellular factor that binds to TAR DNA and represses HIV-1 gene expression (34, 74). Our analysis indicates that the binding of these proteins to TAR DNA is mutually exclusive. TAR DNA is also one of the sites of nucleosome binding in the HIV-1 LTR (67). Upon phorbol ester activation of latently infected ACH2 cells, HIV-1 gene expression is markedly increased, with concurrent displacement of the nucleosome. Thus, it is likely that either cellular DNA-binding proteins or nucleosomes which bind to TAR DNA can prevent the assembly or elongation of transcription complexes which are the target for activation by the transactivator protein, Tat.

TDP-43 is related to a group of cellular proteins which bind to a variety of different RNA molecules (3, 18, 36, 49). Despite the presence of RNP motifs and an RGG sequence motif in TDP-43, both of which have been shown to confer RNA-binding capability, we were unable to demonstrate the binding of TDP-43 to TAR RNA. However, TDP-43 bound weakly and without specificity to several unrelated single-stranded RNAs. Deletion mutagenesis demonstrated that the RNP motif of TDP-43 was critical for its ability to bind to TAR DNA. Several recent studies have confirmed that proteins containing RNP motifs can bind to double-stranded DNA (40, 66, 73). In fact, it has recently been demonstrated that nucleolin, which contains an RNP-binding domain, can repress the expression of a specific acute-phase response gene by interactions with specific upstream binding sites (73). Interestingly, the DNA-binding motif to which nucleolin binds is composed of alternating pyrimidine residues, although it is not clear whether its RNP motif mediates this binding. Finally, another cellular factor, NSEP-1, which also contains an RNP motif, was also found to bind to polypyrimidine-rich single- and double-stranded DNA sequences in the c-*myc* promoter (39, 40). Thus, a number of proteins containing RNP motifs have been demonstrated to bind to polypyrimidine/polypurine-rich double-stranded DNA sequences. It is possible that the duplicated RNP-1 motifs in TDP-43 each bind to one of the two pyrimidine-rich sequences in HIV-1 DNA, thus helping to explain the requirement for two sets of pyrimidine-rich nucleotides for optimal binding. However, we cannot rule out that two molecules of TDP-43 may simultaneously bind to each of the pyrimidine-rich sequences in TAR DNA.

The in vitro transcription results with TDP-43 indicated that preincubation of TDP-43 with the HIV-1 template is necessary for maximal repression of both basal and Tat-induced gene expression. Simultaneous addition of nuclear extract and TDP-43 resulted in less repression of basal and Tat-induced transcription. Addition of TDP-43 after the addition of HeLa nuclear extract resulted in very minimal amounts of repression of basal and Tat-induced gene expression. In cotransfection experiments with TDP-43 and an HIV-1 proviral clone, TDP-43 repressed HIV-1 gene expression. In this system, there is little or no Tat produced at early times posttransfection, with resulting low levels of basal HIV-1 gene expression. During this early phase, TDP-43 is a potent inhibitor of HIV-1 gene expression, probably by preventing the assembly of transcription complexes that are responsive to Tat. However, at later times following transfection, TDP-43 is a less effective inhibitor of HIV-1 gene expression, suggesting that transcription complexes which displace or compete with TDP-43 binding to the HIV-1 LTR are assembled. These experimental observations suggest that the mechanism of transcriptional repression of HIV-1 gene expression by TDP-43 may be explained by its ability to block the assembly of a transcriptional initiation complex which is competent to respond to Tat. A potential model for the effect of TDP-43 on HIV-1 gene expression is illustrated in Fig. 13.

It is likely that a variety of different cellular factors regulate the assembly of active transcription complexes. The TATA-binding protein and at least eight associated proteins designated TATA-associated factors are capable of binding to the TATA element (50). Recently, it has been shown that a 150-kDa *Drosophila* TATA-associated factor (68) can bind to sequences which are located downstream of the TATA box and potentially function in activating transcription. In addition, another transcription factor, YY1, has recently been shown to

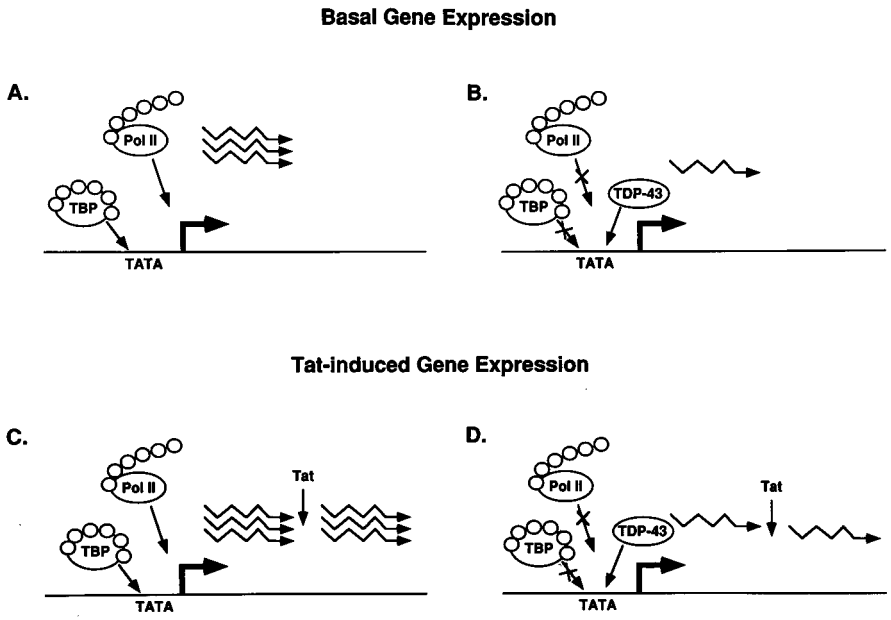

FIG. 13. Potential model of TDP-43 repression of HIV-1 gene expression. The association of the TATA-binding protein (TBP), TATA-associated factors (TAFs), general transcription factors, and RNA polymerase II and associated factors (Pol II) to the HIV-1 promoter is shown. During basal gene expression of the HIV-1 promoter, there is a block to transcriptional initiation and elongation (A) which is relieved by Tat (C). The presence of TDP-43 decreases the number of functional transcription initiation complexes that are assembled by steric effects, resulting in decreased basal (B) and Tat-induced (D) gene expression.

bind to the initiator sequence in HIV-1 LTR (76) and both repress HIV-1 gene expression and delay its growth kinetics (47). One could speculate that the binding of either TDP-43, UBP-1/LBP-1, or YY1 to TAR DNA may prevent the binding of either TATA-binding proteins/TATA-associated factors or general transcription factors to the HIV-1 promoter, thereby precluding the assembly of a transcription initiation complex (19, 54, 58). This will result in defects in the level of both basal and Tat-induced gene expression. The interplay between both positive and negative cellular factors is likely to be critical in the control of HIV-1 gene expression.

ACKNOWLEDGMENTS

We thank Ty Lawrence and Elizabeth Folta for the preparation of the manuscript, Elizabeth Folta for the artwork, and Joe Garcia for initial sequencing and characterization of the different TDP-43 cDNAs.

This work was funded by grants from the NIH and Veterans Administration and the Robert Welch Foundation. S.-H. I. Ou is supported by U. T. Southwestern Medical Scientist Training Program grant GM08014 and the Perot Family Foundation.

REFERENCES

1. **Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff, and G. Dreyfuss.** 1986. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol. Cell. Biol. **6:**2932–2943.
2. **Adams, M. D., M. D. Soares, A. R. Kerlavage, C. Fields, and J. C. Venter.** 1993. Rapid cDNA sequencing from a directionally cloned human infant brain cDNA library. Nat. Genet. **4:**373–380.
3. **Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss.** 1989. RNA-binding proteins as developmental regulators. Genes Dev. **3:**431–437.
4. **Berkhout, B., A. Gatignol, A. B. Rabson, and K.-T. Jeang.** 1990. TAR-independent activation of the HIV-1 LTR: evidence that tat requires specific regions of the promoter. Cell **62:**757–767.
5. **Berkhout, B., and K.-T. Jeang.** 1992. Functional roles for the TATA promoter and enhancers in basal and *tat*-induced expression of the human immunodeficiency virus type 1 long terminal repeat. J. Virol. **66:**139–149.
6. **Berkhout, B., R. H. Silverman, and K. T. Jeang.** 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. Cell **59:**273–282.
7. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250:**1104–1110.
8. **Bourbon, H. M., B. Lapeyre, and F. Amalric.** 1988. Structure of the mouse nucleolin gene. The complete sequence reveals that each RNA binding domain is encoded by two independent exons. J. Mol. Biol. **200:**627–638.
9. **Buratowski, S., M. Sopta, J. Greenblatt, and P. A. Sharp.** 1991. RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. Proc. Natl. Acad. Sci. USA **88:**7509–7513.
10. **Buvoli, M., G. Biamonti, P. Tsoulfas, M. T. Bassi, A. Ghetti, S. Riva, and C. Morandi.** 1988. cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. Nucleic Acids Res. **16:**3751–3770.
11. **Calnan, B. J., S. Biancalana, D. Hudson, and A. D. Frankel.** 1991. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. Genes Dev. **5:**201–210.
12. **Calnan, B. J., B. Tidor, S. Biancalana, D. Hudson, and A. D. Frankel.** 1991. Arginine-mediated RNA recognition: the arginine fork. Science **252:**1167–1171.
13. **Cobianchi, F., R. L. Karpel, K. R. Williams, V. Notario, and S. H. Wilson.** 1988. Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. J. Biol. Chem. **263:**1063–1071.
14. **Cobianchi, F., D. SenGupta, B. Z. Zmudzka, and S. H. Wilson.** 1986. Structure of rodent helix-destabilizing protein revealed by cDNA cloning. J. Biol. Chem. **261:**3536–3543.
15. **Demczuk, S., M. Harbers, and B. Vennstrom.** 1993. Identification and analysis of all components of a gel retardation assay by combination with immunoblotting. Proc. Natl. Acad. Sci. USA **90:**2574–2578.
16. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11:**1475–1489.
17. **Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, and M. A. Skinner.** 1990. HIV-1 tat protein stimulates transcription by binding to a U-rich bulge in the stem of the TAR RNA structure. EMBO J. **9:**4145–4153.
18. **Dreyfuss, G., M. J. Matunis, S. Pinol-Roma, and C. G. Burd.** 1993. hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem. **62:**289–321.
19. **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.
20. **Feinberg, M. B., D. Baltimore, and A. D. Frankel.** 1991. The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation. Proc. Natl. Acad. Sci. USA **88:**4045–4049.
21. **Feng, S., and E. C. Holland.** 1988. HIV-1 tat trans-activation requires the loop sequence within TAR. Nature (London) **334:**165–167.
22. **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.
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. **Gaynor, R.** 1992. Cellular transcription factors involved in the regulation of HIV-1 gene expression. AIDS **6:**347–363.
25. **Ghisolfi, L., G. Joseph, F. Amalric, and M. Erard.** 1992. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. J. Biol. Chem. **267:**2944–2959.
26. **Harrich, D., J. Garcia, R. Mitsuyasu, and R. Gaynor.** 1990. TAR independent activation of the human immunodeficiency virus in phorbol ester stimulated T lymphocytes. EMBO J. **9:**4417–4423.
27. **Harrich, D., C. Yu, E. Race, and R. Gaynor.** 1994. Differential growth kinetics of human immunodeficiency virus type 1 TAR mutant viruses. J. Virol. **68:**5899–5910.
28. **Hauber, J., and B. R. Cullen.** 1988. Mutational analysis of the *trans*-activation-responsive region of the human immunodeficiency virus type 1 long terminal repeat. J. Virol. **62:**673–679.
29. **Ishikawa, F., M. J. Matunis, G. Dreyfuss, and T. R. Cech.** 1993. Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. Mol. Cell. Biol. **13:** 4301–4310.
30. **Jones, K. A., J. T. Kadonaga, P. A. Luciw, and R. Tjian.** 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, SP1. Science **232:**755–759.
31. **Jones, K. A., P. A. Luciw, and N. Duchange.** 1988. Structural arrangements of transcription control domains within the 5′-untranslated leader regions of the HIV-1 and HIV-2 promoters. Genes Dev. **2:**1101–1114.
32. **Kameda, S., and T. Miwa.** 1992. A protein binding to CArG box motifs and to single-stranded DNA functions as a transcriptional repressor. Gene **119:** 229–236.
33. **Kao, S.-Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin.** 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. Nature (London) **330:**489–493.
34. **Kato, H., M. Horikoshi, and R. G. Roeder.** 1991. Repression of HIV-1 transcription by a cellular protein. Science **251:**1476–1479.
35. **Kato, H., H. Sumimoto, P. Pognonec, C.-H. Chen, C. A. Rosen, and R. G. Roeder.** 1992. HIV-1 Tat acts as a processivity factor in vitro in conjuction with cellular elongation factors. Genes Dev. **6:**655–666.
36. **Kenan, D. J., C. C. Query, and J. D. Keene.** 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. **16:**214–220.
37. **Khan, F. A., A. K. Jaiswal, and Z. Szer.** 1991. Cloning and sequence analysis of human type A/B hnRNP protein. FEBS Lett. **290:**159–161.
38. **Kiledjian, M., and G. Dreyfuss.** 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. EMBO J. **11:** 2655–2664.
39. **Kolluri, R., and A. J. Kinniburgh.** 1991. Full length cDNA sequence encoding a nuclease-sensitive element DNA binding protein. Nucleic Acids Res. **19:**4771.
40. **Kolluri, R., T. A. Torrey, and A. J. Kinniburgh.** 1992. A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. Nucleic Acids Res. **20:**111–116.
41. **Kuwabara, M. D., and D. S. Sigman.** 1987. Footprinting DNA-protein complexes in situ following gel retardation assays using 1,10-phenanthroline-copper ion: Escherichia coli RNA polymerase-lac promoter complexes. J. Am. Chem. Soc. **26:**7234–7238.
42. **Laspia, M., A. Rice, and M. Matthews.** 1989. HIV-1 tat protein increases transcriptional initiation and stabilizes elongation. Cell **59:**283–292.
43. **Lu, X., T. M. Welsh, and B. M. Peterlin.** 1993. The human immunodeficiency virus type 1 long terminal repeat specifies two different transcription complexes, only one of which is regulated by Tat. J. Virol. **67:**1752–1760.
44. **Malim, M. H., R. Fenrick, D. W. Ballard, J. Hauber, E. Bohnlein, and B. R. Cullen.** 1989. Functional characterization of a complex protein-DNA-binding domain located within the human immunodeficiency virus type 1 long terminal repeat leader region. J. Virol. **63:**3213–3219.
45. **Marciniak, R. A., B. J. Calnan, A. D. Frankel, and P. A. Sharp.** 1990. HIV-1 Tat protein trans-activates transcription in vitro. Cell **63:**791–802.
46. **Marciniak, R. A., and P. A. Sharp.** 1991. HIV-1 Tat protein promotes formation of more-processive elongation complexes. EMBO J. **10:**4189–4196.
47. **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.
48. **Matsui, T., J. Segall, P. A. Weil, and R. G. Roeder.** 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. **255:**11992–11996.
49. **Mattaj, I. W.** 1989. A binding consensus: RNA-protein interactions in splicing, snRNPs, and sex. Cell **57:**1–3.
50. **Meisterernst, M., and R. Roeder.** 1991. Family of proteins that interact with TFIID and regulate promoter activity. Cell **67:**557–567.
51. **Nabel, G., and D. Baltimore.** 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature (London) **326:**711–713.
52. **Nucifora, G., D. G. Birn, R. Espinosa III, P. Erickson, M. M. LeBeau, D. Roulston, T. W. McKeithan, H. Drabkin, and J. D. Rowley.** 1993. Involvement of the AML1 gene in the t(3;21) in therapy-related leukemia and in chronic myeloid leukemia in blast crisis. Blood **81:**2728–2734.
53. **Olsen, H. S., and C. A. Rosen.** 1992. Contribution of the TATA motif to *tat*-mediated transcriptional activation of human immunodeficiency virus gene expression. J. Virol. **66:**5594–5597.
54. **Ou, S.-H. I., L. F. Garcia-Martinez, E. J. Paulssen, and R. B. Gaynor.** 1994. Role of flanking E box motifs on human immunodeficiency virus type 1 TATA element function. J. Virol. **68:**7188–7199.
55. **Ratnasabapathy, R., M. Sheldon, L. Johal, and N. Hernandez.** 1990. The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters. Genes Dev. **4:**2061–2074.
56. **Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, K. R. Williams, G. Multhaup, K. Beyreuther, H. Werr, B. Henrich, and K. P. Schafer.** 1986. Mammalian single-stranded DNA binding protein UP I is derived from the hnRNP core protein A1. EMBO J. **5:**2267–2273.
57. **Rosen, C. A., J. G. Sodoroski, 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.
58. **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.
59. **Roy, S., M. G. Katze, N. T. Parkin, I. Edery, A. G. Hovanessian, and N. Sonenberg.** 1990. Control of the interferon-induced 68-kilodalton protein kinase by the HIV-1 *tat* gene product. Science **247:**1216–1219.
60. **Selby, M. J., and B. M. Peterlin.** 1990. *trans*-Activation by HIV-1 Tat via a heterologous RNA binding protein. Cell **62:**769–776.
61. **Sheldon, M., R. Ratnasabapathy, and N. Hernandez.** 1993. Characterization of the inducer of short transcripts, a human immunodeficiency virus type 1 transcriptional element that activates the synthesis of short RNAs. Mol. Cell. Biol. **13:**1251–1263.
62. **Sheline, C. T., L. H. Milocco, and K. A. Jones.** 1991. Two distinct nuclear transcription factors recognize loop and bulge residues of the HIV-1 TAR RNA hairpin. Genes Dev. **5:**2508–2520.
63. **Singh, H., J. LeBowitz, A. Baldwin, and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52:**415–423.
64. **Smith, D. B., and K. S. Johnson.** 1988. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene **67:**31–40.
65. **Staudt, L. M., H. Singh, R. Sen, T. Wirth, P. A. Sharp, and D. Baltimore.** 1986. A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. Nature (London) **323:**640–643.
66. **Tay, N., S.-H. Chan, and E.-C. Ren.** 1992. Identification and cloning of a novel heterogeneous nuclear ribonucleoprotein C-like protein that functions as a transcriptional activator of the hepatitis B virus enhancer II. J. Virol. **66:**6841–6848.
67. **Verdin, E., P. Paras, and C. VanLint.** 1993. Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. EMBO J. **12:**3249–3599.
68. **Verrijzer, C. P., K. Yokomori, J.-L. Chen, and R. Tjian.** 1994. Drosophila TAF150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. Science **264:**933–941.
69. **Weeks, K. M., and D. M. Crothers.** 1991. RNA recognition by Tat-derived peptides: interaction in the major groove? Cell **66:**577–588.
70. **Wright, W. E., M. Binder, and W. Funk.** 1991. Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. Mol. Cell. Biol. **11:**4104–4110.
71. **Wu, F., J. Garcia, D. Sigman, and R. Gaynor.** 1991. tat regulates binding of the human immunodeficiency virus trans-activating region RNA loop-binding protein TRP-185. Genes Dev. **5:**2128–2140.
72. **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.
73. **Yang, T. H., W. H. Tsai, Y. M. Lee, H. Y. Lei, M. Y. Lai, D. S. Chen, N. H. Yeh, and S. C. Lee.** 1994. Purification and characterization of nucleolin and its identification as a transcription repressor. Mol. Cell. Biol. **14:**6068–6074.
74. **Yoon, J. B., G. Li, and R. G. Roeder.** 1994. Characterization of a family of related cellular transcription factors which can modulate human immunodeficiency virus type 1 transcription in vitro. Mol. Cell. Biol. **14:**1776–1785.
75. **Yoshimura, T., J. Fujisawa, and M. Yoshida.** 1990. Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent enhancer of HTLV-1: all contain a leucine zipper structure and basic amino acid domain. EMBO J. **9:**2537–2542.
76. **Zenzie-Gregory, B., P. Sheridan, K. A. Jones, and S. T. Smale.** 1993. The HIV-1 core promoter lacks a simple initiator element but contains a bipartite activator at the transcription start site. J. Biol. Chem. **268:**15823–15832.