JOURNAL OF VIROLOGY, July 1992, p. 4488–4496
0022-538X/92/074488-09$02.00/0

Vol. 66, No. 7

Premature Termination and Processing of Human Immunodeficiency Virus Type 1-Promoted Transcripts

MARK KESSLER AND MICHAEL B. MATHEWS*

Cold Spring Harbor Laboratory, 1 Bungtown Road, P.O. Box 100, Cold Spring Harbor, New York 11724-2208

Received 10 December 1991/Accepted 10 April 1992

We have used transient expression assays to study transcription directed by the human immunodeficiency virus (HIV) type 1 promoter. A plasmid containing an HIV-reporter gene fusion and a simian virus 40 origin of DNA replication was transfected into COS-1 cells in the presence or absence of a Tat expression vector. HIV-promoted RNA was analyzed by in vivo labeling, by RNase protection mapping, and in run-on transcription assays. As observed previously, two populations of HIV RNA accumulate in vivo: short, attenuated transcripts and long, polyadenylated mRNA. The short transcripts labeled in vivo were longer and more heterogeneous than expected from RNase protection assays. Moreover, comparison of transcripts labeled in vivo with run-on transcription products revealed that similar, if not identical, short RNAs accumulate in vitro. Utilizing the run-on assay, we show that following transcriptional termination, the attenuated transcripts undergo processing to generate one species of RNA. We also provide evidence that Tat does not act as an antiterminator to relieve a discrete elongation block but instead modifies transcriptional complexes, enabling them to overcome putative pause sites and continue transcription of the template.

The human immunodeficiency virus (HIV) type 1 Tat protein is a potent transactivator of HIV gene expression, capable of greatly stimulating transcription and also implicated in regulating translation (7, 19, 27, 40). Tat acts through the *trans*-activator response element (TAR) situated in the R region of the long terminal repeat (LTR) between residues +1 to +59 relative to the transcription start site (residues +14 to +44 are required for minimal activity [12, 16, 18, 34]). TAR functions at the RNA level in a position- and an orientation-dependent manner (3, 34). RNA encompassing the TAR region folds into a stem-and-loop structure (26) of which three parts are required for transactivation: an intact stem and a 3-nucleotide pyrimidine-rich bulge at +23 to +25, important for Tat binding, and loop sequences (+30 to +35), which can bind cellular proteins (p68 and TRP 185) (3, 8, 11, 13, 18, 31, 32, 34, 42). TAR is dispensable when Tat is brought to the promoter by alternative means (2, 35, 37), suggesting that TAR functions, at least in part, to provide a suitable Tat-binding site.

Tat can modulate transcription by increasing the initiation rate as well as the efficiency of elongation as measured in run-on transcription experiments (10, 20–23). The degree to which Tat stimulates initiation is determined by the basal rate of LTR-directed transcription: Tat stimulates initiation severalfold when the basal rate of transcription is low but has little or no effect on initiation when the basal transcription rate is high (21, 23). The basal transcription rate does not significantly influence Tat's effect on elongation, however. In the absence of Tat, the majority of the transcriptional complexes formed at the HIV promoter do not transcribe beyond the promoter-proximal regions of the template (10, 20–23). In the presence of Tat, although a substantial fraction of the RNA polymerases still appears to stall or disengage within the promoter-proximal region of the transcription unit, the density of complexes in promoter-distal regions of the template increases substantially (10, 23). Tat also stimulates the efficiency of transcriptional elongation in DNA-dependent in vitro transcription systems (24).

Regulation at the level of transcriptional elongation is also reflected in the appearance of two populations of HIV-directed RNAs in the cytoplasm (10, 20, 21, 32, 34). Ribonuclease protection analysis revealed the existence of short, poly(A)$^-$ RNAs terminating at approximately nucleotide +55 to +60, in addition to longer, polyadenylated RNAs representing full-length mRNA (22). The short species predominate in the absence of Tat, whereas the full-length species predominate in its presence. A promoter-proximal element (IST) which induces the generation of short transcripts has been identified, but it is unclear whether the short RNAs are precursors to the long RNAs and whether the two populations are synthesized by similar or distinct transcription complexes (28). Short transcripts terminating in this vicinity (at approximately nucleotides +58 to +61 and +60 to +65) are also generated in cell-free transcription reactions containing partially fractionated nuclear extracts or purified RNA polymerase II (1, 38). These observations suggest that premature termination is an intrinsic property of polymerases initiating on the HIV LTR, especially in the absence of Tat.

Two distinct interpretations of these data have been advanced. In the first, polymerases terminate transcription at or near nucleotide +60, giving rise directly to the short RNAs (20). As a variant of this model, termination might occur at a site some distance downstream, followed by a tightly coupled processing step that yields the mature short transcripts (38). These models are consistent with the in vitro transcription data but are less compatible with other findings, especially with the failure of deletion analysis to detect a termination signal or negative regulatory element in this region. The second interpretation supposes that polymerases terminate at random sites along the gene and that the short RNAs represent stable, processed remnants of these heterogeneously aborted transcripts. The development of a cell-free transcription system that responds to Tat by increased elongational efficiency (24) lends support to this

* Corresponding author.

model. As a variant hypothesis, it is conceivable that the short RNAs might be generated in the cytoplasm as a consequence of mRNA turnover. In either case, the 3′ terminus of the short RNAs might be expected to lie near the base of the TAR stem-loop structure at nucleotide +59.

To extend our understanding of HIV-directed transcriptional elongation, we have directly characterized the short transcripts formed in the presence and absence of Tat. In prior experiments, their presence was inferred from RNase protection analysis of cytoplasmic RNA from transfected cells or from in vitro transcription assays using fractionated nuclear extracts. We demonstrate that the short transcripts that label in vivo are both longer and more heterogeneous than indicated by the RNase protection assays. Moreover, we show that transcripts of a similar size are generated in run-on transcription assays and that the accumulation of at least one of the short transcripts in vivo entails the processing of longer RNA. We also find that Tat does not act as an antiterminator to relieve a discrete elongation block but instead functions to enhance the overall processivity of the transcriptional complex.

MATERIALS AND METHODS

Plasmid and probe construction. The HIV CAT reporter construct, plasmid pSt4, contains the HIV SFII LTR (−643 to +83) driving the chloramphenicol acetyl-transferase (CAT) gene and adjacent to simian virus 40 polyadenylation and splice signals and a simian virus 40 origin of DNA replication (21). The Tat expression vector pRSVTat (pBC12/RSV/t23) (5) was provided by B. Cullen. Plasmid pGEM23, used to synthesize the RNA probe for RNase protection analysis, contains HIV sequences from +83 to −117 under the direction of the SP6 promoter (22). pM13αHIVCAT was prepared by cloning the pU3RIII *Xba*I-*Eco*RI fragment (containing HIV sequences from −643 to +330) into the M13mp18 vector. HIV sense RNA was synthesized by using pEM7, which contains HIV sequences from +1 to +83 under the control of the T7 promoter (15).

Plasmid preparation and transfections. Plasmids were banded twice in CsCl gradients and then treated with RNase A and precipitated with polyethylene glycol. COS-1 cells at 50% confluency (approximately 10^7 cells per 15-cm plate) were transfected by using the DEAE-dextran method (6), with 40 μg of pSt4 reporter plasmid and 30 μg of pRSVTat or carrier.

Labeling and extraction of cellular RNA. At 48 h posttransfection, transfected COS-1 cells (two 15-cm plates per assay) were washed with phosphate-free DMEM (Dulbecco's modified Eagle's medium) and incubated for 12 h with 20 ml of labeling mix containing 10% fetal bovine serum and 5 mCi [^{32}P]phosphate (ICN Radiochemicals) in phosphate-free DMEM. The cells were then washed several times with phosphate-buffered saline (PBS), harvested in 4 ml of PBS, and collected by centrifugation (2 min at 1,500 rpm at room temperature). The cell pellet was resuspended in 1 ml of lysis buffer (50 mM Tris HCl [pH 8.0], 100 mM NaCl, 5 mM $MgCl_2$, 0.5% Nonidet P-40) and incubated on ice for 8 min. After centrifugation for 2 min at 1,500 rpm and 4°C, the nuclear pellet was carefully separated from the cytoplasmic supernatant. Proteinase K (0.2 mg/ml), sodium dodecyl sulfate (SDS) (1%), and EDTA (25 mM) were added to the supernatant. Following incubation for 10 min at 37°C, RNA was extracted with 2 volumes of $CHCl_3$-phenol (1:1) and 1 volume of $CHCl_3$-isoamyl alcohol (24:1) and ethanol precipitated. The nuclear pellet was resuspended in 0.6 ml of CMNT buffer containing 0.3 M NaCl, 30 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.6), 2 mM $CaCl_2$, and 400 U of DNase I (Worthington) and incubated for 5 min at 30°C. Nuclear RNA was isolated by proteinase K treatment and solvent extraction as described above. Both RNA fractions were dissolved in 650 μl of DNase buffer (20 mM HEPES [*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid]-NaOH [pH 7.5], 5 mM $MgCl_2$, 1 mM $CaCl_2$) and incubated with 300 U of DNase I (Worthington) for 30 min at 37°C. The samples were then $CHCl_3$-phenol extracted, $CHCl_3$ extracted twice, and ethanol precipitated.

Run-on transcription assays. Assays were conducted at 72 h posttransfection with two 15-cm plates of transfected COS-1 cells per run-on reaction by using a protocol modified from that of Ucker and Yamamoto (39). Plates of transfected cells were washed three times on ice with 6 ml of A buffer (300 mM sucrose, 10 mM Tris-HCl [pH 8.0], 3 mM $CaCl_2$, 0.5 mM dithiothreitol), permeabilized by treatment with for 2.5 min with 3 mg of digitonin (Sigma) per ml in A buffer, and then washed twice with 6 ml of A buffer. The cells were collected by scraping in 4 ml of A buffer per plate, pelleted by centrifugation for 2 min at 1,500 rpm and room temperature, and then suspended in B buffer (50 mM HEPES-NaOH [pH 7.9], 25% glycerol, 5 mM $MgCl_2$, 0.5 mM dithiothreitol) to 370 μl (total volume). Transcription was started by adding to the cell suspension 400 μl of transcription mix (0.3 M NH_4Cl, 0.1 M HEPES-NaOH [pH 8.0], 8 mM creatine phosphate, 6 mM Mg acetate, 4 mM $MgCl_2$, 4 mM dithiothreitol, 0.8 mM ATP, 0.8 mM CTP, 0.8 mM GTP, 1 μM UTP) and 330 μCi of [α-^{32}P]UTP (3,000 Ci/mM; New England Nuclear). Following 6 min of incubation at 28°C, transcription was stopped by adding α-amanitin to 2 μg/ml, for pulse-labeling, or was chased by adding UTP to 400 μM and incubated for an additional 30 min at 28°C. RNA was isolated according to the protocol of Greenberg (14) with two exceptions: (i) following trichloroacetic acid precipitation, the RNA was treated with 230 U of DNase I (Worthington) for 45 min at 37°C, and (ii) the alkali treatment was omitted.

Hybrid selection of HIV-specific RNA. RNA was dissolved in 0.7 ml of hybridization buffer (70% formamide, 0.3 M NaCl, 10 mM Tris-HCl [pH 7.4], 2 mM EDTA, 0.5% SDS) containing 500 μg of tRNA per ml, heated for 3 min at 85°C, cooled in an NaCl-ice bath, and then added to scintillation vials containing two nitrocellulose filters saturated with 10 μg of single-stranded M13αHIVCAT DNA. Following incubation for 36 to 48 h at 37°C, the filters were washed in hybridization buffer at 37°C for 4 h, with hourly buffer changes. The filters were then rinsed twice with 2 ml of 2 mM EDTA, and HIV-specific transcripts were eluted by heating the filters to 100°C for 1.5 min in 0.8 ml of 2 mM EDTA and cooling them in an NaCl-ice bath. After extraction with $CHCl_3$-phenol and then with $CHCl_3$, the RNA was precipitated with ethanol, suspended in gel loading buffer (90% formamide, 0.5× Tris-borate-EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol), denatured by heating to 100°C for 1 min, and resolved in 6% polyacrylamide–7 M urea sequencing gels.

RESULTS

HIV transcripts of 58 to 70 nucleotides are labeled in vivo. Hitherto, the short HIV-derived transcripts accumulating in vivo have been characterized indirectly, by RNase protection analysis (20–23, 28, 32, 34). Quite apart from the potential artifacts, this assay does not permit precise size estimates. Therefore, in order to visualize the short RNAs

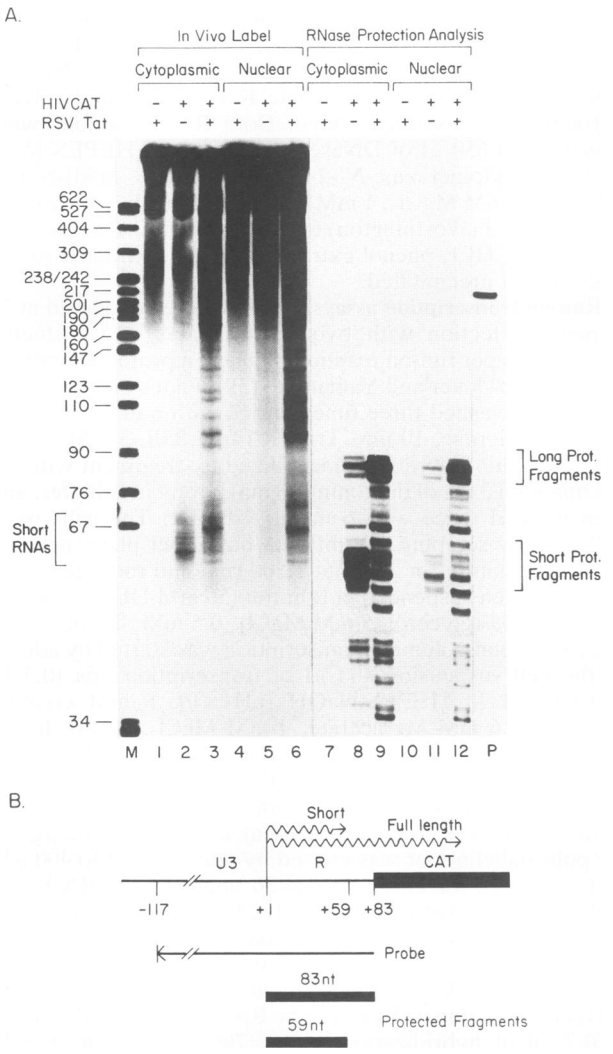
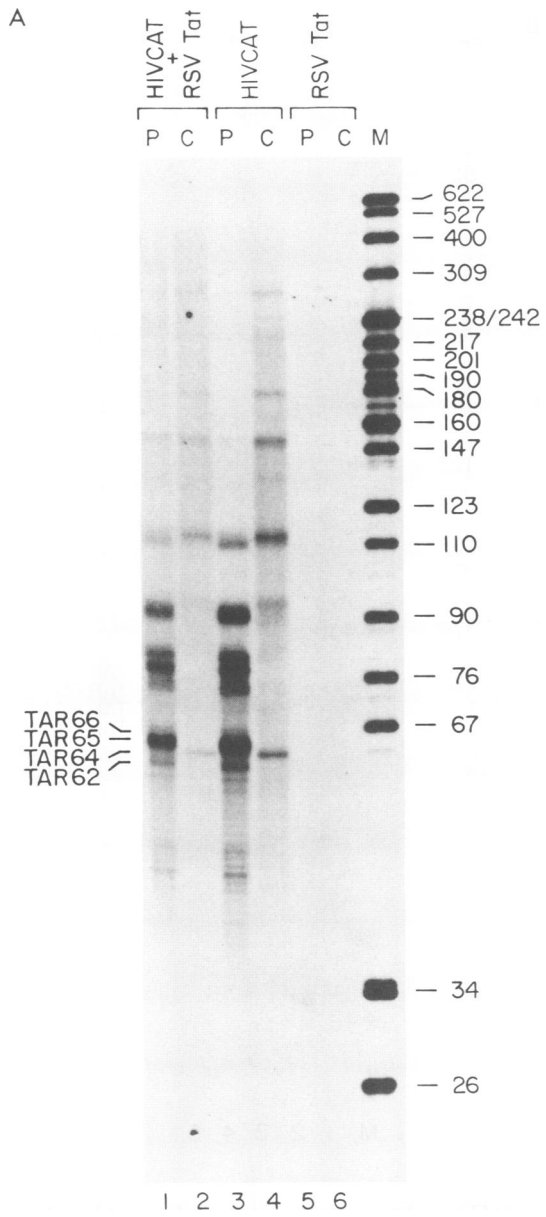


FIG. 1. Analysis of HIV-specific RNA labeled in vivo. (A) COS-1 cells were transfected with pSt4 (HIVCAT), pRSVTat, or both, as indicated at the top of each lane. Cells were labeled with [^{32}P]phosphate from 48 to 60 h posttransfection and fractionated into nucleic and cytoplasmic fractions. HIV-specific RNA, isolated by hybrid selection, was examined directly by gel electrophoresis (lanes 1 to 6) or subjected to RNase protection analysis and then loaded onto the same gel (lanes 7 to 12). Lane M contains pBR322-*Hpa*II markers, whose sizes (in nucleotides) are indicated on the left. Lane P contains the probe used in the RNase protection assays. The positions of the short RNAs labeled in vivo and of the long and short RNase protection fragments are noted. (B) Schematic diagram of the RNase protection analysis depicting the fragments protected by the short RNAs (55 to 60 nucleotides) and full-length RNA (83 nucleotides).

directly and obtain more accurate estimates of their sizes, we labeled cells with [^{32}P]phosphate and isolated RNA by hybridization to HIV DNA immobilized on nitrocellulose filters. Plasmid pSt4, containing the CAT gene under the direction of the HIV LTR, was introduced into COS-1 cells by transfection in the presence or absence of a Tat expression vector (pRSVTat). The hybrid-selected RNAs from both nuclear and cytoplasmic fractions of the transfected cells were displayed in a polyacrylamide gel (Fig. 1A, lanes 1 to 6). Short transcripts migrating near the DNA marker of 67 nucleotides were clearly observed in the cytoplasmic fraction of cells transfected with HIV CAT in the presence or absence of Tat (lanes 2 and 3). Short transcripts were also clearly visible in the nuclei of cells expressing both HIV CAT and Tat (lane 6), but fewer short HIV-encoded RNAs were present in the nuclei of cells transfected with HIV CAT alone (lane 5), suggesting that Tat favors the retention of short transcripts in the nucleus. As expected, there was little or no signal in this region when the HIV CAT plasmid was omitted (lanes 1 and 4). Transcripts of more than about 160 nucleotides were obscured by background in all lanes, but the region between the DNA markers of 76 and 160 nucleotides was well resolved. Transcripts in this size range were abundant in the nuclei and cytoplasm of cells transfected with both plasmids (lanes 3 and 6) but were much less abundant in cells that lacked Tat (lanes 2 and 5) or contained neither plasmid (lanes 1 and 4). Taking these transcripts as representative of HIV mRNA, these observations are consistent with previous data indicating that Tat is required for the formation of long HIV transcripts and that short transcripts are found in the cytoplasm. We also found short transcripts in the nucleus in the presence of Tat.

Detailed examination of the banding pattern of the in vivo-labeled short, HIV-specific RNAs in the gel showed that the distribution of RNAs was influenced by the presence of Tat (compare lanes 2 and 3) but was similar in the nuclei and cytoplasm of cells expressing both HIV CAT and Tat (lanes 3 and 6). We employed a series of homologous and heterologous RNA markers to define the chain lengths of the most abundant species. The markers included TAR RNAs 67 and 83 nucleotides long, synthesized in vitro by T7 RNA polymerase; capped and 5′ monophosphorylated forms of these RNAs; and partial digests of the 5′ end-labeled form of the 67-nucleotide TAR RNA and of adenovirus type 2 VA RNA. By this means we were able to estimate the lengths of the RNA species to within 1 or 2 nucleotides. In the absence of Tat, several RNAs between 57 and 66 nucleotides in length accumulated in the cytoplasm (lane 2). Of these, species of 58 and 59 nucleotides (designated TAR58 and TAR59) were predominant. In Tat-expressing cells (lane 3), transcripts of 56 to 70 nucleotides accumulated cytoplasmically, with a predominant 66-nucleotide transcript (TAR66). The same collection of transcripts was detected in the nuclear fractions of these cells with some strengthening of the TAR65 and TAR66 transcripts (lane 6). Thus, the short HIV-encoded RNAs are distributed over a range of sizes in vivo, and although the population is slightly longer in the presence of Tat than in its absence, it does not differ greatly between nucleus and cytoplasm.

The lengths of these transcripts exceeded, by several nucleotides, the size estimates obtained previously by RNase protection analysis of cytoplasmic RNAs (10, 20, 22, 28, 31, 34). To verify the difference, we annealed portions of the hybrid-selected RNAs to highly radioactive complementary RNA extending from +83 to −117. Following RNase digestion, the protected fragments were resolved in the same polyacrylamide gel (lanes 7 to 12). As with unselected RNA, two populations of protected fragments were obtained: long fragments of about 83 nucleotides representing full-length RNA and short fragments of about 55 to 60 nucleotides representing the truncated RNAs. Thus, the discrepancy does not stem from the hybrid selection procedure itself. Other possibilities include (i) overdigestion of the RNA duplex in the RNase protection assay, (ii) different electrophoretic mobilities of the sense and antisense RNAs, (iii) nontemplated addition of nucleotides to the short transcripts



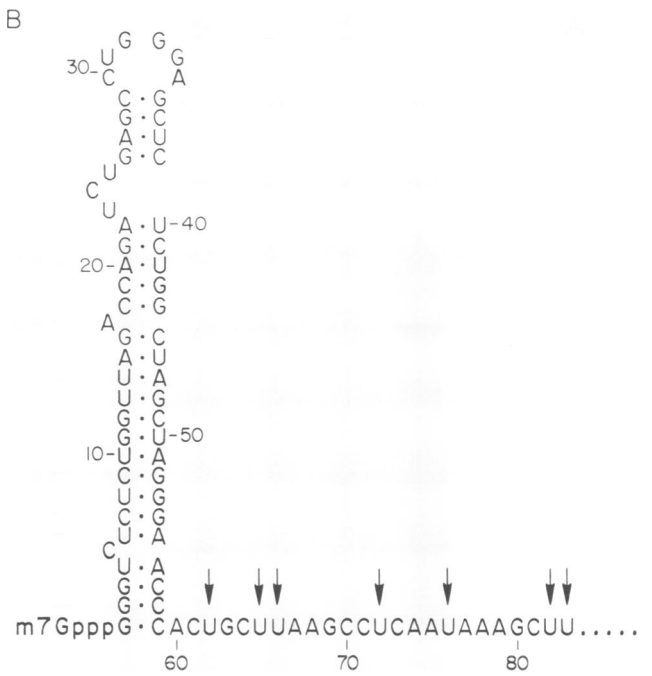


FIG. 2. Size fractionation of HIV-specific run-on transcripts. (A) COS-1 cells were transfected with pSt4 (HIVCAT), pRSVTat, or both, as indicated above the lanes. At 72 h posttransection the nascent transcripts were pulse-labeled for 6 min with [α-^{32}P]UTP (lane P), or pulse-labeling for 6 min was followed by 30 min of chase transcription (lane C). HIV-specific RNA was hybrid selected and resolved in a 6% polyacrylamide–7 M urea sequencing gel. Lane M contains pBR322-*Hpa*II markers, whose sizes (in nucleotides) are indicated on the right. The locations of the TAR62, TAR64, TAR65, and TAR66 run-on transcripts are noted. (B) Predicted RNA secondary structure of the HIV leader sequences. Transcriptional pause sites, based on the gel shown in panel A, are designated.

generated in vivo, and (iv) shortening of the transcripts with time, such that the ^{32}P-labeled RNAs, which are mainly recently synthesized, will be longer than the bulk of the short transcripts which give rise to the protected fragments. In reconstruction experiments with TAR83 synthesized with T7 RNA polymerase, protected fragments both longer and shorter than TAR83 were obtained, perhaps reflecting the distribution and accessibility of the ribonuclease-sensitive sites in the tailed duplex (data not shown). Moreover, there appears to be no significant difference in mobility between complementary strands (data not shown). Although we cannot formally exclude the third possibility, we favor the fourth possibility at present.

Pausing of HIV-promoted complexes at several sites during run-on transcription. In transfected COS-1 cells, complexes traversing the HIV CAT cassette of pSt4 are located predominantly within the first 83 nucleotides of the transcription unit (21). The presence of Tat increases the density of transcriptional complexes in promoter-distal regions without significantly altering the density of complexes within the promoter-proximal region (21). The preponderance of promoter-proximal complexes suggests that elongation is impaired beyond this region, especially in the absence of Tat. To define the precise location of the transcriptional complexes on the template and to determine whether they are accumulating as a result of a block to transcription elongation at a discrete site or sites, we examined the hybrid-selected run-on transcripts in polyacrylamide gels following procedures similar to those used for RNA labeled in vivo.

Several small RNAs accumulated when nascent HIV-promoted transcripts were pulse-labeled with [α-^{32}P]UTP in the absence of Tat (Fig. 2, lane 3), reflecting a predisposition of the transcriptional complexes to stop at specific sites during the run-on transcription reaction. The same small RNAs accumulated in the presence of Tat, as well as longer RNA species represented by the smear of radioactivity in the upper portion of the gel (lane 1). When pulse-labeling was followed by 30 min of chase transcription in the presence of excess unlabeled UTP, the small RNAs that had accumulated during pulse-labeling were absent and were replaced by longer RNA species (lanes 2 and 4), suggesting that the short transcripts represent transcriptional complexes that have paused but not terminated transcription. One 64-nucleotide transcript (designated TAR64) that was not seen in the

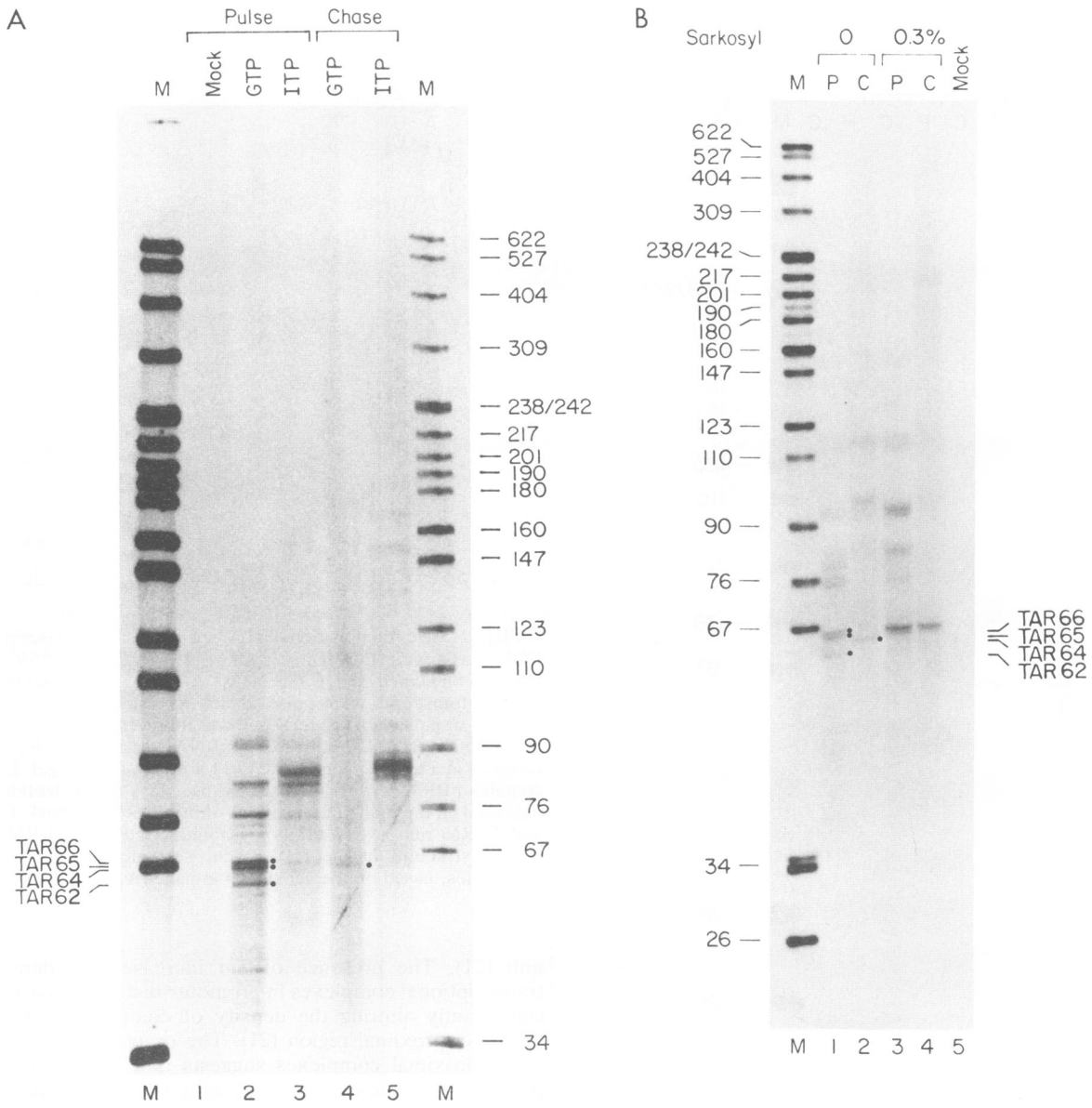


FIG. 3. Run-on transcription conditions influence transcriptional pause sites. (A) ITP modulates pausing at nucleotides +62, +65, and +66 and the processing of TAR64. COS-1 cells were transfected with carrier DNA (lane 1) or with pSt4 (lanes 2 to 5). At 72 h posttransfection nascent transcripts were pulse-labeled for 6 min with [α-^{32}P]UTP (lanes 1 to 3), or pulse-labeling for 6 min was followed by 30 min of chase transcription (lanes 4 and 5). Transcription was carried out in the presence of 400 μM GTP (lanes 1, 2, and 4) or 400 μM ITP (lanes 3 and 5). HIV-specific transcripts were analyzed as for Fig. 2. (B) The effect of 0.3% Sarkosyl on transcription through the promoter-proximal regions of the template. COS-1 cells were transfected with pSt4 (lanes 1 to 4) or carrier DNA (lane 5). Pulse (lanes 1, 3, and 5) and pulse-chase (lanes 2 and 4) transcription were carried out in the absence (lanes 1, 2, and 5) or presence (lanes 3 and 4) of 0.3% Sarkosyl. HIV-specific RNA was analyzed as described above. Lanes M contain pBR322-*Hpa*II markers, whose sizes (in nucleotides) are indicated. The positions of TAR62, TAR64, TAR65, and TAR66 are noted.

pulse-labeled RNAs (lanes 1 and 3) accumulated during the chase reaction (lanes 2 and 4), implying that it is associated with complexes that are very stably paused or, more likely, is a terminated product. None of these RNAs were observed following pulse- or pulse-chase-labeling of nascent RNA in COS cells transfected with only pRSVTat, confirming that they are HIV-promoted RNAs.

Careful sizing of the paused RNAs indicated that most of the short run-on transcripts that accumulated (including TAR62, TAR65, and TAR66, as well as the longer 70- to 83-nucleotide transcripts) correspond to RNAs whose 3′ ends are at U residues within the leader or neighboring CAT sequences (Fig. 2B). Since the probe used to hybrid select the HIV-specific transcripts extends well beyond the TAR region in both directions, it was necessary to show that the short transcripts initiate at the expected site. The 5′ and 3′ termini of TAR64, TAR65, and TAR66 were mapped by eluting the transcripts from the sequencing gel and subjecting them to RNase protection assays with probes extending from residue +83 to +22 or from +83 to −117 but lacking

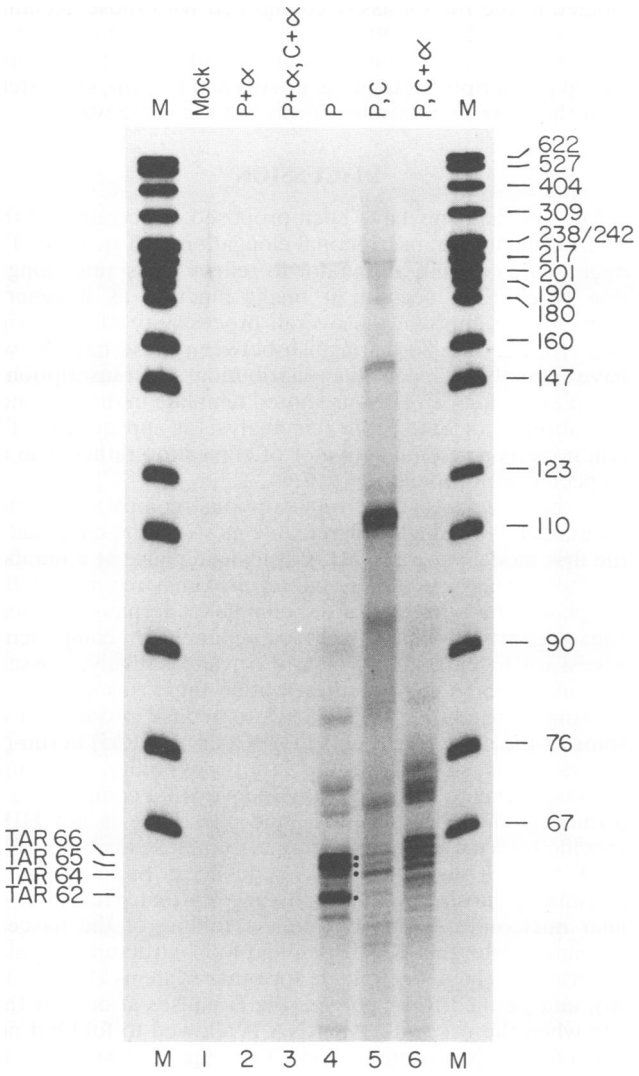


FIG. 4. Effect of RNA polymerase II inhibition on short transcripts. Run-on transcription assays were conducted at 72 h after transfection of COS-1 cells with pSt4 (lanes 2 to 6) or carrier DNA (lane 1). RNA was pulse-labeled for 6 min (lanes 1, 2, and 4), or pulse-labeling for 6 min was followed by 30 min of chase transcription (lanes 3, 5, and 6). The inhibitor α-amanitin was present at a concentration of 2 μg/ml throughout transcription (lanes 2 and 3) or only during chase incubation (lane 6). Lanes M contain pBR322-*Hpa*II markers, whose sizes (in nucleotides) are indicated on the right. The positions of TAR62, TAR64, TAR65, and TAR66 are noted.

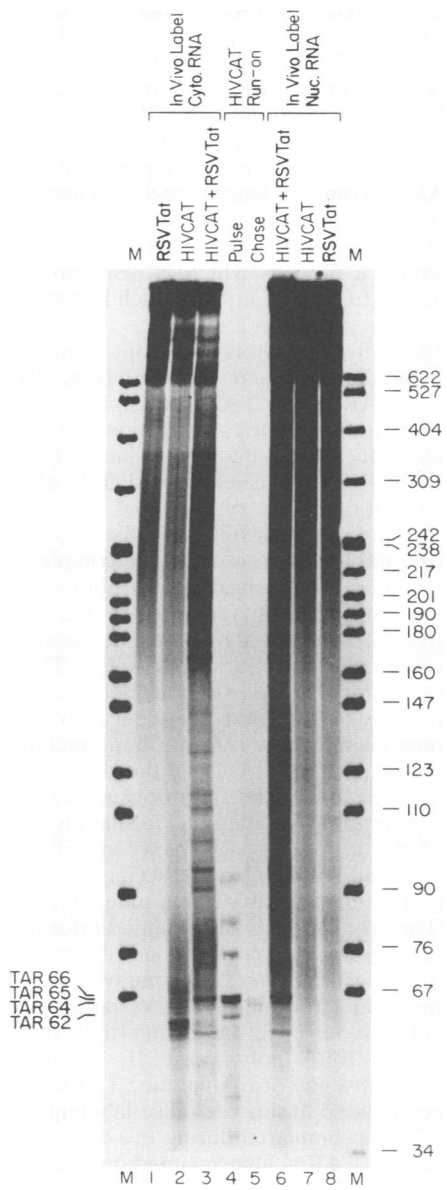


FIG. 5. Comparison of HIV-promoted RNA labeled in vivo and in vitro. COS-1 cells were transfected with pRSVTat (lanes 1 and 8), pSt4 (lanes 2, 4, 5, and 7), or pSt4 and pRSVTat (lanes 3 and 6). Cell labeling and isolation of HIV-specific cytoplasmic (lanes 1 to 3) and nuclear (lanes 6 to 8) transcripts were carried out as described in the legend to Fig. 1. Pulse (lane 4) and pulse-chase (lane 5) run-on transcription were conducted as described in the legend to Fig. 2. Lane M contains pBR322-*Hpa*II markers, whose sizes (in nucleotides) are indicated on the right. The positions of TAR62, TAR64, TAR65, and TAR66 are noted.

residues +35 to +38. On the basis of their measured sizes, the protected products obtained with these two probes were consistent with these transcripts being initiated at nucleotide +1 and terminating at nucleotides +64, +65, and +66, respectively (data not shown).

Short-transcript formation is modulated by transcription complex and nascent RNA structure. The 3′ termini of TAR62, TAR65, and TAR66 mapped to A residues on the template, raising the possibility that these pause sites might be determined, at least in part, by the limiting concentration of UTP present when [α-^{32}P]UTP is used as the label in the run-on assay. Several observations suggest that this is not the sole cause of their pausing. First, when run-on transcription reactions were conducted in the presence of a high concentration of unlabeled UTP and with [α-^{32}P]GTP in place of [α-^{32}P]UTP, the accumulation of TAR62, TAR65, and TAR66 still occurred, albeit to a lesser extent than in reactions containing [α-^{32}P]UTP (data not shown). Second, short-transcript formation was profoundly influenced by substituting ITP for unlabeled GTP in reactions labeled with [α-^{32}P]UTP (Fig. 3A). In the presence of ITP, accumulation

of TAR62, TAR65, and TAR66 was suppressed and transcripts of 80 to 83 nucleotides accumulated. Moreover, in contrast to the GTP-containing reaction (lane 4), the short transcripts formed in the ITP-containing reaction were only elongated by 1 or 2 nucleotides during a 30-min chase incubation with excess unlabeled UTP (lane 5). Transcription on the presence of ITP also suppressed the accumulation of TAR64 (compare lanes 4 and 5). These data suggest that the formation of these short transcripts is influenced by the folding of the nascent TAR RNA into a stable stem-and-loop structure, a process which is disfavored by the ITP analog because I:C base pairs are much less stable than G:C pairs.

Third, the pattern of short-transcript formation was also influenced by the detergent Sarkosyl (Fig. 3B). Reduced quantities of TAR62 and TAR65 accumulated in the presence of 0.3% Sarkosyl, with a corresponding increase in the accumulation of TAR66 (compare lanes 1 and 3). The quantity of TAR66 diminished only slightly after 30 min of chase transcription in the presence of Sarkosyl and unlabeled UTP (lane 4), suggesting that the detergent strongly enhances the stability of transcriptional complexes paused at this site or induces the paused transcriptional complexes to terminate transcription. In contrast to the control chase reaction (lane 2), TAR64 RNA did not accumulate in the chase reaction containing Sarkosyl (lane 4). Thus, the arrest of the transcriptional complexes at nucleotides 64, 65, and 66 is influenced by TAR RNA secondary structure and by protein components of the transcription machinery, as well as by the concentration of UTP in the run-on reaction.

Involvement of RNA processing in short-transcript metabolism. To determine whether the formation of short transcripts includes a processing step, we specifically inhibited RNA polymerase II-mediated transcription with a low concentration of α-amanitin at various times during the run-on reaction. The products that accumulated prior to and following inhibition were compared by gel electrophoresis. No HIV-specific transcripts were detected when α-amanitin was present throughout the pulse (Fig. 4, lane 2) or pulse and chase reactions (lane 3), indicating that the RNAs are synthesized by RNA polymerase II. As previously observed, in the absence of α-amanitin, TAR62, TAR65, and TAR66 accumulated during the pulse-labeling reaction (lane 4) and TAR64 accumulated during the chase reaction (lane 5). Surprisingly, TAR64 also accumulated, and the quantity of TAR65 and TAR66 as well as of other longer transcripts diminished, when α-amanitin was added at the start of the chase reaction (lane 6). The accumulation of TAR64 in the presence of α-amanitin in chase conditions, but not during the pulse, indicates that it is probably formed by a posttranscriptional event. Since TAR64 did not accumulate in reactions that contained α-amanitin from the start of run-on transcription (lanes 2 and 3), it is unlikely to result from RNA polymerase I or III activity or from an alternative labeling reaction such as end labeling. This suggests that TAR64 is derived from a longer transcript(s) that was labeled during the pulse reaction and underwent processing during the chase period.

To relate these findings to observations made in vivo, HIV-specific transcripts labeled in transfected cells and by run-on transcription were directly compared by cofractionation in sequencing gels. RNA labeled in cells expressing (Fig. 5, lanes 3 and 6) or not expressing (lanes 2 and 7) Tat and RNA pulse-labeled in the run-on assay (lanes 4 and 5) gave patterns similar to those seen previously (compare with Fig. 1 and 2). The TAR62, TAR65, and TAR66 species labeled in the run-on assay comigrated with those accumulating in vivo. Since both sets have been shown to initiate at nucleotide +1, we conclude that similar short HIV-promoted transcripts accumulate in vivo and in vitro, consistent with the view that they are formed in the same way.

DISCUSSION

Two mechanisms have been proposed to account for the enhancement of transcriptional elongation by Tat. First, Tat might act as an anti-attenuator to relieve a discrete elongation block (20); second, it might function as a general activator, to improve the overall processivity of transcription (10, 22, 23). To distinguish between these models, we have directly mapped the distribution of transcriptional complexes along an HIV-promoted template in the presence and absence of Tat. From this analysis it appears that Tat functions as a general activator of elongation rather than as a specific antiterminator.

In the absence of Tat, transcriptional complexes do not accumulate at a single discrete site as would be expected if the first model were correct, but instead pause at a number of sites throughout the promoter-proximal regions of the template. The tendency of the complexes to pause at these sites suggests that they lack a component or components necessary to elongate efficiently or, alternatively, possess additional component(s) that render them susceptible to pausing. The ability of Sarkosyl to reduce production of some of the short transcripts (TAR62 and TAR65) in run-on assays lends support to the latter possibility. Although pausing seems to be directed by the way that complexes are formed at the HIV promoter, the pause site is not HIV-specific because the transcriptional complexes pause within CAT as well as within HIV sequences. Suppression of pausing by inosine substitution suggests that pausing at or near nucleotide +60 is due to the folding of the nascent transcript into the TAR stem-and-loop structure. Similar observations have been made for other systems (17, 29, 33, 38), and purified RNA polymerase II pauses at or near this site when the nascent TAR RNA is allowed to fold but not when folding is inhibited (1, 38). The sequences surrounding the downstream pause sites do not share any obvious common primary or secondary structure, however, implying that complexes may pause for a number of different reasons. This view is strengthened by the observation that factors such as the limiting nucleotide, the presence of nucleotide analogs, and the presence of Sarkosyl in the run-on assay modify pausing differentially at the several sites. A simple explanation would be that, because of their inherent elongational inefficiency, these complexes are prone to stop at sites that "normal," elongation-competent complexes would bypass; on this basis, changing the transcription conditions might simply enable the recognition of a different set of putative pause sites.

The appearance of long transcripts when transcription is carried out in the presence of Tat suggests that this viral protein modifies transcriptional complexes, either directly or indirectly, allowing them to transcribe to the 3′ end of the template. Since transcriptional complexes accumulated at the same promoter-proximal sites in the presence or absence of Tat, it seems that Tat does not stabilize all of the complexes formed at the promoter. Similarly, Laspia et al. (23) observed that in run-on assays Tat principally increased the density of transcriptional complexes within the first 800 nucleotides of the transcription unit, and Marciniak et al. (24) found that the addition of purified Tat to in vitro

transcription reactions stimulated transcription in a length-dependent manner. One possible interpretation of these results is that Tat does not stabilize the complexes fully, so that some still tend to pause. Alternatively, it is possible that two classes of complexes are formed in the presence of Tat: one competent in transcriptional elongation and the other incompetent. Consistent with this model, Ratnasabapathy et al. (28) showed that some of the transcription complexes formed on a U2-promoted template containing the TAR region are capable of reading through a U1 3′ termination signal in the presence of Tat, whereas complexes formed in the absence of Tat are not.

What is the relationship between the pausing of transcription complexes that occurs in run-on transcription reactions in vitro, and the accumulation of short transcripts in the cytoplasm of transfected cells? Since most of the short transcripts can be chased in vitro into longer RNAs, it seems that they are paused rather than terminated in the run-on reaction. It is conceivable that this pausing could lead to termination in vivo. TAR64 does appear to be terminated and processed in the run-on reaction: its precursor could be TAR65, TAR66, or something longer. The cytoplasmic RNAs are presumably all terminated, and they appear also to be processed. These RNAs are longer and more heterogeneous than the corresponding RNase protection fragments, suggesting that the transcripts may be shortened with time. Attempts to monitor short-transcript formation in vivo, by labeling the RNA in transfected cells and isolating the HIV-specific transcripts at increasing times after blocking transcription with α-amanitin, did not indicate a specific pattern of processing (data not shown). However, it is conceivable that the processing reaction might be inhibited when α-amanitin is present.

While other interpretations are surely possible, we propose that short-transcript formation is a multistep process. Transcription complexes tend to pause at several sites and (perhaps because they are stalled) to terminate before reaching the end of the transcription unit. Such attenuated transcripts, generated by the premature termination of transcription at heterogeneous sites, undergo processing to generate RNAs terminating just beyond the 3′ base of the TAR stem-and-loop structure (encompassing nucleotides 1 to 59). These RNAs are then processed further to generate the 50- to 55-nucleotide species that are detected in RNase protection assays. The involvement of 3′ processing to generate the short transcripts is consistent with our observation that in digitonin-treated cells, the TAR64 run-on transcript is the product of an RNA-processing event and with the observation that the formation of short HIV-2 transcripts in fractionated nuclear extracts may involve a cotranscriptional processing event (38).

It has been proposed that the short HIV-promoted transcripts accumulating in the cell may serve a biological function. Since these RNAs can be found in the nucleus, at least under some circumstances, it is possible that they play a role in this compartment. For example, as suggested by Ratnasabapathy et al. (28), they might sequester Tat and thereby dampen HIV transcription during latent infection. More attention has been focused on the possibility that they function in the cytoplasm. TAR83 RNA transcribed in vitro has been reported to activate the protein kinase double-stranded RNA-activated inhibitor (DAI) of translation (9, 36). DAI, in its active form, inhibits translation by phosphorylating a protein synthesis initiation factor. DAI synthesis is induced by interferon, and its activation is part of the cellular antiviral defense pathway. In contrast to these reports, however, we find that purified TAR83 inhibits, rather than activates, DAI (15). If the HIV-promoted short transcripts inhibit DAI activity in the infected cell, they could serve to disarm this cellular antiviral mechanism. Both Epstein-Barr virus and adenovirus express small highly structured RNAs that serve similar roles during viral infection (reviewed in reference 25).

ACKNOWLEDGMENTS

We thank Michael Laspia and Nouria Hernandez for valuable suggestions and discussion and Ronnie Packer for cell work.

These studies were supported by NIH grants CA13106, AI27270, and AI31802 and by NIH fellowship AI07877 to M.K.

REFERENCES

1. **Bengal, E., and Y. Aloni.** 1991. Transcriptional elongation by purified RNA polymerase II is blocked at the *trans*-activation-responsive region of human immunodeficiency virus type 1 in vitro. J. Virol. **65**:4910–4918.
2. **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**:7257–7267.
3. **Berkhout, B., and K.-T. Jeang.** 1989. *trans*-activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the *trans*-activating responsive hairpin: a quantitative analysis. J. Virol. **63**:5501–5504.
4. **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.
5. **Cullen, B. R.** 1986. *trans*-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell **46**:973–982.
6. **Cullen, B. R.** 1987. Use of eukaryotic expression technology in the functional analysis of cloned genes. Methods Enzymol. **152**:684–704.
7. **Cullen, B. R.** 1990. The HIV-1 tat protein: an RNA sequence-specific processivity factor? Cell **63**:655–657.
8. **Dingwall, C., I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, M. A. Skinner, and R. Valerio.** 1989. Human immunodeficiency virus 1 *tat* protein binds *trans*-activation-responsive region (TAR) RNA *in vitro*. Proc. Natl. Acad. Sci. USA **86**:6925–6929.
9. **Edery, I., R. Petryshyn, and N. Sonenberg.** 1989. Activation of double-stranded RNA-dependent kinase (dsI) by the TAR region of HIV-1 mRNA: a novel translational control mechanism. Cell **56**:303–312.
10. **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.
11. **Feng, S., and E. C. Holland.** 1988. HIV-1 Tat transactivation requires the loop sequence within TAR. Nature (London) **334**:165–167.
12. **Garcia, J. A., D. Harrich, R. Mitsuyasu, and R. B. Gaynor.** 1988. Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. EMBO J. **7**:3143–3147.
13. **Garcia, J. A., D. Harrich, E. Soultanakis, F. Wu, R. Mitsuyasu, and R. C. Gaynor.** 1989. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. EMBO J. **8**:765–778.
14. **Greenberg, M. E.** 1989. Identification of newly transcribed RNA: nuclear runoff transcription in mammalian cells, p. 4.10.1–4.10.4. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. Green Publishing and Wiley-Interscience, New York.
15. **Gunnery, S., A. P. Rice, H. D. Robertson, and M. B. Mathews.** 1990. HIV-1 TAR RNA can prevent activation of the protein kinase DAI. Proc. Natl. Acad. Sci. USA **87**:8687–8691.
16. **Hauber, J., and B. R. Cullen.** 1988. Mutational analysis of the *trans*-activation-responsive region of the human immunodefi-

ciency virus type 1 long terminal repeat. J. Virol. **62:**673–679.
17. **Hay, N., and Y. Aloni.** 1984. Attenuation in SV40 as a mechanism of transcription-termination by RNA polymerase B. Nucleic Acids Res. **12:**1401–1414.
18. **Jakobovits, A., D. H. Smith, E. B. Jakobovits, and D. J. Capon.** 1988. A discrete element 3′ of human immunodeficiency virus (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV *trans*-activator. Mol. Cell. Biol. **8:**2555–2561.
19. **Jones, K.** 1990. HIV trans-activation and transcription control mechanisms. New Biol. **1:**127–135.
20. **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.
21. **Kessler, M., and M. B. Mathews.** 1991. Tat transcription of the human immunodeficiency virus 1 promoter is influenced by basal promoter activity and the simian virus 40 origin of DNA replication. Proc. Natl. Acad. Sci. USA **88:**10018–10022.
22. **Laspia, M. F., A. P. Rice, and M. B. Mathews.** 1989. HIV-1 Tat protein increases transcription initiation and stabilizes elongation. Cell **59:**283–292.
23. **Laspia, M. F., A. P. Rice, and M. B. Mathews.** 1990. Synergy between HIV-1 Tat and adenovirus E1A is due to stabilization of transcriptional elongation. Genes Dev. **4:**2397–2408.
24. **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.
25. **Mathews, M. B., and T. Shenk.** 1991. Adenovirus virus-associated RNA and translational control. J. Virol. **65:**5657–5662.
26. **Muesing, M. A., D. H. Smith, and D. J. Capon.** 1987. Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. Cell **48:**691–701.
27. **Pavlakis, G. N., and B. K. Felber.** 1990. Regulation of expression of human immunodeficiency virus. New Biol. **2:**20–31.
28. **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.
29. **Resnekov, O., and Y. Aloni.** 1989. RNA polymerase II is capable of pausing and prematurely terminating transcription at a precise location *in vivo* and *in vitro*. Proc. Natl. Acad. Sci. USA **86:**12–16.
30. **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 (*tat-III*, *trs*) of HIV-1. Cell **41:**813–823.
31. **Roy, S., U. Delling, C.-H. Chen, C. A. Rosen, and N. Sonenberg.** 1990. A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat mediated trans-activation. Genes Dev. **4:**1365–1373.
32. **Roy, S., N. T. Parkin, C. Rosen, J. Itovitch, and N. Sonenberg.** 1990. Structural requirements for *trans* activation of human immunodeficiency virus type 1 long terminal repeat-directed gene expression by *tat*: importance of base pairing, loop sequence, and bulges in the *tat*-responsive sequence. J. Virol. **64:**1402–1406.
33. **Seiberg, M., M. Kessler, A. Levine, and Y. Aloni.** 1987. Human RNA polymerase II can prematurely terminate transcription of the adenovirus type 2 late transcription unit at a precise site that resembles a prokaryotic termination signal. Virus Genes **1:**97–116.
34. **Selby, M. J., E. S. Bain, P. A. Luciw, and B. M. Peterlin.** 1989. Structure, sequence, and position of the stem-loop in TAR determine transcriptional elongation by Tat through the HIV-1 long terminal repeat. Genes Dev. **3:**547–558.
35. **Selby, M. J., and B. M. Peterlin.** 1990. *trans*-activation by HIV-1 Tat via a heterologous RNA binding protein. Cell **62:**769–776.
36. **SenGupta, D. N., and R. H. Silverman.** 1989. Activation of interferon-regulated dsRNA-dependent enzymes by human immunodeficiency virus-1 leader RNA. Nucleic Acids Res. **17:**969–978.
37. **Southgate, C., M. L. Zapp, and M. R. Green.** 1990. Activation of transcription by HIV-1 Tat protein tethered to nascent RNA through another protein. Nature (London) **345:**640–642.
38. **Toohey, M. G., and K. A. Jones.** 1989. *In vitro* formation of short RNA polymerase II transcripts that terminate within the HIV-1 and HIV-2 promoter-proximal downstream regions. Genes Dev. **3:**265–282.
39. **Ucker, D. S., and K. R. Yamamoto.** 1984. Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. J. Biol. Chem. **259:**7416–7420.
40. **Vaishnav, Y. N., and F. Wong-Staal.** 1991. The biochemistry of AIDS Annu. Rev. Biochem. **60:**577–630.
41. **Wright, C. M., B. K. Felber, H. Paskalis, and G. N. Paviakis.** 1986. Expression and characterization of the trans-activator of HTLV-III/LAV virus. Science **234:**988–992.
42. **Wu, F., J. Garcia, D. Sigman, and R. Gaynor.** 1991. tat regulates binding of the human immunoefficiency virus trans-activating region RNA loop binding protein TRP-185. Genes Dev. **5:**2128–2140.