

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

SOPHIE ROY,¹ NEIL T. PARKIN,¹ CRAIG ROSEN,² JOSEPH ITOVITCH,¹ AND NAHUM SONENBERG^{1,3*}

Department of Biochemistry¹ and McGill Cancer Center,³ McGill University, Montreal, Canada, H3G 1Y6, and Department of Molecular Oncology, Roche Institute for Molecular Biology, Nutley, New Jersey 07110²

Received 5 September 1989/Accepted 17 November 1989

In order to elucidate the molecular mechanisms of action of the *tat*-responsive sequence, mutational analysis of the *tat*-responsive sequence was carried out. The most critical region comprised nucleotides +18 to +44 and included the 3-nucleotide bulge at positions +23 to +25, the loop sequence, and an intact stem. In addition, base pairing up to nucleotide +52 was required for the full magnitude of the *trans*-activation response. Single-nucleotide bulges at positions +5 to +17 were dispensable. Analysis of truncated and full-length transcripts demonstrated that a transcriptional antitermination model does not fully account for *trans* activation.

Two viral genes encode regulatory proteins essential for human immunodeficiency virus type 1 (HIV-1) replication: *tat* (1, 4, 9, 11, 28, 33, 34) and *rev* (7, 28, 32, 35). Both genes *trans* activate HIV-1 gene expression. The *tat*-responsive sequence (TAR) was originally mapped to the viral long terminal repeat between nucleotides -17 and +80 and thus includes sequences present at the 5' end of all HIV-1 mRNAs (27). Both nucleotide substitutions and 3' deletions were used to delimit the TAR region to nucleotides -17 to +44 (15), +19 to +42 (16), +18 to +44 (10), or +14 to +45 (30). The mRNA transcribed from the TAR region can form a stable stem-and-loop structure in vitro (22), which has been proposed to have functional significance (2, 8, 22, 24).

To assess the requirements for intact base pairing in the stem structure of the TAR region, we made several sets of mutations in predicted duplex regions, including and extending beyond the region previously shown to be important for TAR function (see Fig. 1A). Three sets of 3- or 4-base mutations were introduced by site-directed mutagenesis (41) on either side of the stem and are predicted to decrease the stability of the stem when present alone. The mutations were designed so that when individual mutations are combined with those at the corresponding position on the opposite side of the stem, complete base pairing is restored. This approach allows a clear distinction between changes in primary nucleotide sequence and secondary structure. The mutations were subcloned in the HIV-1 long terminal repeat-driven reporter gene pU3RIII (34). The *trans*-activation activity, measured after transient transfection by the DEAE-dextran method (12), was defined as the chloramphenicol acetyltransferase (CAT) activity obtained in an established HeLa-*tat* cell line (26) compared with that in the parental cell line. Results, expressed as the average of two to five independent experiments, are summarized in Table 1. Mutations predicted to disrupt base pairing resulted in a moderate (TARI, TARI), severe (TARI'), or near-complete (TARIII, TARIII', TARI,III, and TARI',III') loss of *trans*-activation activity,

with the exception of the TARI' mutant, which had a slight stimulatory effect. These results indicate that nucleotides in TAR outside of the boundaries previously described (10, 15, 16, 30) are important for full activity. Possible explanations for this discrepancy are discussed below. In all cases, compensatory mutations predicted to restore base pairing also restored TAR function to wild-type levels. These sets of mutations emphasize the requirement for secondary structure in the TAR region for *trans* activation, especially in mutant TARI,III,I',III', which has 16 nucleotide changes

TABLE 1. Effects of TAR mutations on *trans* activation

Construct	Mutated nucleotides	ΔG_0 of stem-loop structure (kcal/mol) ^a	% <i>trans</i> activation of wild-type ^b (no. of experiments)
TAR (wt)	None	-38.4	100 (6)
TARI	+9 to +12	-26.6	31 (4)
TARI'	+49 to +52	-26.6	8 (5)
TARI,I'	+9 to +12, +49 to +52	-44.8	78 (5)
TARII	+14 to +16	-28.6	30 (4)
TARII'	+45 to +47	-28.6	127 (3)
TARII,II'	+14 to +16, +45 to +47	-38.9	113 (3)
TARIII	+18 to +21	-19.3	0.3 (2)
TARIII'	+41 to +44	-19.3	0.5 (2)
TARIII,III'	+18 to +21, +41 to +44	-38.9	100 (2)
TARI,III	+9 to +12, +18 to +21	-7.5	0.5 (2)
TARI',III'	+41 to +44, +49 to +52	-7.5	0.3 (2)
TARI,III,I',III'	+9 to +12, +18 to +21, +41 to +44, +49 to +52	-45.4	169 (2)
TAR5C	+5 (deleted)	-41.2	100 (1)
TAR17A	+17 (deleted)	-41.2	100 (1)
TAR24UCU	+23 to +25 (deleted)	-42.8	0.3 (2)
TAR31/34	+31 to +34	-38.4	1.0 (2)

^a Calculated according to the energy rules of Salser (29).

^b The average fold *trans* activation was 750.

* Corresponding author.

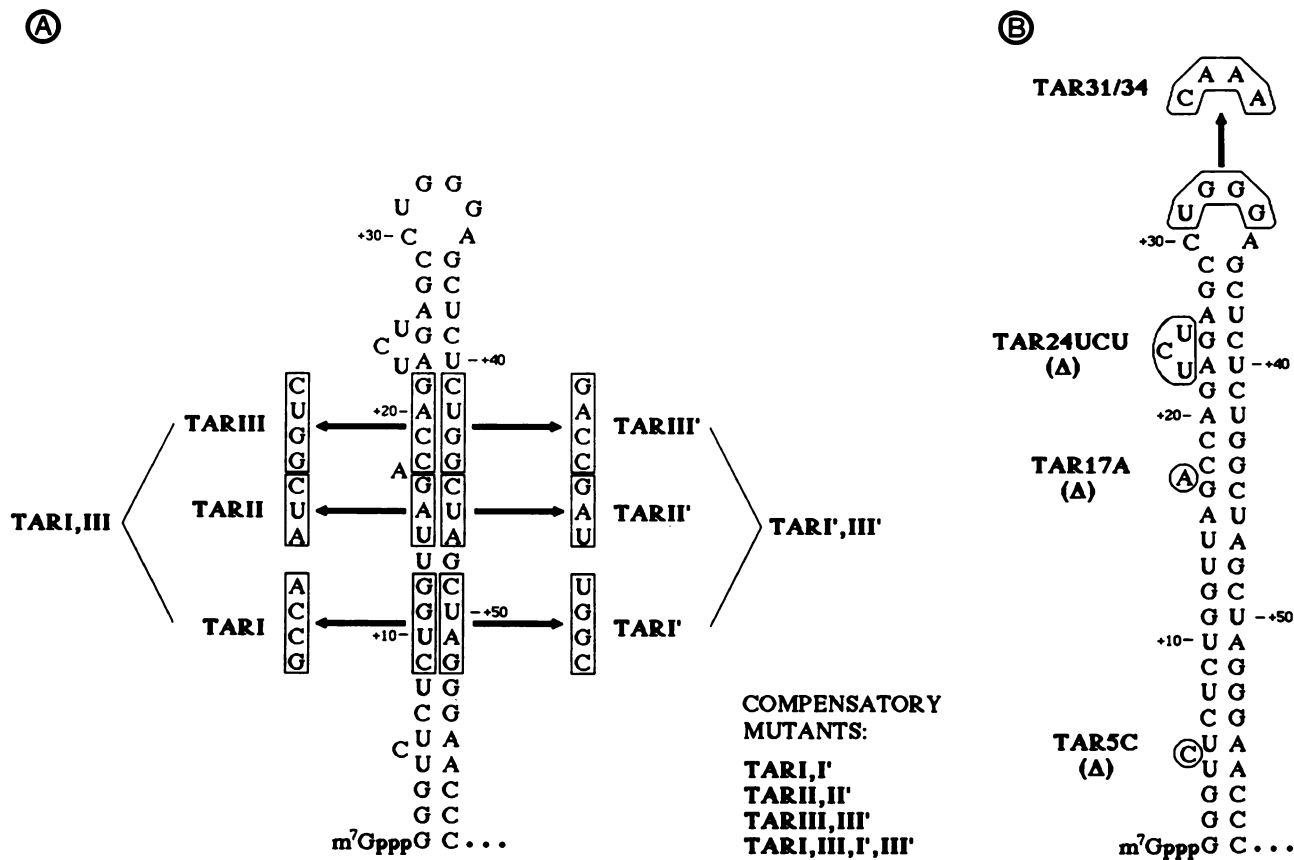


FIG. 1. (A) Secondary structure of the TAR region and stem mutants. Nucleotides shown in boxes were changed to those indicated by site-directed mutagenesis. Mutations were combined by further mutagenesis (TARI,III and TARI',III') or by subcloning, using the *Sac*I site at +34 (all other double mutants). The names of double mutants which restore base pairing in the stem are indicated (bottom, right). (B) Bulge and loop mutants. Bulged nucleotides at positions +5, +17, and +23 to +25 were deleted, and the sequence of the loop at positions +31 to +34 was mutated to the indicated sequence.

but can still be as well *trans* activated by *tat* as the wild-type construct.

One possible explanation for the discrepancy between this report and others is the type of transfection protocol used. All other reports investigating the boundaries of the TAR region (10, 15, 16, 30) have used cotransfection of the reporter gene with a *tat*-expressing plasmid, often in systems which allow efficient replication of the transfected plasmids (15, 30), whereas we have used transfection of the reporter gene into cells stably expressing *tat*. To determine whether there were major differences between the two protocols, we cotransfected the wild-type pU3RIII or mutant TARI, TARI', or TARI,I' with a simian virus 40-based *tat* expression vector, in COS cells. In this system, the mutations had little or no effect (data not shown). However, in cotransfected HeLa cells, in which the plasmid remains in low copy number, there were no qualitative differences in the results obtained compared with those described above, as evidenced by the inhibitory effect of TARI' and the restoration of activity in TARI,I' (data not shown). Thus, it is likely that the copy number of the transfected plasmids has effects on the phenotypes of some TAR mutants. Although unlikely, we cannot exclude the possibility that the mutant phenotypes of TARI and TARI' are species dependent.

A second set of mutations in the TAR region was made to determine the importance of structural features other than

base pairing in the stem (Fig. 1B). Several studies have shown that bulged nucleotides in other RNA stem-and-loop structures are required for protein binding (3, 14, 25, 39). To determine whether such nucleotides are important for TAR function, bulged nucleotides at positions +5, +17, and +23 to +25 were deleted. In addition, the sequence of the loop was mutated (positions +31 to +34). *trans*-Activation activity of these mutants, as assessed by transfection in the HeLa-*tat* cell line, is summarized in Table 1. TAR5C and TAR17A, in which single-nucleotide bulges at positions +5 and +17, respectively, were deleted, had no effect on *trans* activation. However, both TAR24UCU and TAR31/34 had very low activities (less than 5% of wild type) in HeLa-*tat* cells. These results indicate that the 3-nucleotide bulge near the top of the stem is also critical for TAR function. We are currently performing additional mutagenesis in order to better define the requirement for the 3-nucleotide bulge, by changing the sequence of these nucleotides and assaying for *trans* activation. In this context, it is important that conservation of the UCU bulge is not absolute among different isolates of HIV-1, since HIV-1_{BRU} contains UUU instead (38). Also, the TAR region of HIV-2, which contains loop sequences identical to those in HIV-1 (CUGGG), has a bulge of 2 uridine residues in one of the proposed stem-loop structures (8, 13). Such differences may reflect a modulation of the ability of these viruses to respond to *trans* activation

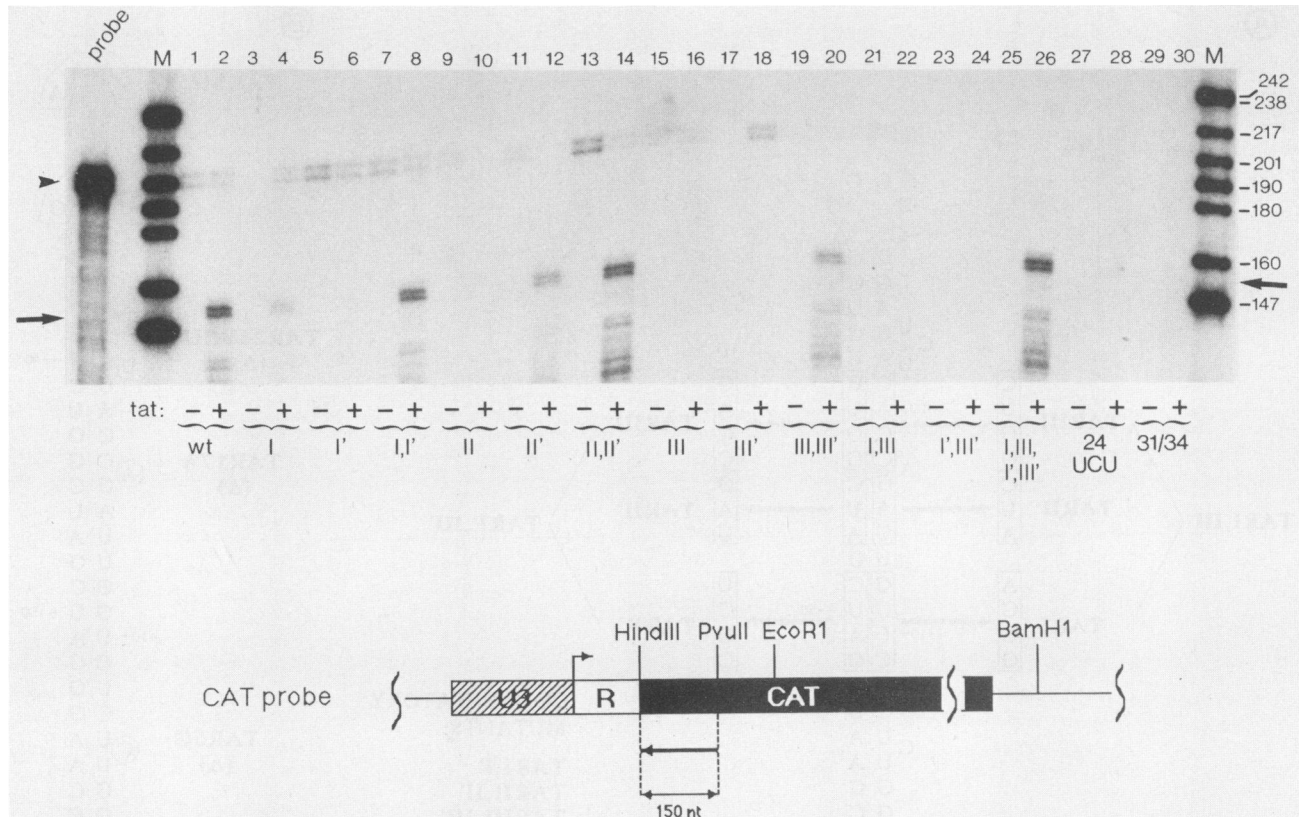


FIG. 2. Quantitative RNase protection analysis of TAR-CAT mRNA in transfected cells. Cells were transfected and 20 μ g of cytoplasmic RNA was analyzed as described in the text. The unprotected probe migrated as a doublet at approximately 200 nucleotides (\blacktriangleright). Lane M, DNA size markers (end-labeled pBR322 cut with *Msp*I); odd-numbered lanes, transfected HeLa cells (-); even-numbered lanes, transfected HeLa-*tat* cells (+). Arrows indicate the position of the protected probe fragment (150 nucleotides). An autoradiogram from a 3-day exposure is shown. The plasmid used for transfection is indicated below the autoradiogram. The position of the probe relative to the structure of the TAR-CAT mRNA is also diagrammed. wt, Wild type.

or may reflect reduced structural or primary sequence requirements within the bulges.

To define the level at which TAR mutations affect gene expression, we analyzed CAT RNA levels in transfected HeLa and HeLa-*tat* cells, using an RNase protection assay (40). Cytoplasmic RNA was isolated by using the Nonidet P-40 lysis procedure (20), including a DNase I digestion step. The probe was derived from the coding portion of the CAT gene (*Hind*III to *Pvu*II fragment, 150 base pairs) and subcloned in antisense orientation in pSP65 (21) and is expected to protect 150 nucleotides of the CAT mRNA from RNase digestion (Fig. 2). The 32 P-labeled probe (approximate specific activity, 6×10^8 cpm/ μ g) was gel purified on a 5% polyacrylamide-urea gel, eluted overnight at 4°C in 0.5 M ammonium acetate-1 mM EDTA-0.1% sodium dodecyl sulfate, extracted with phenol-chloroform, and ethanol precipitated. The optimal hybridization temperature for the CAT probe was determined to be 55°C. Representative results are shown in Fig. 2 (note that the undigested probe and the protected CAT band are doublets, due to some heterogeneity in the probe). All protection assays were standardized by including a probe for the endogenous γ -actin mRNA (6) at the same time as the CAT probe; the signal for γ -actin was observed to be relatively constant between samples (with one exception; see below). We were not able to detect the CAT mRNA in the absence of *tat* (Fig. 2, odd-numbered lanes), presumably because of the very low level of expres-

sion. The signal obtained from HeLa-*tat* cells transfected with most plasmids (Fig. 2, even-numbered lanes) roughly (within a factor of two) paralleled the level of *trans* activation, with the exception of TARI (Table 1). For this construct, the CAT mRNA was barely detectable (Fig. 2, lane 10) although it had only a moderate inhibitory effect on *trans* activation. This discrepancy may be accounted for by a reduction in the amount of RNA used in the assay, since the signal for γ -actin was weaker in this sample than in others (data not shown).

We also examined the effect of TAR mutations on the reported antitermination activity of *tat* (17, 30, 36). This activity can be assayed by comparing the cytoplasmic levels of short RNA species of 59 to 65 nucleotides, suggested to be prematurely terminated transcripts, with those of full-length or read-through transcripts. Figure 3 shows the results of an RNase protection assay, using cytoplasmic RNA from HeLa or HeLa-*tat* cells transfected with wild-type TAR (lanes 1 and 2) or mutants TARI (lanes 3 and 4), TARI' (lanes 5 and 6), and TARI,I' (lanes 7 and 8) and corresponding antisense probes spanning the first 85 nucleotides of the HIV-1 untranslated region (*Pvu*II-*Hind*III fragment of pU3RIII), including the TAR element. The optimal hybridization temperature, for both short and full-length transcripts, was found to be 50 to 55°C. A low level of attenuated transcripts was detectable in HeLa cells transfected with pU3RIII (wild type; Fig. 3, lane 1) or with the double mutant TARI,I' (Fig.

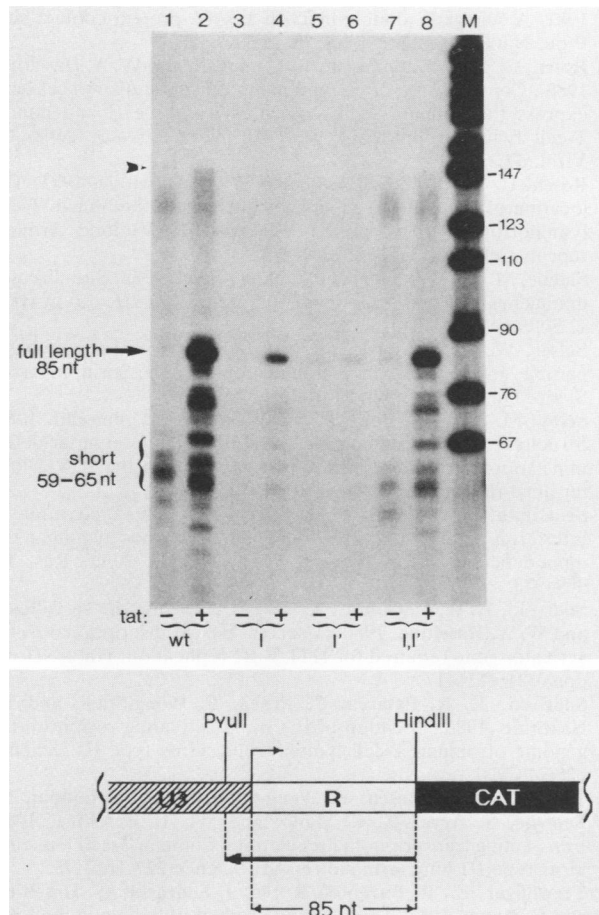


FIG. 3. RNase protection analysis of 5' ends of TAR-CAT mRNAs. The relative abundance of full-length (85-nucleotide) and short (59- to 65-nucleotide) transcripts in cytoplasmic RNA preparations from HeLa (-) or HeLa-*tat* (+) cells transfected with pU3RIII (wild type [wt]) or mutant TARI, TARI', or TARI,I' was determined by using homologous RNA probes spanning the transcriptional start site (see diagram at bottom). A 10- μ g amount of cytoplasmic RNA was used in the assay. Protected fragments were resolved on an 8% polyacrylamide-urea gel; an autoradiogram from a 7-day exposure is shown. Lane M, *Msp*I-digested, end-labeled pBR322. The migration position of the undigested probe is indicated (\blacktriangleright).

3, lane 7). However, these short RNA species were not detected in cells transfected with TARI (Fig. 3, lane 3) or TARI' (Fig. 3, lane 5), indicating that secondary structure in this region is important for the generation and/or stability of the short transcripts. In HeLa-*tat* cells, full-length transcripts, whose relative amounts parallel those obtained by using the CAT probe, are readily seen (Fig. 3, even-numbered lanes). It is important, however, that there was an overall increase in the total amount of RNA detected in HeLa-*tat* cells compared with that in HeLa cells, with a preferential up-regulation in the expression of full-length transcripts. Therefore, unless there were major differences in the stability of the short transcripts in HeLa cells, relative to that of the full-length transcripts in HeLa-*tat* cells, the antitermination mechanism does not account for the full effect of *tat* in this system.

Since short transcripts were not detected in cells transfected with mutant TARI or TARI' in the absence of *tat* and

were barely detectable in the presence of *tat*, it is possible that these mutations affect the termination site, the transcription initiation efficiency, or the stability of terminated RNAs. One interesting possibility, for which evidence is being sought, is based on the recent observation that induction of beta interferon transcription by double-stranded RNA is mediated in part by NF- κ B (19, 37). If the double-stranded nature of the TAR region could also induce NF- κ B activity, perhaps via the double-stranded RNA-dependent p68 kinase (5, 31, 37), then this mechanism may contribute to HIV long terminal repeat-directed transcription, since the long terminal repeat contains two NF- κ B-binding sites (23). The mutant RNAs TARI or TARI', however, would be expected to be less potent activators of NF- κ B and therefore to be transcribed less efficiently (Fig. 2 and 3).

This study, designed to analyze the structural requirements for TAR function, suggests that responsiveness to *tat* requires the bulge at positions +23 to +25, the loop sequence, and proper base pairing. The effects of mutations in nucleotides +9 to +12 and +49 to +52, although not as dramatic as those within the upper portion of the stem, are nevertheless significant. The importance of the lower region may not have been previously realized because it appears to be dependent on the transfection protocol and/or the cell type used to examine the *trans*-activation response.

Since the original submission of the manuscript, Laspija et al. reported similar results with respect to the contribution of antitermination to *trans* activation, using a recombinant adenovirus vector (18).

The first two authors contributed equally to this work.

We thank Beverly Akerman for technical assistance.

This research was supported by grants from The American Foundation for AIDS Research (grant 000946-7-RGR:AmFAR/Drexel Burnham Lambert research grant) and The Cancer Research Society of Montreal to N.S. N.S. is a recipient of an MRC Scientist Award of the Medical Research Council of Canada. S.R. is a recipient of a Fonds de Recherche en Santé du Québec studentship, and N.P. is a recipient of a Medical Research Council of Canada studentship.

LITERATURE CITED

1. Arya, S. K., C. Guo, S. F. Josephs, and F. Wong-Staal. 1985. *trans*-activator gene of human T-lymphotropic virus type III (HTLV-III). *Science* 229:69-73.
2. 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.
3. Christiansen, J., S. R. Douthwaite, A. Chistensen, and R. Garrett. 1985. Does unpaired adenosine-66 from helix II of Escherichia coli 5S RNA bind to protein L18? *EMBO J.* 4:1019-1024.
4. Dayton, A. I., J. G. Sodroski, C. A. Rosen, W. C. Goh, and W. A. Haseltine. 1986. The *trans*-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell* 44:941-947.
5. 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.
6. Enoch, T., K. Zinn, and T. Maniatis. 1986. Activation of the human β -interferon gene requires an interferon-inducible factor. *Mol. Cell. Biol.* 6:801-810.
7. Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal. 1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. *Cell* 46:807-817.
8. Feng, S., and E. C. Holland. 1988. HIV-1 *tat trans*-activation requires the loop sequence within tar. *Nature (London)* 334:165-168.
9. Fisher, A. G., M. B. Feinberg, S. F. Josephs, M. E. Harper,

- L. M. Marselle, G. Reyes, M. A. Gonda, A. Aldovini, C. Debouk, R. C. Gallo, and F. Wong-Staal. 1986. The *trans*-activator gene of HTLV-III is essential for virus replication. *Nature (London)* **320**:367-371.
10. 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.
 11. Goh, W. C., C. Rosen, J. Sodroski, D. D. Ho, and W. A. Haseltine. 1986. Identification of a protein encoded by the *trans* activator gene *tat*III of human T-cell lymphotropic retrovirus type III. *J. Virol.* **59**:181-184.
 12. Grosschedl, R., and D. Baltimore. 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* **41**:885-897.
 13. Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon. 1987. Genome organization and *trans*-activation of the human immunodeficiency virus type 2. *Nature (London)* **326**:662-669.
 14. Haile, D. J., M. W. Hentze, T. A. Rouault, J. B. Harford, and R. D. Klausner. 1989. Regulation of interaction of the iron-responsive element binding protein with iron-responsive RNA elements. *Mol. Cell. Biol.* **9**:5055-5061.
 15. Hauber, J., and B. R. Cullen. 1988. Mutational analysis of the *trans*-activation-responsive region of the human immunodeficiency virus type I long terminal repeat. *J. Virol.* **62**:673-679.
 16. Jakobovits, A., D. H. Smith, E. B. Jakobovits, and D. J. Capon. 1988. A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV *trans* activator. *Mol. Cell. Biol.* **8**:2555-2561.
 17. 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.
 18. Laspia, M. F., A. P. Rice, and M. B. Mathews. 1989. HIV-1 *tat* protein increases transcriptional initiation and stabilizes elongation. *Cell* **59**:283-292.
 19. Lenardo, M. J., C.-M. Fan, T. Maniatis, and D. Baltimore. 1989. The involvement of NF- κ B in β -interferon gene regulation reveals its role as a widely inducible mediator of signal transduction. *Cell* **57**:287-294.
 20. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 21. Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035-7056.
 22. 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.
 23. Nabel, G., and D. Baltimore. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. *Nature (London)* **326**:711-713.
 24. Okamoto, T., and F. Wong-Staal. 1986. Demonstration of virus-specific transcriptional activator(s) in cells infected with HTLV-III by an *in vitro* cell-free system. *Cell* **47**:29-35.
 25. Peattie, D. A., S. Douthwaite, R. A. Garrett, and H. F. Noller. 1981. A "bulged" double helix in a RNA-protein contact site. *Proc. Natl. Acad. Sci. USA* **78**:7331-7335.
 26. Rosen, C. A., J. G. Sodroski, K. Campbell, and W. A. Haseltine. 1986. Construction of recombinant murine retroviruses that express the human T-cell leukemia virus type II and human T-cell lymphotropic virus type III *trans* activator genes. *J. Virol.* **57**:379-384.
 27. Rosen, C. A., J. G. Sodroski, and W. A. Haseltine. 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. *Cell* **41**:813-823.
 28. Sadaie, M. R., T. Betner, and F. Wong-Staal. 1988. Site-directed mutagenesis of two *trans*-regulatory genes (*tat-III*, *trs*) of HIV-1. *Science* **239**:910-913.
 29. Salsler, W. 1977. Globin mRNA sequences: analysis of base-pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **42**:985-1002.
 30. Selby, M. J., E. S. Bain, P. Luciw, and B. M. Peterlin. 1989. Structure, sequence, and position of the stem-loop in *tat* determine transcriptional elongation by *tat* through the HIV-1 long terminal repeat. *Genes Dev.* **3**:547-558.
 31. 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.
 32. Sodroski, J., W. C. Goh, C. A. Rosen, A. Dayton, E. Terwilliger, and W. A. Haseltine. 1986. A second post-transcriptional *trans*-activator gene required for HTLV-III replication. *Nature (London)* **321**:412-417.
 33. Sodroski, J., R. Patarca, C. Rosen, F. Wong-Staal, and W. Haseltine. 1985. Location of the *trans*-activating region on the genome of human T-cell lymphotropic virus type III. *Science* **229**:74-77.
 34. Sodroski, J., C. Rosen, F. Wong-Staal, S. Z. Salahuddin, M. Popovic, S. Arya, R. C. Gallo, and W. A. Haseltine. 1985. *trans*-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. *Science* **227**:171-173.
 35. Terwilliger, E., R. Burghoff, R. Sia, J. Sodroski, W. Haseltine, and C. Rosen. 1988. The *art* gene product of human immunodeficiency virus is required for replication. *J. Virol.* **62**:655-658.
 36. 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.
 37. Visvanathan, K. V., and S. Goodburn. 1989. Double-stranded RNA activates binding of NF- κ B to an inducible element in the human β -interferon promoter. *EMBO J.* **8**:1129-1138.
 38. Wain-Hobson, S., S. P. Sonigo, O. Danos, S. Cole, and M. Alizon. 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**:9-17.
 39. Wu, H.-N., and O. C. Uhlenbeck. 1987. Role of a bulged A residue in a specific RNA protein interaction. *Biochemistry* **26**:8221-8227.
 40. Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β -interferon gene. *Cell* **34**:865-879.
 41. Zoller, M. J., and M. Smith. 1982. Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Res.* **10**:6487-6500.