Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4858–4862, July 1989
Biochemistry

Specific binding of a HeLa cell nuclear protein to RNA sequences in the human immunodeficiency virus transactivating region

(chemical nuclease imprinting/untranslated RNA binding protein 1)

RICHARD GAYNOR*†‡, EMMANUEL SOULTANAKIS*†, MICHIO KUWABARA*§, JOSEPH GARCIA*†, AND DAVID S. SIGMAN*§

*Molecular Biology Institute, and †Division of Hematology-Oncology, Department of Medicine, and §Department of Biological Chemistry, School of Medicine, and ‡Wadsworth Veterans Hospital, University of California, Los Angeles, CA 90024

Communicated by Paul D. Boyer, March 27, 1989

ABSTRACT **The transactivator protein, tat, encoded by the human immunodeficiency virus is a key regulator of viral transcription. Activation by the tat protein requires sequences downstream of the transcription initiation site called the transactivating region (TAR). RNA derived from the TAR is capable of forming a stable stem–loop structure and the maintenance of both the stem structure and the loop sequences located between +19 and +44 is required for complete *in vivo* activation by tat. Gel retardation assays with RNA from both wild-type and mutant TAR constructs generated *in vitro* with SP6 polymerase indicated specific binding of HeLa nuclear proteins to the TAR. To characterize this RNA–protein interaction, a method of chemical "imprinting" has been developed using photoactivated uranyl acetate as the nucleolytic agent. This reagent nicks RNA under physiological conditions at all four nucleotides in a reaction that is independent of sequence and secondary structure. Specific interaction of cellular proteins with TAR RNA could be detected by enhanced cleavages or imprints surrounding the loop region. Mutations that either disrupted stem base-pairing or extensively changed the primary sequence resulted in alterations in the cleavage pattern of the TAR RNA. Structural features of the TAR RNA stem–loop essential for tat activation are also required for specific binding of the HeLa cell nuclear protein.**

Gene expression of the human immunodeficiency virus (HIV) is dependent on multiple cis-acting sequences in the long terminal repeat (LTR) which serve as the binding sites for cellular proteins that are important in viral gene regulation (1–13). In addition, at least three viral proteins—tat, rev, and nef—are also involved in modulating gene expression (14–26). One of these HIV-encoded proteins, tat, requires for its action a region of the HIV LTR extending from −17 to +80, known as the transactivating region (TAR) (1, 5–10).

TAR sequences may be involved in tat regulation of HIV transcription at both the DNA and RNA levels. For example, it is the recognition site for cellular DNA binding proteins UBP-1 (12) and UBP-2 (10) (UBP, untranslated binding protein). Mutagenesis of multiple UBP-1 binding sites in the TAR resulted in decreased activation by the tat protein in transient gene expression assays with the HIV LTR (9, 10). On the other hand, RNA from the TAR is capable of forming a stable stem–loop structure (4). Mutagenesis of the TAR indicated that maintenance of stem base-pairing between +19 and +44 was critical for tat activation (6, 10). However, the primary sequence of the loop region between +28 and +34 was also critical for this activation (6, 10).

The mechanism(s) by which the tat protein activates HIV gene expression is not known. Although the tat protein has been shown to be involved in increasing the steady-state level of RNA, it has also been reported to increase the translation of RNA (15, 17, 22). The possibility that increased levels of RNA are attributable to an antitermination mechanism has also been raised (5, 18–20, 27). In contrast to other transactivating proteins, such as E1A (28), which activate a number of viral and cellular promoters, tat activation is specific for the HIV LTR. It is of interest to identify features of the TAR responsible for this selectivity.

Since TAR RNA secondary structure appeared to be one component required for complete activation by the tat protein, the possibility that a cellular protein may mediate the interaction of tat with TAR RNA has been explored. The methods used were gel-retardation analysis (29) and a new method of examining RNA–protein interactions in which light-activated uranyl acetate is used as the nucleolytic agent. This reagent, which was previously shown to cut DNA in a sequence-independent reaction (30), cleaves RNA in a primary and secondary structure-independent manner under physiological conditions. Both methods demonstrate specific binding of a HeLa nuclear protein which we designate as untranslated RNA binding protein 1 (URBP-1) to TAR RNA.

MATERIALS AND METHODS

Plasmid Constructs and Labeling of mRNAs. Wild-type and mutant HIV mRNAs shown in Fig. 1 were constructed by taking a *Pvu* II/*Xho* I fragment (−18 to +80) from pJGFCAT derivatives (10) and inserting them into a pGEM derivative in which the *Pst* I site was converted into an *Xho* I site by linker insertion. Each pGEM construct was digested with *Hin*dIII, and RNA synthesis, labeling, and purification were performed using the reagents and procedures of the Riboprobe System II (Promega) (29).

Gel Retardation Assays and Chemical Footprinting of RNA. Gel retardation assays were performed as described by Leibold and Munro (29). Internally labeled mRNA (50,000 cpm) was incubated with 1–5 μg of a heparin-agarose column-purified HeLa nuclear extract (31) in 10 μl for 20 min (5, 30). When included, 1 pmol of the indicated unlabeled competitor mRNA was added just prior to the addition of labeled mRNA.

For chemical cleavage, RNA was 5′-end-labeled (100,000 cpm) and incubated as in the gel retardations except 50 mM NaCl was used instead of 100 mM NaCl and reaction volumes were 100 μl. Then, 10 μl of 20 mM uranyl acetate was added followed by exposure to short-wave UV light (7000 μW/cm^2) for 30 min. RNA was extracted, ethanol precipitated, and loaded onto a 10% polyacrylamide sequencing gel. Footprinting lanes were calibrated with partial ribonuclease T1 digests and autoradiography was performed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; LTR, long terminal repeat; TAR, transactivating region; UBP, untranslated binding protein; URBP-1, untranslated RNA binding protein 1.

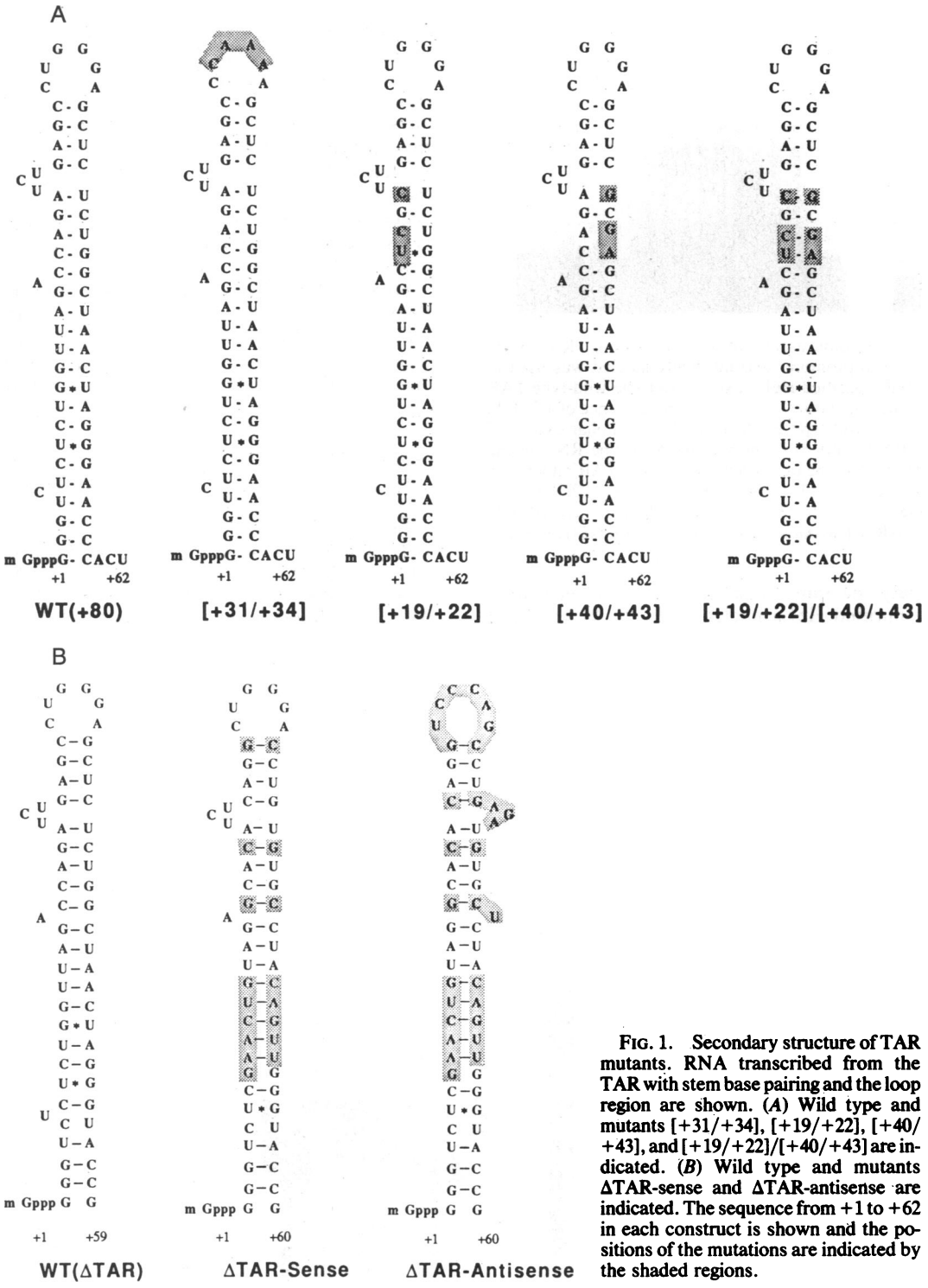

FIG. 1. Secondary structure of TAR mutants. RNA transcribed from the TAR with stem base pairing and the loop region are shown. (*A*) Wild type and mutants [+31/+34], [+19/+22], [+40/+43], and [+19/+22]/[+40/+43] are indicated. (*B*) Wild type and mutants ΔTAR-sense and ΔTAR-antisense are indicated. The sequence from +1 to +62 in each construct is shown and the positions of the mutations are indicated by the shaded regions.

RESULTS

Gel-Retardation Analysis of Wild-Type and Mutant TAR RNAs. Gel retardation assays with wild-type and mutant TAR RNAs (see Fig. 1) were carried out to determine whether cellular proteins present in HeLa cell nuclear extracts could bind to this region. The method of Leibold and Munro (29), which utilizes internally labeled RNA and involves digestion of the protein–RNA extract with T1 ribonuclease prior to polyacrylamide gel electrophoresis, provided a reproducible gel-retardation system.

A gel-retarded species was seen with wild-type TAR-generated RNA in the presence of partially purified HeLa extract (Fig. 2*A*). This species specifically competed with a 50-fold excess of unlabeled wild-type RNA, but not with a corresponding amount of a nonspecific lac mRNA competitor. A similar gel-retarded species was seen with a wild-type TAR RNA that was initiated at the HIV cap site (data not shown). The TAR binding factor, known as untranslated RNA binding protein 1 (URBP-1), is sensitive to heat treatment to 60°C for 5 min and is insensitive to micrococcal nuclease treatment, suggesting that this factor(s) is a protein (data not shown).

Labeled RNA from the TAR mutants shown in Fig. 1 was also used in gel-retardation assays. These RNAs were labeled

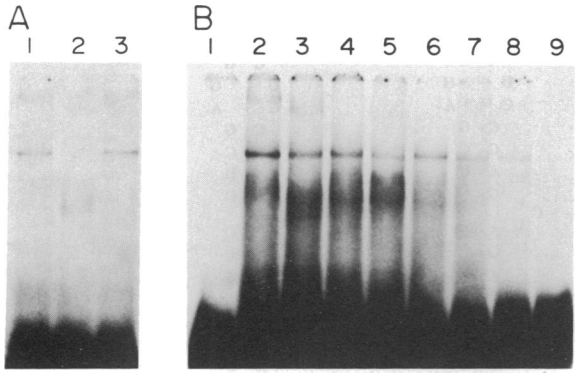

FIG. 2. Gel retardation assays of TAR mutations. RNA transcribed from the SP6 promoter and internally labeled was used in assays with partially purified HeLa extract. (*A*) The wild-type TAR RNA in the presence of extract (lane 1) or in the presence of a 50-fold excess of unlabeled wild-type RNA (lane 2) or a similar excess of unlabeled lac mRNA (lane 3) is shown. (*B*) Wild-type RNA in the absence (lane 1) or presence of extract (lane 2) and TAR mutants in the presence of extract [+31/+34] (lane 3), [+19/+22] (lane 4), [+40/+43] (lane 5), [(+19/+22)(+40/+43)] (lane 6), ΔTAR-wild type (lane 7), ΔTAR-sense (lane 8), and ΔTAR-antisense (lane 9) are shown.

to approximately the same specific activity and incubated with partially purified HeLa nuclear extract. As shown in Fig. 2*B*, all of the constructs except ΔTAR-antisense gave the same gel-retarded species. The ΔTAR-antisense construct preserves the stem–loop structure and stem energy but extensively changes the primary structure of both the stem and the loop. Even though all the mutant constructs shown in Fig. 2*B* except for ΔTAR-antisense gave rise to gel-retarded species, competition experiments using unlabeled RNAs for the +31/+34, +19/+22, +40/+43, and ΔTAR-antisense constructs with labeled wild-type RNA indicated that these mutant RNAs competed poorly for cellular protein binding factor(s) (data not shown). These results suggest that the preservation of a stem–loop structure of fixed dimensions is not sufficient for binding; primary sequence of TAR RNAs is also a critical determinant for stable binding of URBP-1 as determined by gel-retardation assays.

Studying RNA–Protein Interactions with the Nuclease Activity of Uranyl Acetate. In addition to gel-retardation assays, sequence-specific protein–RNA interactions may be detected by nuclease protection assays. However, most nucleases available for studying protein–RNA interactions are too specific in their scission of the nucleic acid. For example, T1 ribonuclease reacts only at G residues in single-stranded regions; ribonuclease A reacts only at pyrimidine residues in single-stranded regions; α-sarcin reacts in both double-stranded and single-stranded regions at purine residues.

Recently, the light-dependent sequence-independent scission of DNA by uranyl acetate has been described (30). Although this reagent is functionally similar to ferrous EDTA for footprinting protein–DNA interactions (32), its reactivity with DNA differs in one important mechanistic aspect from that of ferrous EDTA. It is not inhibited by dithiothreitol and glycerol, whereas ferrous EDTA, a hydroxyl radical generator, is strongly inhibited by these reagents. This difference

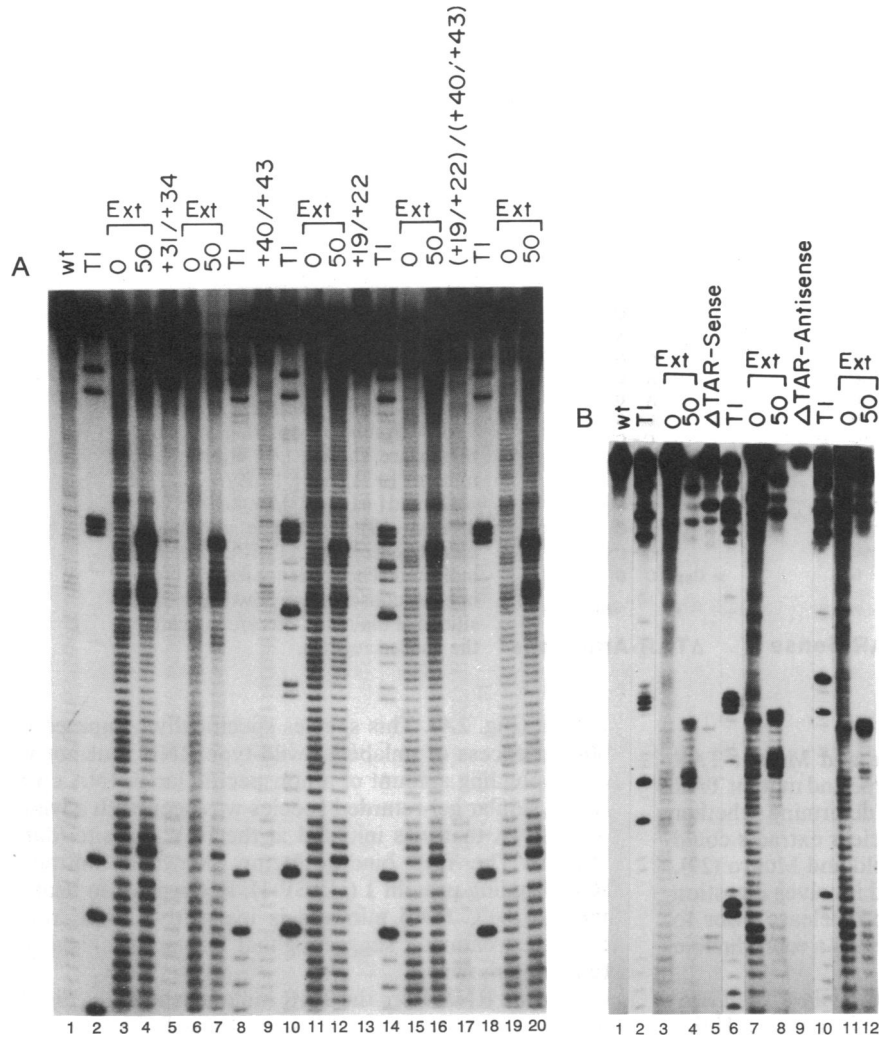

FIG. 3. Uranyl acetate cleavage of TAR RNA. For each series of constructs, 5′-end-labeled RNA in the absence of treatment (lane 1), cleaved with T1 ribonuclease (lane 2), cleaved with uranyl acetate (lane 3), and cleaved with uranyl acetate in the presence of 50 μg of partially purified HeLa extract are shown (lane 4). (*A*) Wild-type (lanes 1–4), +31/+34 (lanes 5–8), +40/+43 (lanes 9–12), +19/+22 (lanes 13–16, and [(+19/+22)/(+40/+43)] (lanes 17–20) construct RNAs are shown. (*B*) ΔTAR-wild type (lanes 1–4), ΔTAR-sense (lanes 5–8), and ΔTAR-antisense (lanes (9-12) construct RNAs are shown in the same order as in *A*.

in reactivity prompted us to examine the reactivity of uranyl acetate with RNA because ferrous EDTA has not proved to be efficient in degrading RNA. Using RNA from the wild-type construct and each of the TAR mutants that was generated from SP6 vectors and labeled at the 5′ end with polynucleotide kinase and [γ-^{32}P]ATP, we found that strand scission with uranyl acetate is achieved by irradiating a solution with UV light for 30 min (Fig. 3). An important feature of cleavage by uranyl acetate is that it generates a regular ladder exhibiting no preference for nucleotide, single-stranded, or double-stranded regions. For example, the scission patterns of the different RNAs are similar in the sequence positions where mutations have altered both the nucleotide composition and secondary structure (Fig. 3).

These RNAs were incubated with uranyl acetate in both the presence and absence of cellular extract and the reactions were analyzed by gel electrophoresis and autoradiography (Figs. 3 and 4). T1 ribonuclease digestion of each RNA is shown next to the uranyl acetate reaction to indicate the position of the loop sequences that are digested with this ribonuclease due to the single-stranded nature of this region. The digestion pattern of each of the RNAs was subjected to densitometry and the intensity of cleavage was compared in both the presence and absence of cellular extract (Fig. 4). The wild-type RNA gives enhanced cleavages in two regions between +22/+25 and +28/+33 in lanes containing extract as compared to lanes lacking extract (Figs. 3*A* and 4*A*). This enhancement of cleavages varies between 2- and 10-fold (Fig. 4*A*). In other regions of wild-type TAR RNA, there were minimal differences between the extract-containing and no extract-containing lanes (Figs. 3*A* and 4*A*). A mutant that changes the loop sequences between +31/+34 also reveals enhanced cleavages in the presence of extract between +17/+33, but the pattern is different from that seen with the wild-type RNA, in that enhanced cleavages are seen throughout this region (Figs. 3*A* and 4*A*). Mutations that disrupt the stem [(+19/+22) and (+40/+43)] result in enhanced cleavages with extract between +22 and +33 but, again, with a different pattern than that seen with the wild-type RNA (Figs. 3*A* and 4*A*). However, combining these two mutations [(+19/+22)(+40/+43)], which reforms stem base pairing, results in a pattern of enhanced cleavages more similar to that seen with the wild-type RNA (Figs. 3*A* and 4*A*). This would suggest that the maintenance of an intact stem–loop structure is required to obtain a wild-type cleavage pattern.

In addition, two mutants that extensively change TAR RNA sequences but preserve stem–loop structure and stem energy were also tested. The mutant ΔTAR-sense extensively changes primary sequence and eliminates UBP-1 binding sites in DNA but maintains stem base-pairing, the loop sequence, and stem energy of the RNA (Fig. 1). In this case, the presence of the extract induces two clusters of increased cleavages between +23/+25 and +29/+32 similar to that seen with the wild-type RNA (Figs. 3*B* and 4*B*). However, with the mutant ΔTAR-antisense, in which the loop sequence is the complementary sequence to the wild-type loop, there is a marked decrease in the intensity of cleavages between +23/+32 in the presence of extract (Figs. 3*B* and 4*B*). Thus,

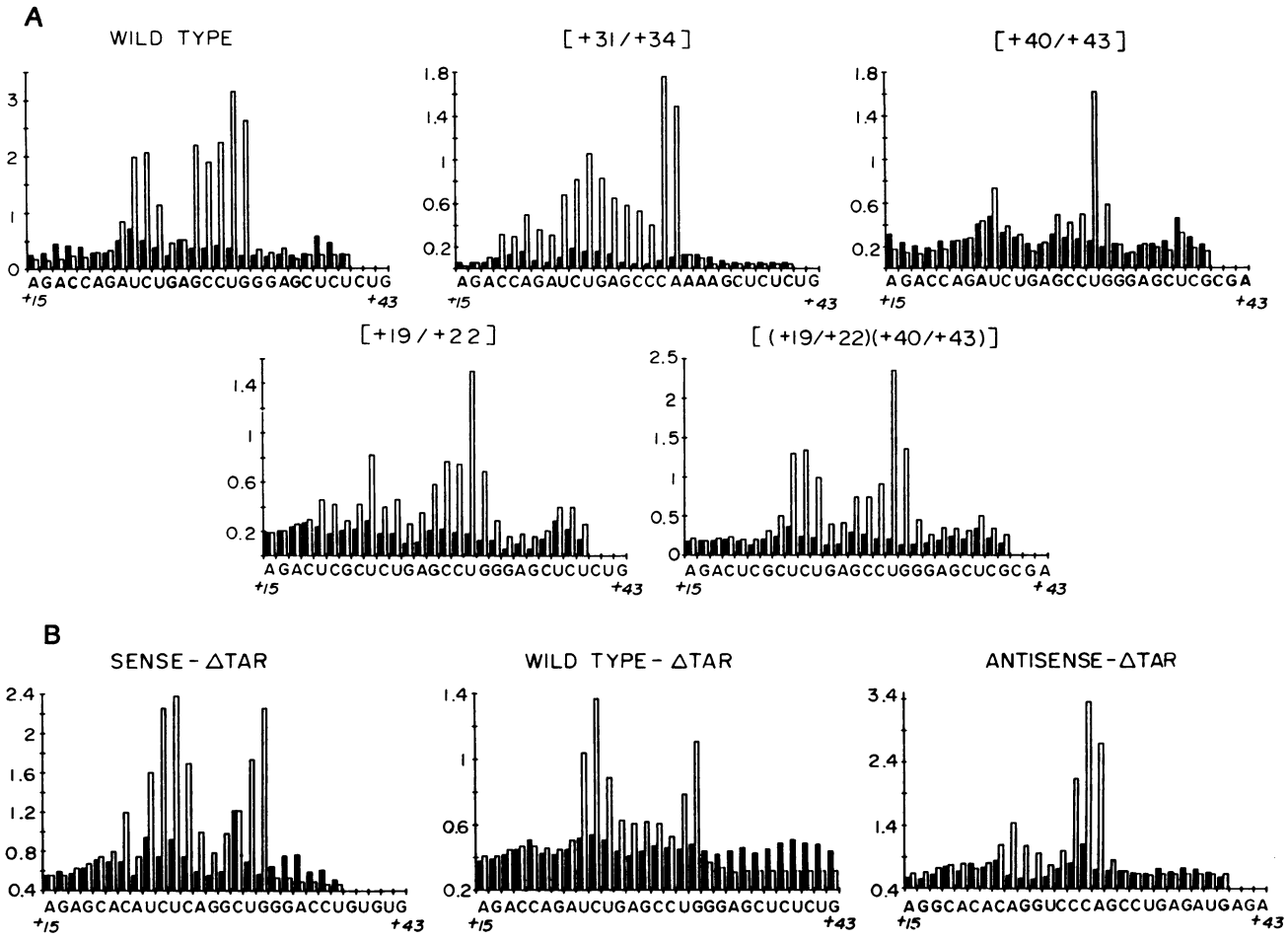

FIG. 4. Densitometry of uranyl acetate cleavage. Uranyl acetate cleavage of RNA in both the presence and the absence of cellular extract was analyzed by densitometry. Solid bars indicate intensity of cleavage in the absence of extract; open bars indicate intensity of cleavage in the presence of abstract. TARs extending from +15 to +43 and intensities of cleavage are indicated for wild type, [+31/+34], [+40/+43], [+19/+22], and [(+19/+22)(+40/+43)] (*A*) and for sense-ΔTAR, wild type-ΔTAR, and antisense-ΔTAR (*B*).

mutants that alter loop sequences (+31/+34 and ΔTAR-antisense) indicate that the primary sequence of the loop is clearly an important determinant required for the wild-type cleavage pattern.

DISCUSSION

Gel retardation and nuclease digestion of RNA–protein complexes demonstrate that a nuclear protein isolated from HeLa cells (URBP-1) binds to the TAR of HIV. The gel-retardation method, which was successfully used in this study, was initially used in a study of structure–function relationships in ferritin mRNA (29). It requires the use of internally labeled RNA, because radioactivity associated with either 5′ or 3′ labeling is readily lost due to T1 ribonuclease digestion. Even though the inability to use end-labeled RNAs prohibits the use of in-gel footprinting techniques (33), the binding of a series of cognate RNAs can be studied by competition. Another approach for studying RNA–protein interactions utilizes a new reagent for RNA scission, photoactivated uranyl acetate. In the presence of UV light, uranyl acetate cleaves the phosphodiester backbone, presumably by oxidative attack on the ribose moiety, at all four nucleotides A, U, G, and C. This reactivity is independent of primary and secondary structure. Its lack of nucleotide specificity contrasts with that observed with other reagents used to cut RNA.

The two methods of studying RNA–protein interactions indicate that both the loop region and the stem structure contribute to the stability of the binding. Since the binding of the nuclear protein induces a marked enhancement of cleavage in two regions of the wild-type TAR sequences between +22 and +33, an "imprint" rather than a footprint provides evidence for a specific RNA–protein interaction. The sequence positions 23–26 (AUCU) and 29–33 (GCCUG) are more rapidly cleaved in the presence of the protein than in its absence. Both of these sequences comprise all or part of postulated single-stranded regions in the stem–loop structure (Fig. 1). UCU (positions 24–26) constitutes the bulge region, whereas CUG (positions 31–33) is the sequence comprising the first 3 bases of the loop structure. If uranyl acetate binds to the phosphate in an analogous fashion to magnesium ion, as has been proposed in the earlier studies with DNA (30), the enhanced scission could be due to protein-induced changes in the conformation of the phosphodiester backbone that facilitate cation binding. Although the detailed chemistry of the scission reaction is not yet known, the increased rate of scission could also be partially due to alterations in the conformation of the ribose moiety accompanying protein binding.

A wild-type cleavage pattern was found in a TAR mutant [(+19/+22)(+40/+43)], which was also strongly inducible to tat activation in transient gene expression assays (10). Other constructs [(+31/+34), (+19/+22), (+40/+43), ΔTAR-antisense] that were severely defective for tat induction did not give the same imprint pattern of enhanced scission observed with the wild-type RNA (10). The only exception to this correlation between *in vivo* tat activation and a wild-type imprint pattern was a mutant (ΔTAR-sense), which diminished transcription owing to a disruption of multiple UBP-1 binding sites in the TAR DNA (10). The tat defective construct (+31/+34), in which U-31 is changed to C and G-32, G-33, and G-34 are changed to A, is also an instructive mutant for emphasizing the specificity of URBP-1 binding. It retains all features of the stem–loop structure, yet it fails to induce changes in the RNA structure that parallel those observed with the wild type. Secondary structure alone is therefore not sufficient for wild-type patterns of imprinting; base-specific interactions must also play a central role in precise binding.

The correspondence between tat activation and URBP-1 binding is unlikely to be fortuitous but probably arises from a functional relationship between the two processes. Possibly, tat may bind to the URBP-1–TAR RNA complex and this structure could facilitate interaction with RNA polymerase or other cellular factors involved in RNA elongation. Alternatively, URBP-1 and tat may compete for binding to TAR RNA with different ensuing effects on gene expression. Although the URBP-1 activity described here was purified from nuclear extract, cytoplasmic forms of URBP-1 may also be important in gene expression. URBP-1 binding to the TAR RNA could block the activation of the double-stranded RNA-dependent protein kinase (34, 35) and thus prevent inhibition of host cell translation. Characterization and purification of URBP-1 will be vital for testing these different models.

This research was supported by grants from the U.S. Public Health Service (GM39558, GM38815, AI25288, CA30891) and the California Universitywide Task Force on AIDS.

1. Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. (1985) *Cell* **41**, 813–823.
2. Jones, K. A., Kadonga, J. T., Luciw, P. A. & Tjian, R. (1986) *Science* **232**, 755–759.
3. Nabel, G. & Baltimore, D. (1987) *Nature (London)* **326**, 711–713.
4. Muesing, M., Smith, D. H. & Capon, D. J. (1987) *Cell* **48**, 691–701.
5. Garcia, J. A., Wu, F. K., Mitsuyasu, R. & Gaynor, R. B. (1987) *EMBO J.* **6**, 3761–3770.
6. Feng, S. & Holland, E. C. (1988) *Nature (London)* **334**, 165–167.
7. Jakobovits, A., Smith, D. H., Jakobovits, E. B. & Capon, D. J. (1988) *Mol. Cell. Biol.* **8**, 2555–2561.
8. Hauber, J. & Cullen, B. R. (1988) *J. Virol.* **62**, 637–679.
9. Jones, K. A., Luciw, P. A. & Duchange, N. (1988) *Genes Dev.* **2**, 1101–1114.
10. Garcia, J., Harrich, D., Soultanakis, E., Wu, F. K., Mitsuyasu, R. & Gaynor, R. B. (1989) *EMBO J.* **8**, 765–778.
11. Wu, F. K., Garcia, J., Mitsuyasu, R. & Gaynor, R. B. (1988) *J. Virol.* **62**, 218–225.
12. Wu, F. K., Garcia, J. A., Harrich, D. & Gaynor, R. B. (1988) *EMBO J.* **7**, 2117–2129.
13. Kawakami, K., Scheidereit, C. & Roeder, R. G. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4700–4704.
14. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. (1985) *Science* **229**, 69–73.
15. Cullen, B. (1986) *Cell* **46**, 973–982.
16. Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. (1988) *Cell* **44**, 941–947.
17. Feinberg, M. B., Jarret, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. (1986) *Cell* **46**, 807–817.
18. Hauber, J., Perkins, A., Heimer, E. P. & Cullen, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6364–6368.
19. Kao, S. Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. (1987) *Nature (London)* **330**, 489–493.
20. Rice, A. P. & Mathews, M. B. (1988) *Nature (London)* **332**, 551–553.
21. Rosen, C. A., Sodroski, J. G., Goh, W. C., Dayton, A. I., Lippke, J. & Haseltine, W. A. (1986) *Nature (London)* **319**, 555–559.
22. Sodroski, J. G., Goh, W. C., Rosen, C., Dayton, A., Terwilliger, E. & Haseltine, W. (1986) *Nature (London)* **321**, 412–417.
23. Garcia, J., Harrich, D., Mitsuyasu, R. & Gaynor, R. B. (1989) *EMBO J.* **7**, 3143–3147.
24. Fisher, A. G., Feinberg, M. B., Josephs, S. F., Harper, M. E. & Wong-Staal, F. (1986) *Nature (London)* **320**, 367–371.
25. Ahmad, W. & Venketesan, J. (1988) *Science* **241**, 1481–1485.
26. Okamoto, T. & Wong-Staal, F. (1986) *Cell* **47**, 29–35.
27. Peterlin, B. M., Luciw, P. A., Barr, P. J. & Walker, M. D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9734–9738.
28. Berk, A. J. (1986) *Annu. Rev. Genet.* **20**, 45–80.
29. Leibold, E. A. & Munro, H. N. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2171–2175.
30. Nielson, P. E., Jeppesen, C. & Burchardt, O. (1988) *FEBS Lett.* **235**, 122–124.
31. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
32. Tullius, T. D. & Dombroski, B. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5469–5473.
33. Kuwabara, M. D. & Sigman, D. S. (1987) *Biochemistry* **26**, 7234–7238.
34. Edery, I., Petryshyn, R. & Sonnenberg, N. (1989) *Cell* **56**, 303–312.
35. Sen Gupta, D. & Silverman, R. H. (1989) *Nucleic Acids Res.* **17**, 969–978.