Probing the Proximity of the Core Domain of an HIV-1 Tat Fragment in a Tat−TAR Complex by Affinity Cleaving[†]

Ikramul Huq and Tariq M. Rana*

Department of Pharmacology, Robert Wood Johnson (Rutgers) Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854

Received April 30, 1997; Revised Manuscript Received July 2, 1997[⊗]

ABSTRACT: Transactivation of human immunodeficiency virus (HIV) gene expression depends upon the interaction of the viral regulatory protein Tat with the transactivation responsive region (TAR) RNA, a 59-base stem−loop structure located at the 5′-end of all mRNAs. We have used a site-directed RNA-cleaving strategy to determine the neighborhood of the core domain of a Tat fragment in the Tat−TAR complex. We synthesized a 35-amino acid fragment containing arginine-rich RNA-binding domain of Tat(38−72) and attached an EDTA analog to its amino terminus. A derivative of (*p*-aminobenzyl)-EDTA tetra-*tert*-butyl ester was synthesized and attached to the amino terminus of the Tat peptide by standard peptide coupling methods. Cleavage from the resin and deprotection of the peptide were carried out in trifluoroacetic acid which also generated unprotected metal binding EDTA moieties. We used this EDTA−Tat conjugate to form a specific complex with TAR RNA. This sequence-specific RNA-binding peptide was converted into a sequence-specific RNA-cleaving peptide by the addition of Fe(II) salt, ascorbate, and H_2O_2. Hydroxyl radicals generated from the tethered Fe(II) cleaved the TAR RNA backbone in two localized regions. Site-specific cleavage of TAR RNA was observed at the bulge residues (U23, C24, and U25), in the loop region (G34 and A35), and at the strand opposite the bulge (U40 and C41). These results demonstrate that, in the three-dimensional structure of the Tat−TAR complex, the Phe38 of Tat(38−72) is located in the proximity of the bulge region and two nucleotides from the loop sequence.

The promoter of the human immunodeficiency virus type 1 (HIV-1),[1] located in the U3 region of the viral long terminal repeat (LTR), is an inducible promoter which can be stimulated by the transactivator protein, Tat (*1*). As in other lentiviruses, Tat protein is essential for transactivation of viral gene expression (*2*−*6*). In the absence of Tat, most of the viral transcripts terminate prematurely, producing short RNA molecules ranging in size from 60 to 80 nucleotides. The Tat protein is a small, cysteine-rich nuclear protein containing 86 amino acids and an arginine-rich RNA-binding domain (amino acids 49−57) located in the carboxyl-terminal half of the molecule (*1, 6*). HIV-1 Tat protein acts by binding to the TAR (transactivation responsive) RNA element, a 59-base stem−loop structure located at the 5′-end of all nascent HIV-1 transcripts (*7*). Upon binding to the TAR RNA sequence, Tat causes a substantial increase in transcript levels (*8*−*12*). The increased efficiency in transcription may result from preventing premature termination of the transcriptional elongation complex (*13*−*17*). TAR RNA was originally localized to nucleotides +1 to +80 within the viral long terminal repeat (LTR) (*18*). Subsequent deletion studies have established that the region from +19 to +42 incorporates the minimal domain that is both necessary and sufficient for Tat responsiveness *in vivo* (*19*−*21*). TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge which separates two helical stem regions (*7, 10, 18, 19*). The trinucleotide bulge is essential for high-affinity and specific binding of the Tat protein (*22, 23*). The loop region is required for *in vivo* transactivation but is not involved in Tat binding (*23*−*27*).

RNA molecules can fold into extensive structures containing regions of double-stranded duplex, hairpins, internal loops, bulged bases, and pseudoknotted structures (*28, 29*). Due to the complexity of RNA structure, the rules governing sequence-specific RNA−protein recognition are not well understood. RNA−protein interactions are vital for many regulatory processes, especially, in gene regulation where proteins specifically interact with binding sites found within RNA transcripts. Understanding the principles of Tat−TAR interactions is a crucial step for drug design. Although high-resolution NMR information is limited to the TAR RNA component, a structure of the RNA−protein complex is still missing. Therefore, new methods are needed to determine the topology of RNA−protein complexes under physiological conditions. We have used an affinity cleaving method to map the proximity of the core domain of a Tat fragment in the Tat-TAR complex.

Interaction of transition metal complexes with nucleic acids and proteins makes them a useful tool in molecular biology. Metal complexes can be synthesized for specific recognition and cleavage of nucleic acids or they can be used as nonspecific nucleases (for an excellent review, see ref *30*). An iron−EDTA complex cleaves DNA nonspecifically and can be applied to determine the helical periodicity of DNA and structural details of bent DNA (*31, 32*). In the presence of ascorbate and H_2O_2, the Fe(II)−EDTA complex generates

[†]This work was supported in part by National Institutes of Health Grants AI 34785, and TW00702 and a Research Career Development Awrad to T.M.R.

* To whom correspondence should be addressed. Phone: (732) 235-4082. Fax: (732) 235-4073. E-mail: rana@umdnj.edu.

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1997.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; TAR, transactivation response element; Tat, transcription antitermination; EDTA, ethylenediaminetetraacetic acid.

S0006-2960(97)01011-8 CCC: $14.00

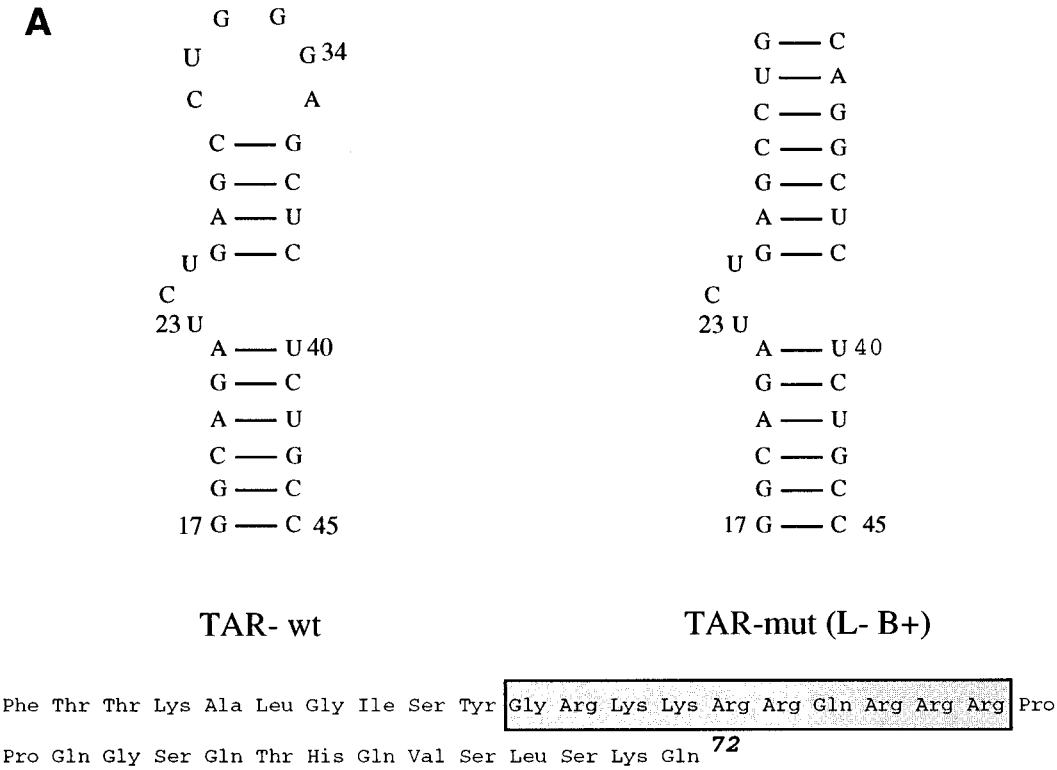

FIGURE 1: (A) Secondary structures of wild-type and mutant TAR RNAs used in this study. Wild-type TAR RNA spans the minimal sequences that are required for Tat responsiveness *in vivo* (*19*) and for *in vitro* binding of Tat-derived peptides (*26*). Wild-type TAR contains two non-wild-type base pairs to increase transcription by T7 RNA polymerase. Mutant TAR RNA contains a wild-type trinucleotide bulge, and the loop sequence is substituted with Watson−Crick base pairs. The numbering of nucleotides in the RNA corresponds to their positions in wild-type TAR RNA. (B) The Tat peptide, amino acids 38−72, contains the RNA-binding domain of Tat. The arginine-rich region of Tat, corresponding to the Tat(48−57) peptide, is highlighted.

hydroxyl radicals which are responsible for DNA scission (*33*). Cleavage of DNA and RNA by metal chelates is an important new approach for characterizing specific structural features of nucleic acids and their complexes in solutions (*34*−*36*). The metal chelate attachment converts sequence-specific DNA-binding protein or oligonucleotide to sequence-specific DNA-cleaving molecules that function under physiological pH, temperature, and salt conditions (*37*−*41*). In a recent report, DNA cleavage mediated by (*p*-bromoacetamidobenzyl)-EDTA−Fe has been used to demonstrate the arrangement and contacts of α-subunits of *Escherichia coli* RNA polymerase on the upstream element DNA (*42*). Recently, RNA−protein proximities have been probed by hydroxyl radicals generated by Fe(II)−EDTA attached to a protein in a ribonucleoprotein complex (*43*−*45*).

In this report, we have used the affinity cleaving method to study Tat−TAR interactions in solution. It has been shown by a number of groups that Tat-derived peptides which contain the basic arginine-rich region of Tat are able to form *in vitro* complexes with TAR RNA (*26, 27, 46*−*51*). To achieve specific RNA binding by a Tat fragment, we synthesized a Tat peptide (amino acids 38−72) which contained an RNA-binding domain and part of the core domain of the Tat protein (Figure 1). This sequence-specific RNA-binding peptide was converted into a sequence-specific RNA-cleaving peptide by covalent attachment of an EDTA analog. Fe(II)−EDTA-modified Tat peptide cleaved TAR RNA at specific positions upon addition of Fe(II) salt, ascorbate, and H_2O_2. Sites of RNA cleavage were mapped by sequencing reactions, revealing the proximity of the NH_2 terminus of the peptide when bound to RNA.

EXPERIMENTAL PROCEDURES

Buffers. All buffer pH values refer to measurements at room temperature: TK buffer, 50 mM Tris-HCl (pH 7.4), 20 mM KCl, 0.1% Triton X-100; Transcription buffer, 40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 0.01% Triton X-100, and 5 mM DTT; TBE buffer, 45 mM Tris-borate (pH 8.0) and 1 mM EDTA; sample loading buffer, 9 M urea, 1 mM EDTA, and 0.1% bromophenol blue in 1× TBE buffer; binding buffer, 25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM $MgCl_2$, and 0.1% Triton X-100; hydrolysis buffer, 50 mM Na_2CO_3/$NaHCO_3$ (pH 9.2); elution buffer, 1× TBE and 10% sodium acetate (3 M, pH 5.5).

Oligonucleotide Synthesis. (a) *DNAs.* All DNAs were synthesized on an Applied Biosystems ABI 392 DNA/RNA synthesizer. The template strand encodes the sequence for the duplex TAR RNA (Figure 1). The top strand is a short piece of DNA complementary to the 3′-end of all template DNAs having the sequence 5′TAATACGACTCACTAT-AG3′. DNA was deprotected in NH_4OH at 55 °C for 8 h and then dried in a Savant speedvac. The samples were resuspended in sample loading buffer and were purified on 20% acrylamide/8 M urea denaturing gels (50 cm × 0.8 mm). Gels were run for 3 h at 30 W until the xylene cyanol tracking dye was 5 cm from the bottom of the gel. DNAs were visualized by UV shadowing, excised from the gel, and eluted in 50 mM Tris, 16 mM boric acid, 1 mM EDTA, and 0.5 M sodium acetate. DNAs were ethanol precipitated and resuspended in DEPC (diethyl pyrocarbonate), treated water. Concentrations of DNAs were determined by measuring absorbance at 260 nm in a Shimadzu UV spectrophotometer.

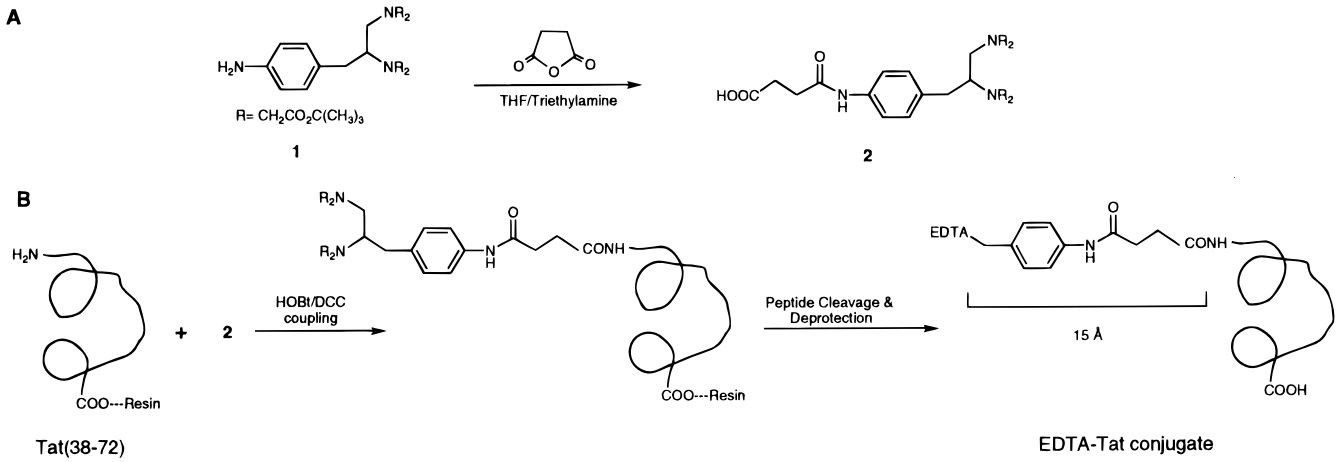

FIGURE 2: (A) Synthesis of a bifunctional chelating agent, compound **2**. (B) Schematic representation for site-specific chemical modification of a Tat fragment with an EDTA analog, compound **2**. The distance between the EDTA molecule and the amino terminus of the Tat peptide is ≈15 Å.

(b) RNAs. All RNAs were prepared by *in vitro* transcription (*52*, *53*). The template strand of DNA was annealed to an equimolar amount of top strand DNA, and transcriptions were carried out in transcription buffer and 4.0 mM NTPs at 37 °C for 2−4 h. For reaction mixtures (20 μL) containing 8.0 pmol of template DNA, 40−60 units of T7 polymerase (Promega) was used. Transcription reactions were stopped by adding an equal volume of sample loading buffer. RNA was purified on 20% acrylamide/8 M urea denaturing gels as described above. RNAs were stored in DEPC water at −20 °C.

RNAs were 5′-dephosphorylated by incubation with calf intestinal alkaline phophatase (Promega) for 1 h at 37 °C in 50 mM Tris-HCl (pH 9.0), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, and 1 mM spermidine. The RNAs were purified by multiple extractions with Tris-saturated phenol and one extraction with 24:1 chloroform/isoamyl alcohol followed by ethanol precipitation. The RNAs were 5′-end-labeled with 0.5 μM [γ-^{32}P]ATP (6000 Ci/mmol) (ICN) per 100 pmol of RNA by incubating with 16 units of T4 polynucleotide kinase (New England Biolabs) in 70 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, and 5 mM DTT (*53*, *54*). RNAs were gel purified on a denaturing gel, visualized by autoradiography, and recovered from gels as described above.

RNA Sequencing. Alkaline hydrolysis of RNAs was carried out in hydrolysis buffer for 8−12 min at 85 °C. All ribonucleases were purchased from Pharmacia. RNAs were incubated with 0.1 unit of RNase from *Bacillus cereus* per picomole of RNA for 4 min at 55 °C in 16 mM sodium citrate (pH 5.0), 0.8 mM EDTA, and 0.5 mg/mL yeast tRNA (Gibco-BRL). This enzyme yields U- and C-specific cleavage of RNA. For RNase T1 digestion, RNAs were incubated with 0.05 unit of RNase per picomole of RNA for 10 min at 37 °C in 40 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 40 μg of yeast tRNA. T1 cuts RNA at G nucleotides. Sequencing products were resolved on 20% denaturing gels and visualized by phosphor image analysis.

*Synthesis of an EDTA Analog (Compound **2**).* (*p*-Aminobenzyl)-EDTA tetra-*tert*-butyl ester (compound **1**) (100 mg, 0.16 mmol) was dissolved in THF (15 mL), and succinic anhydride (200 mg, 2 mmol) was added followed by triethylamine (3 mL). The mixture was stirred for 2 h at room temperature. After evaporation of the solvent, the residue was balanced between dichloromethane (30 mL) and aqueous $NaHCO_3$ (20 mL). The organic solvent was then washed with water (15 mL) and brine (20 mL). After drying over Na_2SO_4, the solvent was evaporated, and the product was obtained (100 mg, 0.138 mmol, 95% yield). MS (M + H = 722.3).

^{1}H NMR in $CDCl_3$: δ 7.44 (2H, d, J = 8.42 Hz), 7.11 (2H, d, J = 8.42 Hz), 3.48 (4H, s), 3.43 (4H, s), 3.03 (1H, m), 2.81 (2H, m), 2.71 (4H, s), 2.55 (2H, m), 1.43 (18H, s), 1.40 (18H, s).

^{13}C NMR: δ 176.42, 171.88, 171.26, 171.16, 171.07, 169.52, 130.54, 129.98, 120.94, 68.96, 65.27, 56.69, 55.35, 53.92, 53.83, 35.69, 32.56, 30.79, 28.57.

Peptide Synthesis. All Fmoc-amino acids, piperidine, 4-(dimethylamino)pyridine, dichloromethane, *N*,*N*-dimethylformamide, 1-hydroxybenzotriazole (HOBT), 2-(1*H*-benzotriazo-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), diisopropylethylamine, and HMP-linked polystyrene resin were obtained from the Applied Biosystems Division of Perkin-Elmer. Trifluoroacetic acid, 1,2-ethanedithiol, phenol, and thioanisole were from Sigma. Tat-derived peptide (from amino acids 38 to 72) was synthesized on an Applied Biosystems 431A peptide synthesizer using standard FastMoc protocols. Conjugation of the (*p*-aminobenzyl)-EDTA to the amino terminus of the Tat(38−72) peptide was also accomplished on the automated synthesizer. The Tat peptide (0.01 mmol) was placed in the reaction vessel, and 0.05 mmol of compound **2** (Figure 2A) was transferred to an amino acid cartridge for coupling reaction. Standard FastMoc protocols were used to couple compound **2** to the Tat peptide. Cleavage and deprotection of the peptide were carried out in 2 mL of Reagent K for 6 h at room temperature. Reagent K contained 1.75 mL of TFA, 100 μL of thioanisole, 100 μL of water, and 50 μL of ethanedithiol (*55*). After cleavage from the resin, peptide was purified by HPLC on a Zorbax 300 SB-C_8 column. The masses of fully deprotected and purified peptides were confirmed by FAB mass spectrometry: mass for Tat(38−72) = 4083.7 (M + H); mass for EDTA−Tat(38-72) = 4563.3 (M + H).

Site-Specific Cleavage of RNA. A typical cleavage reaction mixture (10 μL) contained 5′-end-labeled TAR RNA (0.04 μM) and 0.2 μM EDTA−Tat(38−72) or unmodified Tat(38−72) in a buffer containing 50 mM Tris-HCl (pH 7.4) and 20 mM KCl. After the RNA−protein complexes were

incubated at room temperature for 1 h, the cleavage reaction was initiated by the addition of 2 μL of 100 μM $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$ followed by 1 μL of 20 mM sodium ascorbate and 1 μL of 0.6% H_2O_2. The reaction was allowed to proceed for 5 min, quenched by the addition of 10 μL of 2× sample loading buffer, and electrophoresed on a 7 M urea/20% polyacrylamide gel. Cleavage products were visualized by a Phosphor Image analysis (Molecular Dynamics).

RESULTS

Synthesis of an Iron-EDTA-Labeled Tat(38−72). The experimental strategy for site-specific chelate conjugation of Tat(38−72) is outlined in Figure 2. A protected EDTA analog, (*p*-aminobenzyl)-EDTA tetra-*tert*-butyl ester (compound **1**), was synthesized as described by Song and Rana (*56*). Reagent **2** [succinamide derivative of a *tert*-butyl-protected (aminobenzyl)-EDTA] was prepared by treating compound **1** with succinic anhydride in THF and triethylamine. Reagent **2** contains only one free carboxylic acid group which can be attached to biomolecules. A Tat-derived peptide (from amino acids 38 to 72) was synthesized by using standard HOBT/HBTU FastMoc protocols (*57*). EDTA attachment to the N terminus of the Tat(38−72) peptide was achieved by using compound **2** and standard peptide coupling reagents. After cleavage from the resin and deprotection, peptide was purified by reverse phase high-performance liquid chromatography. Since peptide cleavage and deprotection were carried out in trifluoroacetic acid, *tert*-butyl protecting groups of compound **2** were also cleaved and the free EDTA moiety was obtained. The mass of the fully deprotected and HPLC pure EDTA−peptide conjugate was confirmed by fast atom bombardment (FAB) mass spectrometry.

Determination of Dissociation Constants. To further characterize and evaluate the binding capabilities of the EDTA−peptide conjugate, we determined the dissociation constants for EDTA−peptide and compared them with those of the wild-type peptide [Tat (38−72)]. Equilibrium dissociation constants of the Tat(38−72)−TAR complexes were measured using direct and competition electrophoretic mobility assays (*22, 23, 51, 58*). For direct mobility shift assays, the fractional saturation of 50 nM 5′-^{32}P-end-labeled TAR RNA was measured as a function of wild-type and EDTA-modified Tat(38−72). A typical gel of these experiments is shown in Figure 3. Dissociation constants were calculated from multiple sets of experiments which showed that Tat(38−72) and Fe−EDTA−Tat bind TAR RNA with a K_d of 0.14 and 0.18 μM, respectively. A relative dissociation constant (K_{rel}) can be determined by measuring the ratios of wild-type Tat peptide to the Fe−EDTA−Tat dissociation constants (K_d) for TAR RNA. Our results demonstrate that the calculated value for K_{rel} was 0.77, indicating that the metal chelate attachment did not significantly alter the TAR binding affinities of the Tat peptide.

Site-Specific Cleavage of TAR RNA. EDTA-modified Tat(38−72) peptide was used to form a complex with 5′-^{32}P-end-labeled TAR RNA at room temperature in TK buffer, and metal-catalyzed cleavage reactions were performed as described in Experimental Procedures. Cleavage products were separated by denaturing polyacrylamide gel electrophoresis. Results of these experiment are shown in Figure 4. The cleavage sites in TAR RNA were assigned by comparison with 5′-end-labeled products of alkaline hydrolysis, random Fe−EDTA cleavage ladder of the RNA, and ribonuclease *B. cereus* and T1 reactions. As shown in Figure 4, cleavage by Fe−EDTA−Tat was not random and few specific sites were cleaved. Specific cleavage of TAR RNA was observed at the bulge residues (U23, C24, and U25), in the loop region (G34 and A35), and at the strand opposite the bulge (U40 and C41). Intensities of the cleavage product bands on the gel showed that U23 and C24 were cleaved most efficiently while A35 and C41 were cleaved with the lowest yields. The overall order of the cleavage efficiency was U23 and C24 > C25, G34, and U40 > A35 and C41.

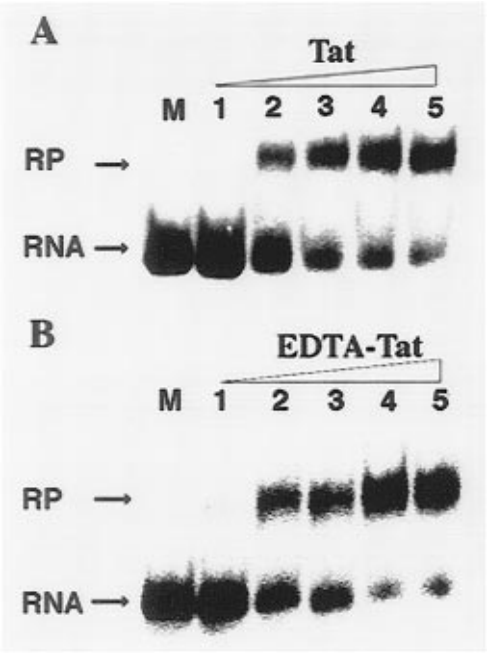

FIGURE 3: Binding of TAR RNA to the Tat(38−72) peptide (A) or the Fe(II)−EDTA−Tat conjugate (B). Binding reaction mixtures contained 0.05 μM 5′-^{32}P-end-labeled TAR RNA and 0.1, 0.2, 0.3, 0.6, and 0.7 μM Tat(38−72) or Fe(II)−EDTA−Tat conjugate (lanes 1−5), respectively. Lane M was a control lane without the peptide. Complex formation was performed in the TK buffer and incubation at room temperature for 1 h. Complexes were separated from unbound RNA by electrophoresis in nondenaturing 8% polyacrylamide gels containing 0.1% Triton X-100. Gels were run in a cold room at 300 V for 2 h. The relative amounts of free and bound RNA were determined by phosphor imaging. RNA−peptide complexes are shown as RP.

Figure 4 shows that no detectable cleavage of TAR RNA was observed when the RNA was subjected to cleavage reaction conditions in the presence of 20 μM Fe−EDTA complex without the addition of EDTA−Tat conjugate (lane 3). Lane 2 shows that TAR RNA was stable during experimental procedures and no background RNA degradation was detected. Control experiments showed that ascorbate and hydrogen peroxide were necessary for the cleavage reaction because the EDTA−Tat conjugate could not cleave RNA without the cleavage buffer (lane 4). Further control experiments established that unmodified Tat(38−72) did not cleave TAR RNA under standard cleavage conditions (data not shown). The specificities of the cleavage reactions were determined by competition experiments. Addition of unlabeled wild-type TAR RNA abolished the cleavage, while the addition of unlabeled mutant (G26−C39 to C26−G39)

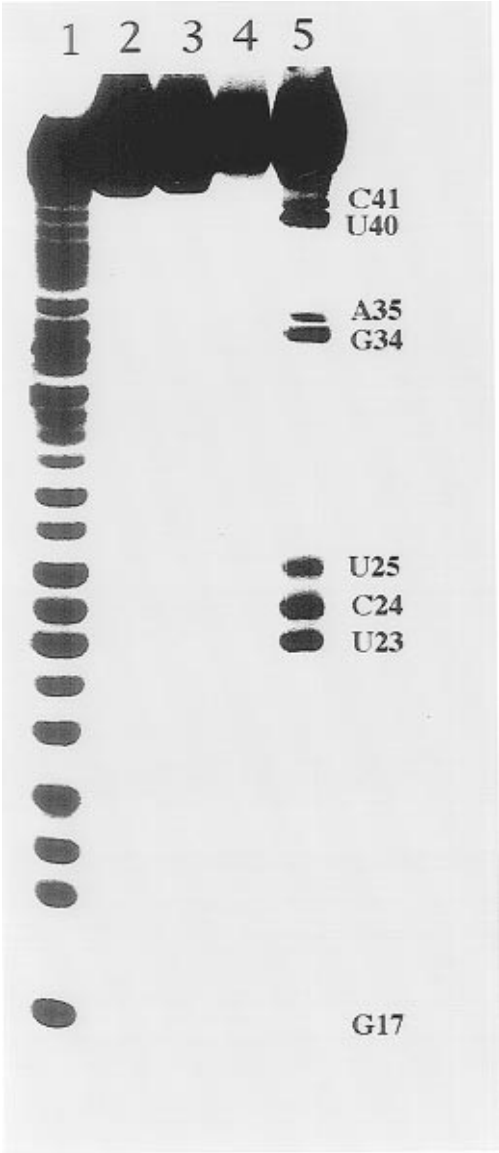

FIGURE 4: Cleavage of 5′-end-labeled wild-type TAR RNA by the Fe(II)−EDTA−Tat conjugate: lane 1, alkaline hydrolysis of RNA; lane 2, RNA without peptide or cleavage reaction; lane 3, RNA subjected to the cleavage reaction in the presence of 20 μM Fe−EDTA complex without the addition of Fe(II)−EDTA−Tat conjugate; lane 4, RNA and Fe(II)−EDTA−Tat conjugate incubated together witout the addition of ascorbate and H_2O_2; and lane 5, RNA and Fe(II)−EDTA−Tat conjugate subjected to the cleavage reaction. Major sites of cleavage are marked.

TAR RNA had no effect on the cleavage reaction (data not shown). These results demonstrate that the Fe−EDTA chelate attached to the aminoterminus of Tat(38−72) lies in close proximity to the bulge region and two nucleotides (G34 and A35) from the loop sequence also come in contact with the iron chelate.

Cleavage in the loop sequence at G34 and A35 is intriguing and posed the question of whether this cleavage was due to the bend structure of TAR RNA or the result of a flexible loop structure which could bring specific residues into the vicinity of the iron chelate. To answer this question, we synthesized a mutant TAR RNA without the loop structure but with a wild-type bulge sequence (Figure 1). Nucleotides above the bulge region were base-paired in mutant RNA. EDTA-modified Tat(38−72) peptide was used to form a complex with 5′-^{32}P-end-labeled mutant TAR RNA at room temperature in TK buffer, and metal-catalyzed

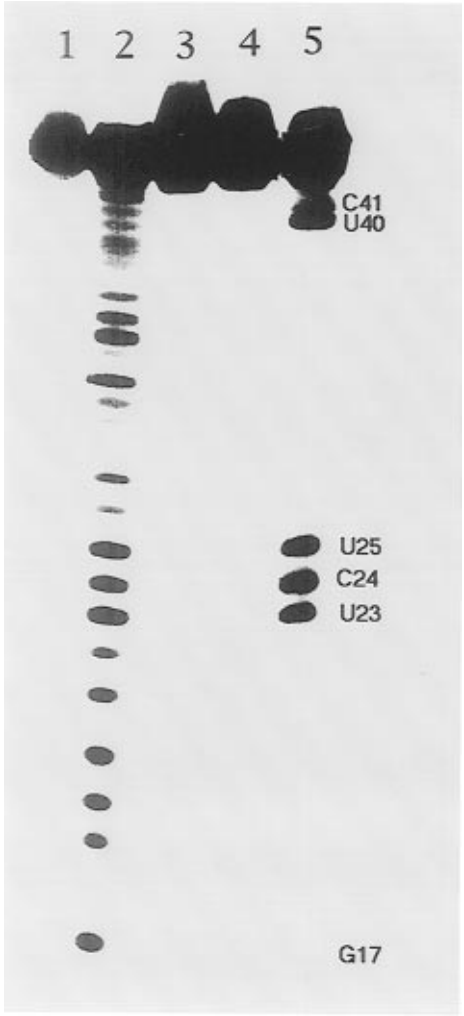

FIGURE 5: Cleavage of 5′-end-labeled mutant TAR RNA by the Fe(II)−EDTA−Tat conjugate: lane 1, RNA without peptide or cleavage reaction; lane 2, alkaline hydrolysis of RNA; lane 3, RNA subjected to the cleavage reaction in the absence of Fe(II)−EDTA−Tat conjugate; lane 4, RNA and Fe(II)−EDTA−Tat conjugate incubated together without the addition of ascorbate and H_2O_2; and lane 5, RNA and Fe(II)−EDTA−Tat conjugate subjected to the cleavage reaction. Major sites of cleavage are marked.

cleavage reactions were performed as described above. Figure 5 shows the results of this experiment. Specific cleavage of mutant TAR RNA was observed only at the bulge region (U23, C24, and U25) and at the strand opposite the bulge (U40 and C41). There was no cleavage observed at nucleotide position 34 or 35 in mutant TAR RNA (lane 5). Similar to the results shown in Figure 4, control experiments showed that Fe−EDTA−Tat, ascorbate, and hydrogen peroxide were necessary for the cleavage reaction (lanes 3−5). This result indicates that the loop sequence of wild-type TAR RNA is involved in the observed cleavage at G34 and A35 residues.

DISCUSSION

We have used a site-directed RNA-cleaving strategy to determine the neighborhood of the core domain of a Tat fragment in the Tat−TAR complex. The results of our affinity cleaving experiments demonstrate that, in solution, an iron−EDTA attached at the N terminus of the core domain of Tat(38−72) is located in the proximity of the bulge region and two nucleotides from the loop sequence also come in close contact with the iron chelate.

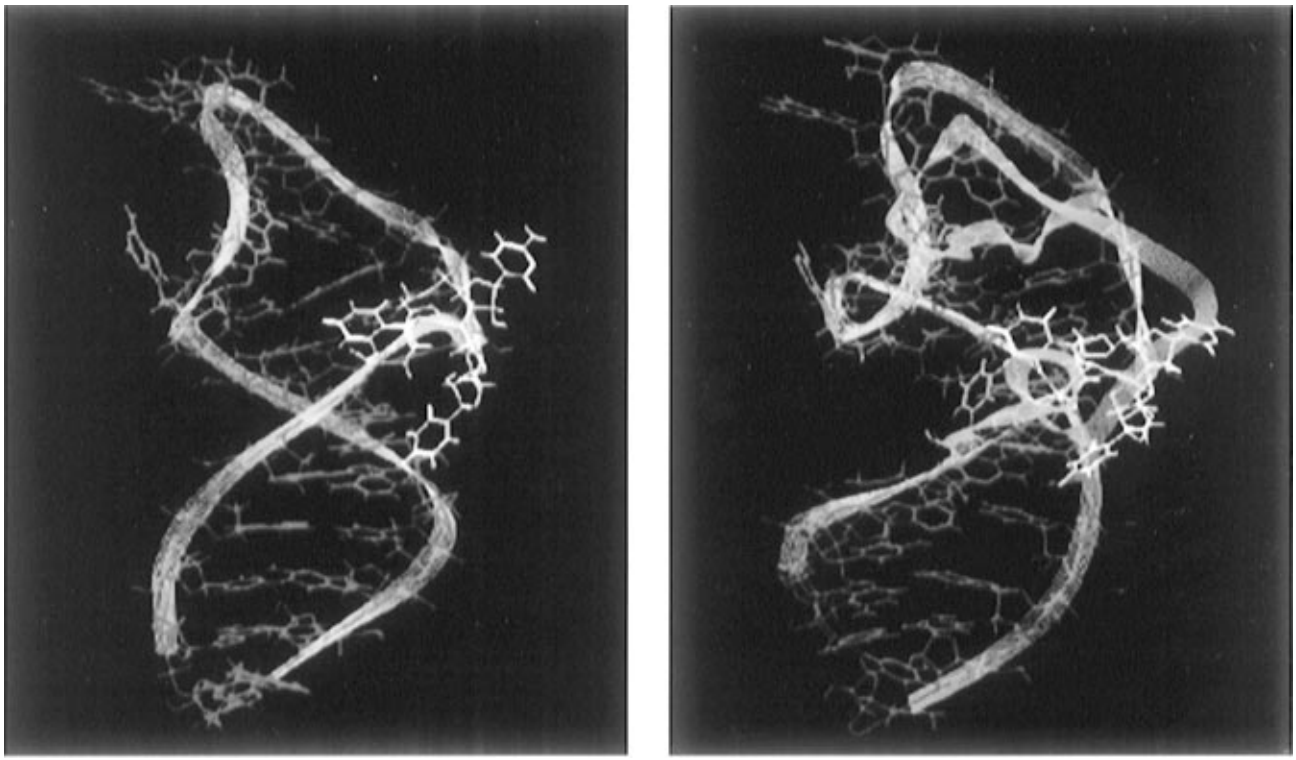

FIGURE 6: Proposed model for Tat peptide−TAR interactions showing the protein orientation and the proximity of the amino terminus to the peptide in the RNA−protein complex. (Left) A view of the TAR RNA structure in the presence of L-argininamide (*63*) displaying nucleotides cleaved by Fe(II)−EDTA−Tat conjugate: trinucleotide bulge residues (white), U40 and C41 (magenta), and G34 and A35 (green). The ribbon structure of TAR RNA is shown in yellow lines, and nucleotides as shown in red. (Right) TAR RNA in the presence of Tat peptide showing that the RNA is wrapped around the peptide and the loop is located in the proximity of the bulge. The ribbon structure of the Tat peptide (in lines) and the amino-terminal Phe38 are shown in cyan. The ribbon structure of the Tat peptide is drawn from the Tat protein structure (*68*). The orientation of the Tat peptide is based on previous photo-cross-linking results indicating that Lys41 and Arg57 are close to U42 and U31, respectively (*65, 76*). Structures of RNA and protein were visualized using Insight II software on an IRIS workstation.

Attachment of an Fe(II)−EDTA at the amino terminus of Tat(38−72) localizes the production of hydroxyl radicals generated by the addition of ascorbate and hydrogen peroxide. Hydroxyl radical species are short-lived, and the cleavage of nucleic acids is observed within a 10 Å radius from the site of production of these radicals (*31, 59*). Since the linker between the amino terminus of Tat(38−72) and the EDTA moiety is about 15 Å, the overall RNA probing radius would be ≈25 Å from the amino terminus of the Tat peptide which is approximately 10 nucleotides in an A-form double helix. If the cleavage is nonspecific or the RNA has no structure, most of the nucleotides in TAR RNA should be cleaved. Our results show that the cleavage reaction is specific and affinity cleaving can be used to probe RNA−protein interactions in small ribonucleoprotein complexes with well-defined structures. Cleavage in the bulge region of TAR RNA indicates that these nucleotides are within the reach of Fe(II)−EDTA tethered to the amino terminus of the Tat fragment. Hydroxyl radicals are readily diffusible, nonspecific species and can attack virtually any region of the RNA backbone; therefore, the intensity of the cleavage products may indicate the extent of proximity. We observed the strongest cleavage at U23 and C24, while cleavage reactions at A35 and C41 were the least efficient. Cleavage at other nucleotides (U25, G34, and U40) resulted in the production of medium-intensity bands. Three possible explanations can be advanced to interpret this result. (1) U23 and C24 in the bulge region are located in the nearest neighborhood of the amino terminus of Tat(38−72). (2) RNA cleavage is not uniform at all the cleaved nucleotides. Due to the RNA conformation at the bulge region, the backbone near U23 and C24 is more exposed to the solvent than the other residues which results in more cleavage at U23 and C24. (3) Dynamics of TAR RNA structure play a role in the observed difference of RNA cleavage efficiencies by bringing certain residues into the vicinity of the Fe(II)−EDTA complex for varying amounts of time. For example, the hairpin loop of TAR RNA has a flexible structure in solution, and there is no compelling evidence that shows the loop is closed by any non-Watson−Crick pairs, such as C·AH$^+$ or G·U base pairs (*60*). This dynamic nature of the loop may form structured regions transiently, removing G34 and A35 from the proximity of the bulge region and therefore resulting in a decreased cleavage at these two nucleotides. Alternatively, loop residues may come in close proximity to the metal chelate for a short period of time. Similarly, breathing of the RNA structure at the bulge region may translate into a reduced cleavage at U40 and C41 nucleotides. Site-specific photo-cross-linking experiments support the notion that the TAR RNA structure at the bulge region is dynamic and U40 is not base-paired with A22 all the time (*53*).

Our results also demonstrate that, in a Tat−TAR compex, two nucleotides (G34 and A35) of the TAR RNA loop sequence are located in close proximity to the trinucleotide bulge region. Since there is a small amount of free Fe(II) in the reaction medium which does not cleave unbound TAR RNA (Figure 4, lane 3), it could be argued that Tat binding causes structural changes in TAR RNA and makes the loop sequence hypersensitive to cleavage by hydroxyl radicals

produced by unchelated Fe(II). To test this possibility, we used an unmodified Tat(38−72) peptide to form a Tat−TAR complex and subjected it to standard cleavage conditions containing unchelated Fe(II) concentrations used in Figure 4. There was no detectable RNA cleavage observed (data not shown). Cleavage in the loop sequence by Fe(II)−EDTA−Tat is an interesting result and raises two possibilities about TAR RNA structure in a Tat−TAR complex. (a) The trinucleotide bulge sequence causes a bend in TAR RNA structure and brings the loop region near the bulge. Previous experiments using gel electrophoresis suggest that the pyrimidine bulge of TAR RNA causes a local bending of the helical axis (*61*). Zacharias and Hagerman (*62*) performed transient electric birefringence measurements which showed that TAR RNA bulge introduces a bend of 50° in the absence of Mg^{2+}, which is straightened by the addition of Arg and Tat-derived peptides. (b) It is the loop sequence which is located in close proximity to the bulge in a tertiary RNA−protein structure, and bending of the helix does not cause cleavage at G34 and A35 residues. Recent NMR models suggested close proximity between the UCU bulge and the apical loop of TAR RNA across the RNA major groove (*63*). To test this hypothesis, we synthesized a mutant TAR RNA containing a wild-type bulge sequence and without a loop structure (Figure 1). We reasoned that, if the loop structure is responsible for the observed cleavage at G34 and A35, replacing the loop sequence with a duplex structure should abolish cleavage at G34 and A35. On the other hand, if helix bending due to the bulge sequence is responsible for cleavage at G34 and A35, we should still observe cleavage in the region of these two nucleotides. The results depicted in Figure 5 (lane 5) indicate that cleavage occurred only at the bulge region. Thus, our results establish that, in a Tat−TAR complex, the TAR RNA loop sequence is positioned in the neighborhood of the UCU bulge region.

How does Tat interact with TAR RNA? Previous studies suggest that Tat protein contacts TAR RNA in a widened major groove (*54, 64*). In a recent study from our laboratory, we used a rhodium complex, $Rh(phen)_2phi^{3+}$, to probe the effect of bulge bases on the major groove width in TAR RNA (*54*). Our studies establish two important factors involved in Tat−TAR recognition. (i) There is a correlation between major groove opening and Tat binding. A bulge of at least two bases is required for major groove widening and other conformational changes to facilitate Tat binding. This cannot be accomplished by a single-base bulge. (ii) A Tat fragment (42−72) occupies the major groove of TAR RNA and abolishes access of the rhodium complex.

We devised a new method based on psoralen photochemistry to determine the relative orientation of the nucleic acid and protein in the Tat−TAR complex (*65*). We synthesized a 30-amino acid fragment containing the arginine-rich RNA-binding domain of Tat(42−72) and chemically attached a psoralen at the amino terminus. Upon near-ultraviolet irradiation (360 nm), this synthetic psoralen−peptide cross-linked to a single site in the TAR RNA sequence. The RNA−protein complex was purified, and the cross-link site on TAR RNA was determined by chemical and primer extension analyses. Our results show that the amino terminus of Tat(42−72) contacts, or is close to, uridine 42 in the lower stem of TAR RNA (*65*). In another study, we synthesized Tat(38−72) and replaced Arg57 with Cys to introduce a unique thiol group (SH) into the peptide (*76*). A psoralen derivative which can react with thiol groups was synthesized and used for specific chemical modification of Cys57−Tat-(38−72). We used this psoralen−Tat conjugate [psoralen−Cys57−Tat(38−72)] to form a specific complex with TAR RNA. Photo-cross-linking and RNA sequencing results revealed that Cys57 of Tat(38−72) is close to U31 of TAR RNA (*76*).

On the basis of the above studies, we propose a model for Tat−TAR recognition in which Tat makes initial contacts with TAR RNA in the widened major groove and causes a structural change in RNA where the amino terminus of the Tat peptide lies in close proximity to the bulge and loop nucleotides (Figure 6). To fit together our Fe(II)−EDTA−Tat cleavage results with previous psoralen photo-cross-linking findings, TAR RNA structure must adopt a conformation such that TAR RNA is folded around the Tat peptide. This model is consistent with recent NMR studies which suggest that RNA in TAR RNA−peptide complexes has a rigid structure which is folded around the basic region of Tat (*63, 66*). Although some α-helical tendency appears within the basic region of an HIV−EIAV tat hybrid peptide (*67*), the basic region of Tat has an extended and disordered chain in solution (*68*). It is possible that the basic region of Tat initiates contacts with TAR RNA in the major groove, causing RNA to undergo a structural change which narrows the major groove making it appear as if the RNA is wrapped around the peptide. During this RNA−protein interaction, the basic region folds into an ordered structure. Short basic peptides which are unfolded in solution show structural changes after binding to TAR RNA (*47, 63*). For illustration purposes, the ribbon structure of Tat(38−72) peptide is based on the Tat protein structure. However, it remains to be established whether Tat peptide models recognize TAR RNA in a fashion similar to that of Tat protein or not.

Mutational analyses have shown that sequences in the loop of TAR RNA are required for transactivation (*24, 25*), and not for Tat binding (*26, 27, 69*). The loop may provide the binding site for cellular factor(s) involved in transactivation (*70*−*73*). Recently, it has been shown that the cellular factor TRP-185 regulates RNA polymerase II binding to HIV-1 TAR RNA (*74*). A group of cellular cofactors that stimulate the binding of RNA polymerase II and TRP-185 to TAR RNA has also been identified (*75*). Tat could also be involved in rearranging the loop structure that can be recognized by cellular factors. Our results show that, in a Tat−TAR complex, the loop is arranged to position G34 and A35 in the neighborhood of the bulge region. Information about the tertiary structure of loop nucleotides when TAR is bound to Tat is very important in defining the role of cellular factors that bind to TAR RNA. These studies are crucial in understanding the mechanism of gene regulation in HIV-1 and designing new drugs to inhibit HIV-1 gene expression.

ACKNOWLEDGMENT

We thank Heather Neenhold and Anne In Song for preliminary experiments.

REFERENCES

1. Jones, K. A., and Peterlin, B. M. (1994) *Annu. Rev. Biochem. 63,* 717−743.
2. Jeang, K.-T., Berkhout, B., and Dropulic, B. (1993) *J. Biol. Chem. 268,* 24940−24949.

3. Gaynor, R. (1992) *AIDS 6*, 347−363.
4. Fisher, A. G., et al. (1986) *Nature 320*, 367−371.
5. Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C., and Haseltine, W. A. (1986) *Cell 44*, 941−947.
6. Cullen, B. R. (1992) *Microbiol. Rev. 56*, 375−394.
7. Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989) *Cell 59*, 273−282.
8. Cullen, B. R. (1986) *Cell 46*, 973−982.
9. Peterlin, B. M., Luciw, P. A., Barr, P. J., and Walker, M. D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9734−9738.
10. Muesing, M. A., Smith, D. H., and Capon, D. A. (1987) *Cell 48*, 691−701.
11. Rice, A. P., and Mathews, M. B. (1988) *Nature 332*, 551−553.
12. Laspia, M. F., Rice, A. P., and Mathews, M. B. (1989) *Cell 59*, 283−292.
13. Marciniak, R. A., and Sharp, P. A. (1991) *EMBO J. 10*, 4189−4196.
14. Kao, S.-Y., Calman, A. F., Luciw, P. A., and Peterlin, B. M. (1987) *Nature 330*, 489−493.
15. Zhou, Q., and Sharp, P. A. (1995) *EMBO J. 14*, 321−328.
16. Graeble, M. A., Churcher, M. J., Lowe, A. D., Gait, M. J., and Karn, J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 6184−6188.
17. Churcher, M. J., Lowe, A. D., Gait, M. J., and Karn, J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2408−2412.
18. Rosen, C. A., Sodroski, J. G., and Haseltine, W. A. (1985) *Cell 41*, 813−823.
19. Jakobovits, A., Smith, D. H., Jakobovits, E. B., and Capon, D. J. (1988) *Mol. Cell. Biol. 8*, 2555−2561.
20. Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989) *Genes Dev. 3*, 547−558.
21. Garcia, J. A., Harrich, D., Soultanakis, E., Wu, F., Zmitsuyasu, R., and Gaynor, R. B. (1989) *EMBO J. 8*, 765−778.
22. Dingwall, C., et al. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6925−6929.
23. Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., and Skinner, M. A. (1990) *EMBO J. 9*, 4145−4153.
24. Feng, S., and Holland, E. C. (1988) *Nature 334*, 165−167.
25. Berkhout, B., and Jeang, K.-T. (1989) *J. Virol. 63*, 5501−5504.
26. Cordingley, M. G., et al. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8985−8989.
27. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., and Sonenberg, N. (1991) *J. Virol. 65*, 5196−5202.
28. Tinoco, I., Jr., Puglisi, J. D., and Wyatt, J. R. (1990) *Nucleic Acids Mol. Biol. 4*, 205−226.
29. Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989) *Bioassays 11*, 100−106.
30. Tullius, T. D. (1989) in *Metal-DNA chemistry* (Tullius, T. D., Ed.) pp 1−23, American Chemical Society, Washington, DC.
31. Tullius, T. D., and Dombroski, B. A. (1985) *Science 230*, 679−681.
32. Price, M. A., and Tullius, T. D. (1992) *Methods Enzymol. 212*, 194−219.
33. Pogozelski, W. K., McNeese, T. J., and Tullius, T. D. (1995) *J. Am. Chem. Soc. 117*, 6428−6433.
34. Dervan, P. B. (1986) *Science 232*, 464−471.
35. Celander, D. W., and Cech, T. R. (1991) *Science 251*, 401−407.
36. Han, H., and Dervan, P. B. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 4955−4959.
37. Chen, C.-H. B., and Sigman, D. S. (1987) *Science 237*, 1197−1201.
38. Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. I., and Dervan, P. B. (1987) *Science 238*, 1129−1132.
39. Moser, H. E., and Dervan, P. B. (1987) *Science 238*, 645−650.
40. Ebright, R. H., Ebright, Y. W., Pendergrast, P. S., and Gunasekera, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2882−2886.
41. Sigman, D. S., Kuwabara, M. D., Chen, C.-H. B., and Bruce, T. C. (1991) *Methods Enzymol. 208*, 414−433.
42. Murakami, K., Kimura, M., Owens, J. T., Meares, C. F., and Ishihama, A. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1709−1714.
43. Heilek, G. M., Marusak, R., Meares, C. F., and Noller, H. F. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 1113−1116.
44. Heilek, G. M., and Noller, H. F. (1996) *Science 272*, 1659−1662.
45. Heilek, G. M., and Noller, H. F. (1996) *RNA 2*, 597−602.
46. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., and Crothers, D. M. (1990) *Science 249*, 1281−1285.
47. Calnan, B. J., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Genes Dev. 5*, 201−210.
48. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Science 252*, 1167−1171.
49. Delling, U., Roy, S., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C. A., and Sonenberg, N. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 6234−6238.
50. Weeks, K. M., and Crothers, D. M. (1993) *Science 261*, 1574−1577.
51. Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. C., Gait, M. J., and Karn, J. (1993) *J. Mol. Biol. 230*, 90−110.
52. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
53. Wang, Z., and Rana, T. M. (1996) *Biochemistry 35*, 6491−6499.
54. Neenhold, H. R., and Rana, T. M. (1995) *Biochemistry 34*, 6303−6309.
55. King, D. S., Fields, C. G., and Fields, G. B. (1990) *Int. J. Pept. Protein Res. 36*, 255−266.
56. Song, A. I., and Rana, T. M. (1997) *Bioconjugate Chem. 8*, 249−252.
57. Knorr, R., Trzeciak, A., Bannwarth, W., and Gillessen, D. (1989) *Tetrahedron Lett. 30*, 1927−1930.
58. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505.
59. Dreyer, G. B., and Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 968−972.
60. Jaeger, J. A., and Tinoco, I., Jr. (1993) *Biochemistry 32*, 12522−12530.
61. Riordan, F. A., Bhattacharyya, A., McAteer, S., and Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305−310.
62. Zacharias, M., and Hagerman, P. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6052−6056.
63. Aboul-ela, F., Karn, J., and Varani, G. (1995) *J. Mol. Biol. 253*, 313−332.
64. Weeks, K. M., and Crothers, D. M. (1991) *Cell 66*, 577−588.
65. Wang, Z., and Rana, T. M. (1995) *J. Am. Chem. Soc. 117*, 5438−5444.
66. Aboul-ela, F., Karn, J., and Varani, G. (1996) *Nucleic Acids Res. 24*, 3974−3981.
67. Mujeeb, A., Bishop, K., Peterlin, B. M., Turk, C., Parslow, T. G., and James, T. L. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8248−8252.
68. Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., and Rösch, P. (1995) *J. Mol. Biol. 247*, 529−535.
69. Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P. J. G., Karn, J., and Gait, M. J. (1993) *J. Mol. Biol. 230*, 111−123.
70. Gatignol, A., Kumar, A., Rabson, A., and Jeang, K.-T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7828−7832.
71. Marciniak, R. A., Garcia-Blanco, M. A., and Sharp, P. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3624−3628.
72. Wu, F., Garcia, J., Sigman, D., and Gaynor, R. (1991) *Genes Dev. 5*, 2128−2140.
73. Sheline, C. T., Milocco, L. H., and Jones, K. A. (1991) *Genes Dev. 5*, 2508−2520.
74. Wu-baer, F., Lane, W. S., and Gaynor, R. B. (1995) *EMBO J. 14*, 5995−6009.
75. Wu, B. F., Lane, W. S., and Gaynor, R. B. (1996) *J. Biol. Chem. 271*, 4201−4208.
76. Wang, Z., Wang, X., and Rana, T. M. (1996) *J. Biol. Chem. 271*, 16995−16998.

BI971011G