Major Groove Opening at the HIV-1 Tat Binding Site of TAR RNA Evidenced by a Rhodium Probe†

Heather R. Neenhold and Tariq M. Rana*

Department of Pharmacology, Robert Wood Johnson Medical School-UMDNJ, and Microbiology and Molecular Genetics Graduate Program at Rutgers, The State University of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854

Received November 10, 1994; Revised Manuscript Received February 7, 1995®

ABSTRACT: Transactivation of human immunodeficiency virus (HIV) gene expression requires the interaction of Tat protein with the *trans*-activation responsive region (TAR) RNA, a 59-base stem−loop structure located at the 5′-end of all mRNAs. The TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge which separates two helical stem regions. The trinucleotide bulge is essential for high affinity and specific binding of the Tat protein. Recently, a rhodium complex, $Rh(phen)_2phi^{3+}$, was discovered which promotes RNA cleavage in the open major groove and triply bonded bases [Chow, C. S., *et al.* (1992) *Biochemistry 31*, 972−982]. This metal complex does not bind double-helical RNA or unstructured single-stranded regions of RNA. Instead, sites of tertiary interaction which are open in the major groove and accessible to stacking are targeted by the complex through photoactivated cleavage. We have used this rhodium probe to investigate the effect of bulge bases on the major groove opening in TAR RNA. The sites targeted by the rhodium complex have been mapped to single nucleotide resolution on wild-type TAR RNA and on several mutants of the TAR RNA containing different numbers of mismatch bases in the bulge region. A strong cleavage at residues C39 and U40 was observed on the wild-type TAR RNA and in mutant TAR RNA containing two mismatch bases in the bulge. No cleavage at C39 and U40 was observed in a bulgeless and a one-base bulge TAR RNA. By varying the number of mismatch bases, we demonstrated that the trinuclear bulge widens the major groove of TAR RNA to facilitate Tat binding. Our studies establish three important factors involved in Tat−TAR recognition: (i) A bulge containing two or more bases permits major groove widening sufficient for Tat binding, and this cannot be accomplished by a single-base bulge; (ii) Tat fragment (42−72) binds in the major groove of TAR RNA because cleavage at C39 and U40 in wild-type TAR RNA was inhibited in the presence of Tat-(42−72); (iii) cleavage efficiency is unaffected by the presence of arginine, indicating that the conformational changes in TAR RNA upon arginine binding have no effect on the widening and accessibility of the major groove.

Human immunodeficiency virus type 1 (HIV-1),[1] like other lentiviruses, encodes a transactivating regulatory protein, Tat, essential for HIV replication (Dayton *et al.*, 1986; Fisher *et al.*, 1986). The Tat protein is a small, cysteine-rich nuclear protein which contains 86 amino acids and comprises three important functional domains. HIV-1 Tat protein acts by binding to the TAR (*trans*-activation responsive) RNA element, a 59-base stem−loop structure located at the 5′-ends of all nascent HIV-1 transcripts (Berkhout *et al.*, 1989; Feng & Holland, 1988; Hauber & Cullen, 1988; Jakobovits *et al.*, 1988; Rosen *et al.*, 1985). Upon binding to the TAR RNA sequence, Tat causes a substantial increase in transcript levels (Cullen, 1986; Laspia *et al.*, 1989; Muesing *et al.*, 1987; Peterlin *et al.*, 1986; Rice & Mathews, 1988). The increased efficiency in transcription may result from preventing premature termination of the transcriptional elongation complex (Kao *et al.*, 1987) or from enhancing initiation of transcription (Sharp & Marciniak, 1989). The TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge separating two helical stem regions (Berkhout *et al.*, 1989; Jakobovits et al., 1988; Muesing *et al.*, 1987; Rosen *et al.*, 1985). As shown in Figure 1, nucleotides spanning positions +19 to +42 are sufficient for Tat responsiveness *in vivo* (Jakobovits *et al.*, 1988). The trinucleotide bulge is essential for high affinity and specific binding of the Tat protein (Dingwall *et al.*, 1990, 1989).

The diversity of RNA structures plays a central role in their specific recognition by proteins. RNA molecules can fold into extensive structures containing regions of double-stranded duplex, hairpins, internal loops, bulged bases, and pseudoknotted structures (Tinoco *et al.*, 1990; Wyatt *et al.*, 1989). Due to the complexity of RNA structure, the rules governing sequence-specific RNA−protein recognition are not well understood. In a regular B-form DNA double helix, the major groove is wide and easily accessible to protein side chains. In contrast to DNA, double-helical regions of RNA tend to adopt an A-form conformation, where the major groove is narrower and cannot accommodate an α-helix of an RNA-binding protein. On the other hand, the minor groove of an A-form RNA helix is shallow and wide, making it easily accessible to proteins for RNA recognition. Recent

† Supported by Research Grant AI 34785 from the National Institutes of Health.

* To whom correspondence should be addressed at the Department of Pharmacology, Robert Wood Johnson Medical School-UMDNJ [telephone, (908) 235-4082 or (908) 235-4590; Fax, (908) 235-4073; E-Mail, rana@mbcl.rutgers.edu].

® Abstract published in *Advance ACS Abstracts,* May 1, 1995.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; TAR, *trans*-activation response element; rhodium probe, $Rh(phen)_2phi^{3+}$; Tat, transcription antitermination.

0006-2960/95/0434-6303$09.00/0

structural studies have demonstrated that RNA-binding proteins interact with RNA in both the minor and major grooves. For example, two tRNA synthetases (alanine and glutamine) interact with the acceptor stems of their cognate tRNAs in the minor grooves (Musier-Forsyth *et al.*, 1991; Rould *et al.*, 1989). Major groove recognition takes place between aspartyl-tRNA synthetase and its cognate tRNAs at a site of local distortion in the RNA helix (Ruff *et al.*, 1991). Bulge loops or bulges (unpaired nucleotides on one strand of a duplex) in RNA helices are potentially important in tertiary folding of RNA and in providing sites for specific RNA−protein interactions, as illustrated by TFIIIA of *Xenopus* (Baudin & Romaniuk, 1989) and the coat protein of phage R17 (Wu & Uhlenbeck, 1987). It has been proposed that bulge bases introduce distortion into the geometry of TAR RNA to enlarge the major groove for protein recognition (Weeks & Crothers, 1991). A similar model has been proposed for U1A snRNP and U1 RNA, where RNA−protein contacts take place at the junction of the RNA stem and its loop. This RNA stem−loop junction has a distorted structure with an enlarged major groove (Jessen *et al.*, 1991). In a recent report, interactions between U1 snRNA and the N-terminal domain of the human U1A protein were mapped by multidimensional heteronuclear NMR studies (Howe *et al.*, 1994). These studies showed that protein−RNA contacts occur at the single-stranded apical loop of the hairpin and also in the major groove of the helical stem at neighboring U·G and U·U non-Watson−Crick base pairs (Howe *et al.*, 1994). The crystal structure of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin also revealed that the loop sequence (AUUGCAC) interacts with the surface of the four-stranded β-sheet (Oubridge *et al.*, 1994).

Recently, Barton and co-workers discovered a rhodium complex, $Rh(phen)_2phi^{3+}$, which promotes RNA cleavage in the open major groove and in triply bonded bases (Chow *et al.*, 1992a,b). This rhodium complex (Figure 1), bis-(phenanthroline)(phenanthrenequinone diimine)rhodium(III) [$Rh(phen)_2phi^{3+}$], cleaves RNA regions which are open in the major groove due to base tilting or propeller twisting (Pyle *et al.*, 1990, 1989). $Rh(phen)_2phi^{3+}$ binds avidly by intercalation in the major groove of DNA and promotes DNA cleavage upon photoactivation (Pyle *et al.*, 1989). The mechanistic studies showed that cleavage results from direct abstraction of the C-3′ H atom from the sugar by the photoactivated metal complex (Sitlani *et al.*, 1992). The recognition criteria of $Rh(phen)_2phi^{3+}$ are extremely useful to probe RNA structure. Double-helical regions of RNA adopt an A-form helix with a deep and narrow major groove inaccessible to stacking required for the rhodium complex to bind. Cleavage studies on $tRNA^{Phe}$ showed that neither double-helical nor single-stranded regions were targeted by $Rh(phen)_2phi^{3+}$ (Chow & Barton, 1990). Consistent with the recognition criteria, $Rh(phen)_2phi^{3+}$ targets sites of tertiary interaction in tRNA, where the major groove is open and accessible for stacking. Specific recognition by the complex depends upon its shape. In contrast to other chemical cleaving reagents, the intercalator probe directly maps the major groove of the helix. Thus, this rhodium complex is a valuable probe for higher order structure of RNA.

We used this rhodium complex [$Rh(phen)_2phi^{3+}$] to investigate the effect of bulge bases on major groove opening at the HIV-1 Tat binding site of TAR RNA. By varying the number of mismatch bases in the bulge region of RNA, we demonstrated that the trinucleotide bulge widens the major groove of TAR RNA to facilitate Tat binding. Our studies establish three important factors involved in Tat−TAR recognition: (i) A bulge containing two or more bases permits major groove widening sufficient for Tat binding, and this cannot be accomplished by a single base bulge; (ii) Tat fragment (42−72) binds in the major groove of TAR RNA because TAR RNA cleavage was inhibited in the presence of Tat(42−72); (iii) cleavage efficiency is unaffected by the presence of arginine, indicating that the conformational changes in TAR RNA upon arginine binding have no effect on the widening and accessibility of the major groove.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis

DNAs. All DNAs were synthesized on an Applied Biosystems ABI 392 DNA/RNA synthesizer. The template strand encodes the sequence for the TAR RNA wild type or mutants (Figure 1). The top strand is a short piece of DNA complementary to the 3′-end of all template DNAs having the sequence 5′TAATACGACTCACTATAG3′. 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 (9 M urea, 2 mM EDTA, 50 mM Tris, 16 mM boric acid, 0.1% bromophenol blue, 0.1% xylene cyanol) 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 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. The concentration of DNA was determined by measuring absorbance at 260 nm in a Shimadzu UV spectrophotometer. Samples were stored at −20 °C.

RNAs. All RNA sequences were prepared by *in vitro* transcription using the method of Milligan *et al.* (1987). The template strand of DNA was annealed to an equimolar amount of top-strand DNA, and transcriptions were carried out in 40 mM Tris-HCl, pH 7.9, 30 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 80 mg/mL PEG, and 4.0 mM NTPs at 37 °C for 2−4 h. For reactions containing 8.0 pmol of template DNA 400−600 units of T7 polymerase (Promega) was used. 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. The RNA was stored in DEPC-treated water at −20 °C.

Mutant and wild-type TAR RNAs were 5′ dephosphorylated by incubation with calf intestinal alkaline phosphatase (Promega) for 1 h at 37 °C in 50 mM Tris-HCl, pH 9.0, 1 mM $MgCl_2$, 0.1 M $ZnCl_2$, 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 (Promega) in 70 mM Tris-HCl,

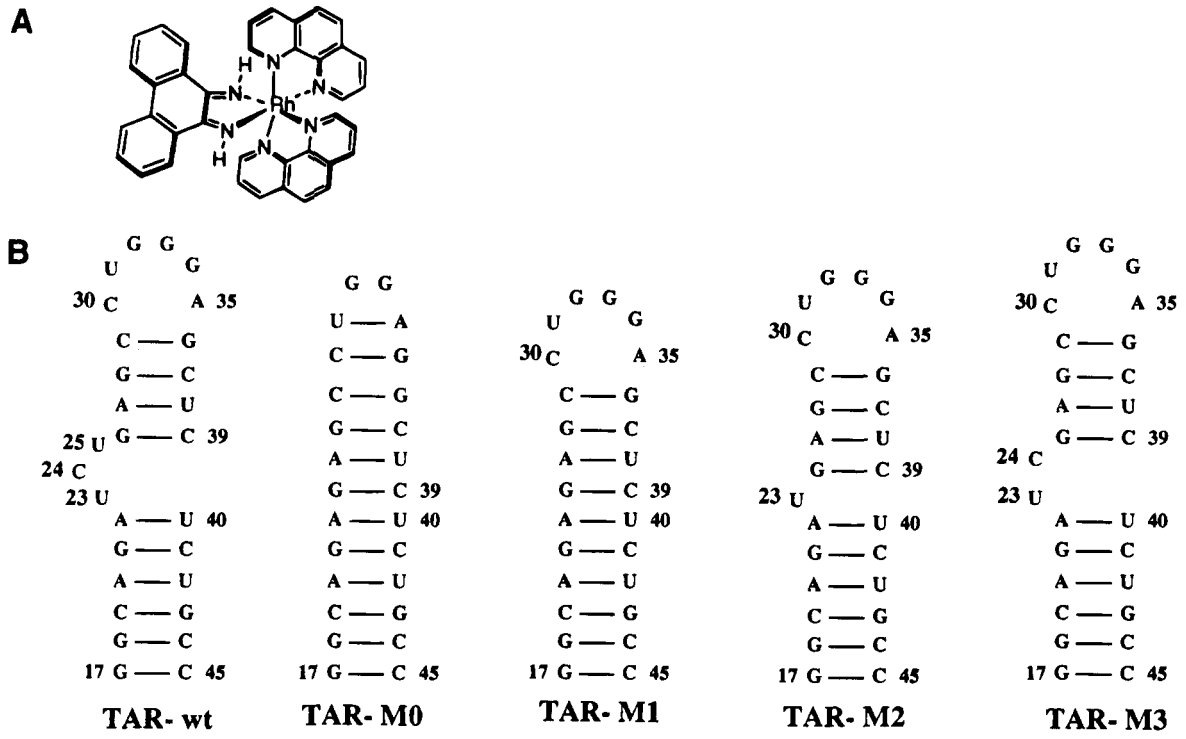

FIGURE 1: (A) Schematic illustration of the rhodium complex $Rh(phen)_2phi^{3+}$, which was synthesized as described by Pyle *et al.* (1990). (B) Secondary structures of wild-type and mutant TAR RNAs. Wild-type TAR RNA spans the minimal sequences that are required for Tat responsiveness *in vivo* (Jakobovits *et al.*, 1988) and for *in vitro* binding of Tat-derived peptides (Cordingley *et al.*, 1990). Wild-type TAR contains two non-wild-type base pairs to increase transcription by T7 RNA polymerase. Mutant TAR RNAs contained different numbers of mismatch bases in the pyrimidine bulge. All RNAs had a six-nucleotide wild-type loop except the mutant M0. Numbering of nucleotides in mutant RNAs was relative to the wild-type TAR.

pH 7.5, 10 mM $MgCl_2$, and 5 mM DTT (Kuchino & Nishimura, 1989). RNAs were gel purified on a denaturing gel, visualized by autoradiography, and recovered from gels as described above.

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 Applied Biosystems Division, Perkin-Elmer. Trifluoroacetic acid, 1,2-ethanedithiol, phenol, and thioanisole were from Sigma. A Tat-derived peptide (from amino acids 42–72) was synthesized on an Applied Biosystems 431A peptide synthesizer using standard FastMoc protocols. 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 (King *et al.*, 1990). After cleavage from the resin, the peptide was purified by HPLC on a Zorbax 300 SB-C_8 column. The mass of the peptides was confirmed by FAB mass spectrometry. Calculated mass for Tat(42–72), 3605; found, 3606 (M + H).

Cleavage Reactions

The rhodium complex, $Rh(phen)_2phi^{3+}$, was synthesized as described by Pyle *et al.* (1990) and freshly diluted in DEPC-treated water. RNA (0.3 μM) was incubated in the presence of 50 mM sodium cacodylate, pH 7.0, and 20 μM $Rh(phen)_2phi^{3+}$ for 30 min at room temperature in the dark. Cleavage was initiated by exposure of samples to ultraviolet light at 3000 Å for 10 min on ice using a 21-W lamp, Rayonet photochemical chamber reactor RPR-100 (Southern New England Ultraviolet). Cleavage products were separated on a 20% acrylamide–8 M urea gel. Cleavage products were identified by sequencing reactions (see below). Products were visualized by phosphorimage analysis (Molecular Dynamics).

Cleavage in the Presence of Arginine and a Tat-Derived Peptide

Preincubation with the Peptide. Wild-type TAR RNA (0.3 μM) was incubated first in the presence of 0.3–1.5 μM peptide, 3.0–30.0 μM L-arginine, or 1.5 μM BSA in 50 mM sodium cacodylate for 30 min at room temperature in the dark. $Rh(phen)_2phi^{3+}$ was then added to a final concentration of 20 μM, and the incubation was allowed to continue for 30 min at room temperature in the dark. Formation of the peptide–TAR complex was confirmed by gel mobility-shift assays as described by Churcher *et al.* (1993).

Preincubation with the Rhodium Complex. Wild-type TAR RNA was first incubated with 20 μM $Rh(phen)_2phi^{3+}$ 30 min at room temperature in the dark, and then peptide was added to a final concentration of 0.3 or 1.5 μM for an additional 30-min incubation. Cleavage was then initiated by exposure to ultraviolet light as described above. Interference of cleavage by the peptide, arginine, or BSA was analyzed on 20% acrylamide–8 M urea gels.

Sequencing Reactions

Alkaline hydrolysis of RNAs was carried out in 50 mM sodium carbonate, pH 9.3, for 5 min at 85 °C. RNAs were incubated with 1 unit of RNase from *Bacillus cereus*

(Pharmacia) per 3 pmol of TAR RNA for 2−3 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. Sequencing products were resolved on 20% denaturing gels and visualized by phosphorimage analysis.

RESULTS AND DISCUSSION

To probe structural flexibility and major groove opening in TAR RNA, we used a transition metal complex, Rh(phen)$_2$phi^{3+} (phen = phenanthroline; phi = phenanthrenequinone diimine), which recognizes its nucleic acid binding sites based on shape selection and promotes RNA strand cleavage upon photoactivation (Chow *et al.*, 1992a). We constructed wild-type TAR RNA, a series of mutants containing different numbers of bases at the bulge position, and one mutant without bulge and loop bases. The structures of the rhodium complex, TAR RNA, and mutant RNAs are shown in Figure 1. Syntheses of the rhodium complex and RNAs are described in Experimental Procedures.

To determine the sites of rhodium cleavage, wild-type TAR RNA was 5′ end-labeled and subjected to the cleavage reaction. The rhodium cleavage sites were assigned by comparison with 5′-end-labeled products of alkaline hydrolysis of RNA and ribonuclease *B. cereus* reactions. Partial alkaline hydrolysis causes RNA cleavage at every base, while RNase *B. cereus* reactions lead to specific cuts at C and U residues. As shown in Figures 2 and 4 (left), cleavage by the rhodium complex was not random, and few specific sites were cleaved. The cleavage was observed near the bulge bases (A22−U25), in the loop region (C30−A35), and at the strand opposite to the bulge (C39 and U40). Intensities of the cleavage were similar in the bulge and loop regions, although nucleotides C39 and U40 were cleaved more efficiently. Cleavage in the bulge bases suggests that the structure of TAR RNA is quite flexible, and therefore the rhodium complex cleaves the accessible open conformation present at that time. It has been shown by NMR spectroscopy that nucleotides C39 and U40 are stacked (Puglisi *et al.*, 1992). A strong cleavage at C39 and U40 suggests that TAR RNA has a tertiary structure in this region where the major groove is open and accessible to stacking by the complex (Chow *et al.*, 1992b). The hairpin loop of TAR RNA has a very flexible structure in solution, and there is no compelling evidence that the loop is closed by any non-Watson−Crick pairs, such as C·AH$^+$ or G·U base pairs (Jaeger & Tinoco, 1993). Cleavage in the loop nucleotides suggests the dynamic structural nature of the loop which forms structured regions transiently, since the rhodium complex does not cleave in double-helical or unstructured single-stranded regions of RNA (Chow *et al.*, 1992a).

It has been shown by competition, chemical interference, and NMR experiments that free arginine interacts with TAR in a manner similar to that of arginine in the context of Tat peptides (Puglisi *et al.*, 1992; Tan & Frankel, 1992; Tao & Frankel, 1992). Puglisi *et al.* (1992) proposed that the TAR RNA structure is significantly altered after binding to peptides containing a single arginine residue or to free argininamide. What happens to the major groove accessibility or widening in the presence of free arginine or Tat peptide? To address this question, we carried out the rhodium cleavage reactions in the presence of arginine and

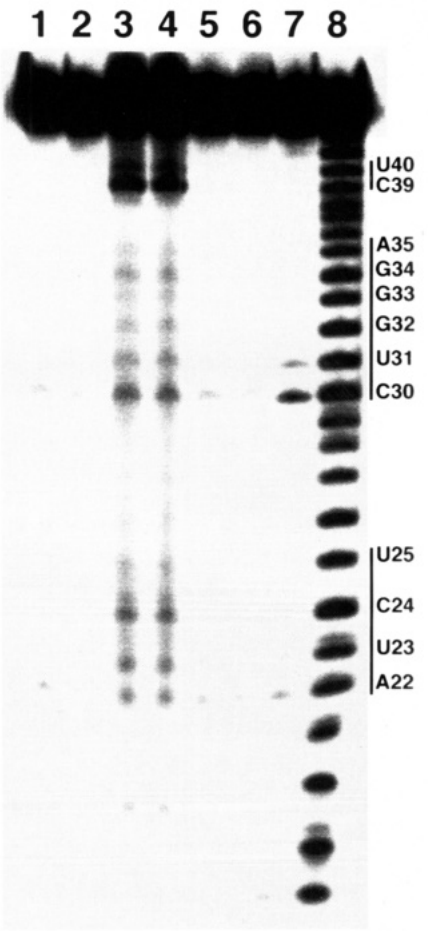

FIGURE 2: Cleavage of ^{32}P 5′-end-labeled wild-type TAR RNA by Rh(phen)$_2$phi^{3+}: lane 1, RNA without metal or irradiation; lane 2, RNA irradiated in the absence of the metal complex; lane 3, RNA cleavage with the rhodium complex and irradiation; lane 4, RNA with a 10-fold excess of arginine; lane 5, RNA with Tat peptide (1:5), preincubated with the rhodium complex; lane 6, RNA with Tat peptide (1:5), with subsequent addition of the rhodium complex; lane 7, C- and U-specific reaction; lane 8, alkaline hydrolysis of RNA. Major sites of cleavage are marked.

a Tat-derived peptide (amino acids 42−72) which specifically binds TAR RNA with high binding affinities (Churcher *et al.*, 1993; Cordingley *et al.*, 1990). As shown in Figure 2 (lanes 4−6), addition of arginine did not affect the specificity or efficiency of the cleavage, while the addition of rhodium complex before or after the Tat-derived peptide resulted in no cleavage. The presence of peptide may sterically block the access of the rhodium complex to its binding site, which leads to a loss of cleavage in the bulge region, while cleavage inhibition in the loop region is presumably due to nonspecific binding of the peptide or disruption of the loop structure upon peptide binding (Mazumder *et al.*, 1992; Pearson *et al.*, 1994). It has been shown by ribonuclease protection experiments that full-length and refolded HIV-1 Tat protein interacts with both the bulge and loop regions of TAR RNA (Harper & Logsdon, 1991). On the other hand, a 500-fold excess of arginine to RNA or addition of bovine serum albumin had no effect on the cleavage efficiencies (data not shown). These results suggest that the arginine binding to the bulge results in a conformation where the major groove is still open and accessible, as probed by the rhodium complex.

To establish whether the RNA bulge widens the major groove, four mutant RNAs (M0, M1, M2, and M3) were

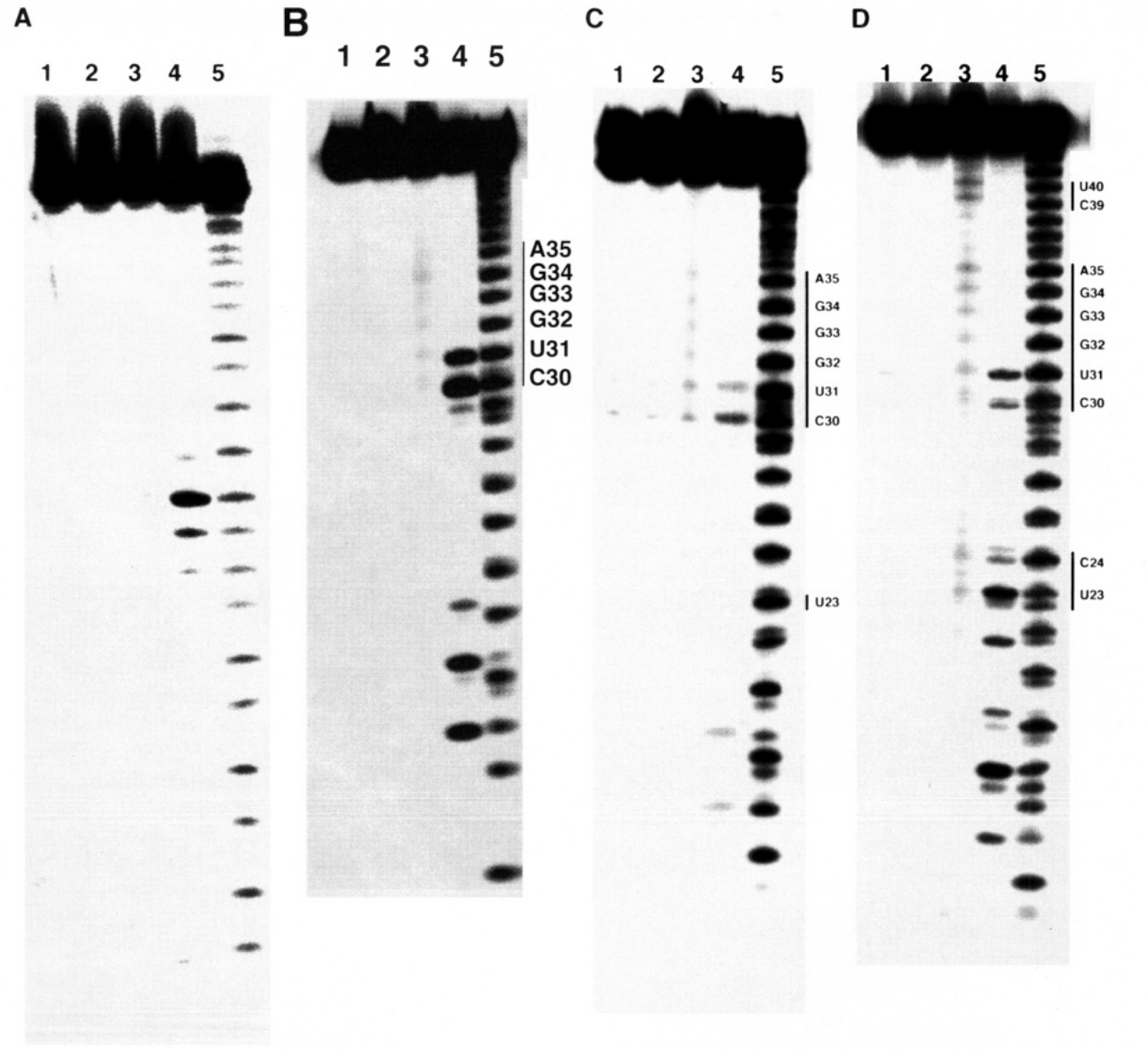

FIGURE 3: Cleavage of ^{32}P 5′-end-labeled mutant RNAs (for sequence, see Figure 1): mutant M0 (A); mutant M1 (B); mutant M2 (C); mutant M3 (D). For all four mutants: RNA without metal or irradiation (lane 1); RNA irradiated in the absence of a metal complex (lane 2); RNA cleavage with the rhodium complex and irradiation (lane 3); C- and U-specific reaction (lane 4); alkaline hydrolysis of RNA (lane 5). Major sites of cleavage are marked.

synthesized and subjected to cleavage reactions with the rhodium complex (Figure 1B). Results of these experiments are shown in Figure 3. Mutant M0 was a control sequence without a loop or bulge; it showed no cleavage, as expected (Figure 3A). Mutant M1 contained a wild-type loop but no bulge and was cleaved only in the loop region (Figure 3B). Mutants M2 (with one-base bulge and wild-type loop) and M3 (with two-base bulge and wild-type loop) showed cleavage in both the loop and bulge regions (Figure 3C,D).

Our cleavage results are consistent with the previous notion of Barton and co-workers that $Rh(phen)_2phi^{3+}$ is a sensitive probe of RNA tertiary structure (Chow *et al.*, 1992a,b). Cleavage by $Rh(phen)_2phi^{3+}$ does not occur in the canonical helical regions of the TAR RNA (Figure 3A) but at regions which are involved in unusual base pairing or bulged residues. Since double-helical RNA has a deep and narrow major groove, the rhodium complex cannot interact in this groove due to steric requirements. However, the complex can stack at sites which are accessible in a more open major groove. We observed a strong cleavage at C39 and U40 in a wild-type TAR RNA, suggesting that these bases form a stacked structure in a wide major groove. A similar stacking structure has been proposed by Puglisi *et al.* (1992) where C39 and U40 are stacked in TAR RNA. Cleavage in the bulge nucleotides was inhibited in the presence of a Tat-derived peptide but not in the presence of arginine. We carried out gel mobility-shift assays to confirm the peptide binding to the TAR RNA, and these experiments showed a gel shift for the RNA−peptide complex (data not shown). These results clearly indicate that the peptide occupies the major groove of TAR RNA. Cleavage in the loop region occurred in wild-type TAR and all mutant TAR RNAs except the mutant M0 (loopless mutant), suggesting that TAR RNA has a structured loop which could be important for cellular protein binding and *in vivo* transactivation (Marciniak *et al.*, 1990; Sheline *et al.*, 1991). Inhibition of cleavage in the loop residues upon peptide binding can be explained by two possible hypotheses: (a) peptide binds nonspecifically to the loop and blocks the access of the complex; (b) the loop has a very flexible and dynamic structure which is disrupted

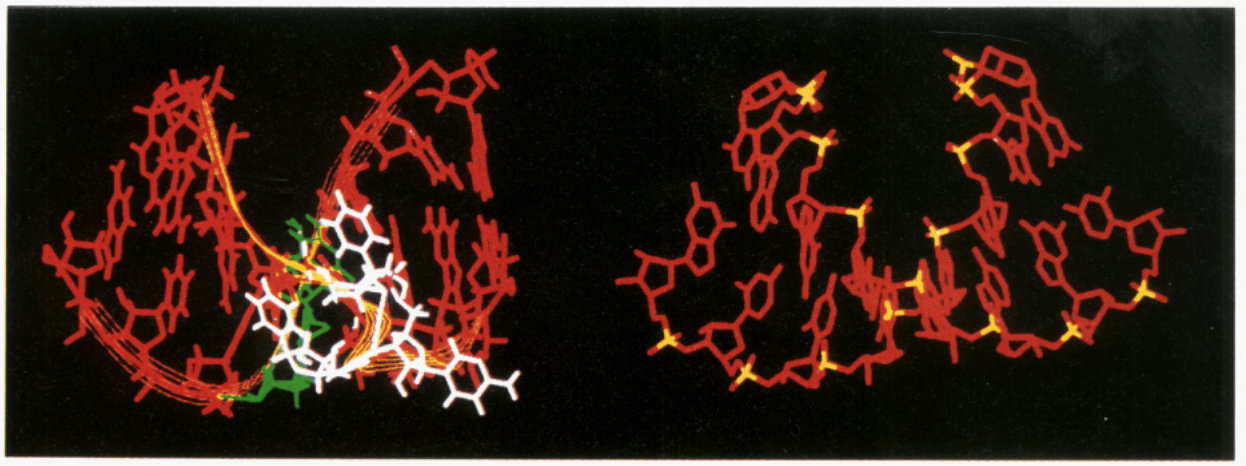

FIGURE 4: Model for the structure of the HIV-1 Tat binding site of TAR RNA showing a comparison of major groove widening in TAR RNA and A-form RNA. (Left) Distorted helical structure of TAR RNA and the cleavage data for $Rh(phen)_2phi^{3+}$. The structure contains two strands of TAR RNA; strand 1 has bases C19−C29, and the strand 2 has bases G36−G43. The phosphodiester backbone is shown in a yellow ribbon. Structures of RNA were visualized using Insight II software on an IRIS work station. The five nucleotides which are cleaved by the rhodium complex in the trinucleotide bulge region are shown: nucleotides U23, C24, and U25 (white); nucleotides C39 and U40 (green). The structure is based on NMR data provided by G. Varani and is similar to that proposed by Loret *et al.* (1992), Weeks and Crothers (1991), and Puglisi *et al.*, (1992). (Right) A-Form helix of a duplex TAR RNA without mismatch nucleotides in the bulge (Arnott *et al.*, 1972). Bases are shown in red and phosphorus atoms in yellow. Viewed from the top, the major groove is more accessible in TAR RNA (left) than the canonical A-form helix of RNA (right).

when peptide binds to a bend structure in the bulge. This suggests that another possible function of Tat is that it binds to the bulge region and arranges the loop bases in a structure that can be recognized by other cellular proteins.

In mutant M2, there was no observable cleavage at C39 and U40 bases. However, cleavage at nucleotides C39 and U40 was observed in mutant M3. These results indicate that at least a two-base bulge is required for major groove widening to allow stacking by the rhodium complex which cannot be achieved by a one-base bulge. It is interesting to note that TAR mutants with a bulge of two uridines, but not one, bind Tat peptides as well as wild-type TAR (Sumner-Smith *et al.*, 1991; Weeks & Crothers, 1991). The major groove of the bulge region in TAR RNA is opened due to the absence of base pair hydrogen bonds. A broadening of the major groove at the trinucleotide bulge could facilitate interactions with Tat protein. Our findings provide the first physical evidence that there is a correlation between major groove opening and Tat binding, suggesting that a two-base bulge permits groove widening and other conformational changes which facilitate Tat binding. This cannot be accomplished by a single-base bulge. On the basis of chemical modification and gel mobility studies, a similar model was suggested earlier by Weeks and Crothers (1991).

Bulge loops are commonly found in helical segments of cellular RNAs. The presence of bulge loops in a double-stranded RNA can introduce points of flexibility or permanent bends. The effect of bulge bases on RNA bending has been studied by gel mobility studies, and results show that the magnitude of this kinking is a function of the number and type of bases present in the bulge (Bhattacharyya *et al.*, 1990; Tang & Draper, 1990). How does a trinucleotide bulge affect TAR RNA structure? Originally, Weeks and Crothers (1991) proposed that the major groove in TAR RNA is enlarged as a consequence of local distortion in the A-form helix from stacking of the bulged uridines into the helix. This model was supported by two observations: (a) in the presence of a bulge, G26 and A27 showed increased reactivity toward diethyl pyrocarbonate (Weeks & Crothers, 1991); (b) gel mobility experiments showed that U-rich bulges markedly bend RNA duplexes (Riordan *et al.*, 1992). In another model, Delling *et al.* (1992) suggested that U23 forms a base triple with G26 and C39 by a bifurcated water molecule, and this structure produces a widened major groove of the RNA duplex with the bulge residue U25 forced out of the helix. On the basis of NMR experiments, Puglisi *et al.* (1992) proposed a structure of free TAR RNA in solution which is similar to that proposed by Weeks and Crothers (1991). However, in the presence of free argininamide or a peptide containing a single arginine, the conformation of TAR RNA is drastically changed to adopt a structure where U23 forms a base triple with A27 and U38, forcing C24 and U25 out of the helix (Puglisi *et al.*, 1992).

In this report, we have established very clear physical evidence for major groove widening in TAR RNA in solution and in the presence of arginine. In previous studies, attempts to determine whether the reactive purines in TAR RNA are protected from diethyl pyrocarbonate reactions in the presence of a Tat peptide were unsuccessful (Weeks & Crothers, 1991). In a recent report, Hamy *et al.* (1993) carried out site-specific functional group modifications on TAR RNA and showed that Tat forms multiple specific hydrogen bonds to a series of dispersed sites displayed in the major groove. We have shown here that a Tat fragment (42−72) occupies the major groove of TAR RNA and abolishes the access of the rhodium complex.

A three-dimensional computer graphic representation is used to show cleavage sites and major groove widening in the bulge region of TAR RNA (Figure 4). The fundamental conclusion established by our studies is that the major groove widens significantly in a bulge of more than one nucleotide. In virtually all lentiviral Tat, Rev, or Rex proteins and bacteriophage N proteins, an arginine- and lysine-rich subdomain is involved in RNA recognition (Lazinski *et al.*, 1989; Mattaj, 1993). There is an intriguing possibility that these proteins recognize their target RNAs in a similar fashion. Bulge-induced widening of major grooves may

provide potential recognition sites for basic subdomain-containing proteins.

ACKNOWLEDGMENT

We thank Kavita Shah for the synthesis of $Rh(phen)_2phi^{3+}$; Anna Marie Pyle for advice for synthesis of the rhodium complex; Weirong Wang for preliminary experiments; Gabriele Varani for coordinates for the ribbon model, discussions, and critical reading of the manuscript; and Marc Gartenberg for helpful discussions.

REFERENCES

Arnott, S., Hukins, D. W. L., & Dover, S. D. (1972) *Biochem. Biophys. Res. Commun. 48*, 1392−1399.
Baudin, F., & Romaniuk, P. J. (1989) *Nucleic Acids Res. 17*, 2043−2056.
Berkhout, B., Silverman, R. H., & Jeang, K. T. (1989) *Cell 59*, 273−282.
Bhattacharyya, A., Murchie, A. I. H., & Lilley, D. M. J. (1990) *Nature 343*, 484−487.
Chow, C. S., & Barton, J. K. (1990) *J. Am. Chem. Soc. 112*, 2839−2841.
Chow, C. S., Behlen, L. S., Uhlenbeck, O. C., & Barton, J. K. (1992a) *Biochemistry 31*, 972−982.
Chow, C. S., Hartmann, K. M., Rawlings, S. L., Huber, P. W., & Barton, J. K. (1992b) *Biochemistry 31*, 3534−3542.
Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. C., Gait, M. J., & Karn, J. (1993) *J. Mol. Biol. 230*, 90−110.
Cordingley, M. G., La Femina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., Le Grow, K., Gotlib, L., Schlabach, A. J., & Colonno, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8985−8989.
Cullen, B. R. (1986) *Cell 46*, 973−982.
Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C., & Haseltine, W. A. (1986) *Cell 44*, 941−947.
Delling, U., Reid, L. S., Barnett, R. W., Ma, M. Y.-X., Climie, S., Summer-Smith, M., & Sonenberg, N. (1992) *J. Virol. 66*, 3018−3025.
Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., & Valerio, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6925−6929.
Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., & Skinner, M. A. (1990) *EMBO J. 9*, 4145−4153.
Feng, S., & Holland, E. C. (1988) *Nature 334*, 165−168.
Fisher, A. G., Feinberg, M. B., Josephs, S. F., Harper, M. E., Marselle, L. M., Reyes, G., Gonda, M. A., Aldovini, A., Debouck, C., Gallo, R. C., & Wong-Staal, F. (1986) *Nature 320*, 367−371.
Harper, J. W., & Logsdon, N. J. (1991) *Biochemistry 30*, 8060−8066.
Hauber, J., & Cullen, B. R. (1988) *J. Virol. 62*, 673−679.
Howe, P. W. A., Nagai, K., Neuhaus, D., & Varani, G. (1994) *EMBO J. 13*, 3873−3881.
Jaeger, J. A., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 12522−12530.
Jakobovits, A., Smith, D. H., Jakobovits, E. B., & Capon, D. J. (1988) *Mol. Cell. Biol. 8*, 2555−2561.
Jessen, T.-H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.
Kao, S.-Y., Calman, A. F., Luciw, P. A., & Peterlin, B. M. (1987) *Nature 330*, 489−493.
King, D. S., Fields, C. G., & Fields, G. B. (1990) *Int. J. Pept. Protein Res. 36*, 255−266.
Kuchino, Y., & Nishimura, S. (1989) *Methods Enzymol. 180*, 154−163.
Laspia, M. F., Rice, A. P., & Mathews, M. B. (1989) *Cell 59*, 283−292.
Lazinski, D., Grzadzielska, E., & Das, A. (1989) *Cell 59*, 207−218.
Loret, E. P., Georgel, P., Johnson, W. C., & Ho, S. P. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9734−9738.
Marciniak, R. A., Garcia-Blanco, M. A., & Sharp, P. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3624−3628.
Mattaj, I. W. (1993) *Cell 73*, 837−840.
Mazumder, A., Gaynor, R. B., & Sigman, D. S. (1992) *Biochem. Biophys. Res. Commun. 187*, 1503−1509.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
Muesing, M. A., Smith, D. H., & Capon, D. A. (1987) *Cell 48*, 691−701.
Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., & Schimmel, P. (1991) *Science 253*, 784−786.
Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., & Nagai, K. (1994) *Nature 372*, 432−438.
Pearson, L., Chen, C.-h. B., Gaynor, R., & Sigman, D. S. (1994) *Nucleic Acids Res. 22*, 2255−2263.
Peterlin, B. M., Luciw, P. A., Barr, P. J., & Walker, M. D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9734−9738.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76−80.
Pyle, A. M., Long, E. C., & Barton, J. K. (1989) *J. Am. Chem. Soc. 111*, 4520−4522.
Pyle, A. M., Chiang, M. Y., & Barton, J. K. (1990) *Inorg. Chem. 29*, 4487−4495.
Rice, A. P., & Mathews, M. B. (1988) *Nature 332*, 551−553.
Riordan, F. A., Bhattacharyya, A., McAteer, S., & Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305−310.
Rosen, C. A., Sodroski, J. G., & Haseltine, W. A. (1985) *Cell 41*, 813−823.
Rould, M. A., Perona, J. J., Soll, D., & Steitz, T. A. (1989) *Science 246*, 1135−1142.
Ruff, M., Krisnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., & Moras, D. (1991) *Science 252*, 1682−1689.
Thierry, J. C., & Moras, D. (1991) *Science 252*, 1682−1689.
Sharp, P. A., & Marciniak, R. A. (1989) *Cell 59*, 229−230.
Sheline, C. T., Milocco, L. H., & Jones, K. A. (1991) *Genes Dev. 5*, 2508−20.
Sitlani, A., Long, E. C., Pyle, A. M., & Barton, J. K. (1992) *J. Am. Chem. Soc. 114*, 2303−2312.
Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., & Sonenberg, N. (1991) *J. Virol. 65*, 5196−5202.
Tan, R., & Frankel, A. D. (1992) *Biochemistry* 31, 10288−10294.
Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232−5237.
Tao, J., & Frankel, A. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2723.
Tinoco, I., Jr., Puglisi, J. D., & Wyatt, J. R. (1990) *Nucleic Acids Res. Mol. Biol. 4*, 205−226.
Weeks, K. M., & Crothers, D. M. (1991) *Cell 66*, 577−588.
Wu, H.-N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221−8227.
Wyatt, J. R., Puglisi, J. D., & Tinoco, I., Jr. (1989) *Bioassays 11*, 100−106.

BI942620G