

Visualizing a Specific Contact in the HIV-1 Tat Protein Fragment and *trans*-Activation Responsive Region RNA Complex by Photocross-linking*

(Received for publication, October 27, 1995, and in revised form, January 17, 1996)

Yiping Liu, Zhuying Wang‡, and Tariq M. Rana§

From the Department of Pharmacology, Robert Wood Johnson Medical School-University of Medicine and Dentistry of New Jersey, and the ‡Biochemistry and Molecular Biology Graduate Program at Rutgers State University, Piscataway, New Jersey 08854

Replication of human immunodeficiency virus type 1 (HIV-1) requires specific interactions of Tat protein with the *trans*-activation responsive region (TAR) RNA, a stem-loop structure containing two helical stem regions separated by a trinucleotide bulge. The Tat protein contains a basic RNA-binding region (amino acids 49–57) located in the carboxyl-terminal half of the protein, and peptides containing this basic domain of Tat protein can bind TAR RNA with high affinities. We synthesized a 31-amino acid Tat fragment (amino acids 42–72) containing the basic region and part of flanking regulatory core domain that formed a specific complex with TAR RNA. Upon UV irradiation (254 nm), this Tat fragment cross-linked covalently with TAR RNA. Sites of cross-links were determined on both the TAR RNA and Tat protein fragment by RNA and protein sequencing, respectively. These results revealed that guanosine 26 of TAR RNA was cross-linked with tyrosine 47 of the Tat peptide. Our results provide the first physical evidence for a direct amino acid-base contact in Tat-TAR complex. Recently, orientation of the Tat-(42–72) was determined in our laboratory by psoralen-Tat-(42–72) conjugate (Wang, Z., and Rana, T. M. (1995) *J. Am. Chem. Soc.* 117, 5438–5444). On the basis of our findings, we suggest a model in which Tat binds to TAR RNA by inserting the basic recognition sequence into the major groove with an orientation where lysine 41 in the core domain of Tat contacts the lower stem and Tyr⁴⁷ is close to G²⁶ of TAR RNA. The knowledge of the orientation of Tat and details of other interactions with TAR RNA in Tat-TAR complex has significant implications for understanding gene regulation in HIV-1.

The role of RNA-protein interactions is vital for many regulatory processes, especially in gene regulation where proteins specifically interact with binding sites found within RNA transcripts. RNA molecules can fold into extensive structures containing regions of double-stranded duplex, hairpins, internal loops, bulged bases, and pseudoknotted structures (1, 2). Due to the complexity of RNA structure, the rules governing sequence-

specific RNA-protein recognition are not well understood. Recent 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 (3, 4). Major groove recognition takes place between aspartyl-tRNA synthetase and its cognate tRNAs at a site of local distortion in the RNA helix (5). 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* (6) and the coat protein of phage R17 (7). In a recent report, interactions between U1 small nuclear RNA and the N-terminal domain of the human U1A protein were mapped by multidimensional heteronuclear NMR studies (8). 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 (8). 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 (9). On the basis of NMR data, it has been shown that TAR¹ RNA in HIV-1 changes its conformation upon arginine binding (10, 11). All of these studies suggest that the diversity of RNA structures plays a central role in their specific recognition by proteins.

The promoter of the human immunodeficiency virus type 1 (HIV-1), located in the U3 region of the viral long terminal repeat, is an inducible promoter that can be stimulated by the *trans*-activator protein, Tat (12). As in other lentiviruses, Tat protein is essential for transactivation of viral gene expression (13–16). In the absence of Tat, most of the viral transcripts terminate prematurely, producing short RNA molecules ranging in size from 60 to 80 nucleotides. Jeang *et al.* (17) reported that integrated HIV-1 promoters did not show a high rate of abortive transcription. Nonetheless, HIV-1 proviruses and integrated long terminal repeats respond efficiently to Tat (17). The Tat protein is a small, cysteine-rich nuclear protein containing 86 amino acids and comprised of 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 (18–22). Upon binding to the TAR RNA sequence, Tat causes a substantial increase in transcript levels (23–27). The increased efficiency in transcription may result from preventing premature termination of the transcriptional elongation

* This work was supported in part by National Institutes of Health Grant AI 34785 and by the State of New Jersey Commission on Cancer Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed. Dept. of Pharmacology, Robert Wood Johnson Medical School-UMDNJ and Biochemistry and Molecular Biology Graduate Program, Rutgers State University, 675 Hoes Lane, Piscataway, NJ 08854. Tel.: 908-235-4082; Fax: 908-235-4073; E-mail: rana@mbcl.rutgers.edu.

¹ The abbreviations used are: TAR, *trans*-activation responsive region; HIV-1, human immunodeficiency virus, type 1.

complex (28) or from enhancing initiation of transcription (29). TAR RNA was originally localized to nucleotides +1 to +80 within the viral long terminal repeat (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* (21, 30, 31). As shown in Fig. 5, the TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge, which separates two helical stem regions (18, 21, 22, 25). The trinucleotide bulge is essential for high affinity and specific binding of the Tat protein (32, 33).

The Tat protein contains a basic RNA-binding region (amino acids 49–57) located in the carboxyl-terminal half of the protein (19, 34–37). Peptides containing the basic domain (residues 49–57) of Tat protein can bind TAR RNA with high affinities (36, 38–44). We used a 31-amino acid Tat fragment (amino acids 42–72) to form a specific complex with TAR RNA. Upon UV irradiation, this Tat fragment formed a covalent cross-link with TAR RNA. Sites of cross-links were determined on both the TAR RNA and Tat protein fragment by RNA and protein sequencing, respectively, which revealed that Tyr⁴⁷ of Tat is close to G²⁶ of TAR RNA. Our results provide the first physical evidence for a direct amino acid-base contact in Tat-TAR complex.

EXPERIMENTAL PROCEDURES

Buffers

All buffer pH values refer to measurements at room temperature. Transcription buffer contained 40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 0.01% Triton X-100, 5 mM dithiothreitol. TBE buffer contained 50 mM Tris, 50 mM boric acid (pH 8.3), 1 mM EDTA. Sample loading buffer contained 9 M urea, 50 mM Tris, 50 mM boric acid (pH 8.3), 1 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol.

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. The top strand is a short piece of DNA complementary to the 3'-end of all template DNAs having the sequence 5'-TAATACGACTCACTATAG-3'. DNA was deprotected in NH₄OH 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 8 M urea-20% acrylamide denaturing gels, 450 × 0.8 mm. Gels were run for 4 h at 30 W until bromophenol blue 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, 50 mM boric acid, 1 mM EDTA, and 0.5 M sodium acetate. DNAs were ethanol-precipitated and resuspended in diethyl pyrocarbonate-treated water. Concentration of DNAs was determined by measuring absorbance at 260 nm in a Shimadzu UV spectrophotometer. Samples were stored at -20 °C.

RNAs—RNAs were prepared *in vitro* by transcription from synthetic DNA templates by T7 polymerase (45). 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 reactions containing 8.0 pmol of template DNA, 40–60 units of T7 polymerase was used. Reactions were stopped by adding an equal volume of sample loading buffer. RNAs were purified by electrophoresis on an 8 M urea, 20% polyacrylamide gel as described above. The sequence of RNAs was determined by base hydrolysis and nuclease digestion. RNA, starting with 300 pmol, was 5'-end-labeled with ³²P to a specific activity of 2000 × 10⁴ cpm/pmol after dephosphorylation with calf intestinal alkaline phosphatase (Promega), as described (46). The labeled RNAs were purified by electrophoresis on an 8 M urea, 20% polyacrylamide gel or by a Sep-Pak C₁₈ cartridge (Waters, Millipore Corp.).

Peptide Synthesis

All Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) amino acids, piperidine, 4-dimethylaminopyridine, dichloromethane, *N,N*-dimethylformamide, 1-hydroxybenzotriazole, 2-(1H-benzotriazo-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, diisopropylethylamine, and HMP-linked polystyrene resin were obtained from the Applied Biosystems Division, Perkin-Elmer. Trifluoroacetic acid, 1,2-ethanedithiol, phenol, and

thioanisole were from Sigma. Two Tat-derived peptides (from amino acids 42–72 and 48–72) were synthesized on an Applied Biosystems 431A peptide synthesizer using standard Fast Moc protocols. Cleavage and deprotection of the peptide was carried out in 2 ml of Reagent K for 6 h at room temperature. Reagent K contained 1.75 ml of trifluoroacetic acid, 100 μl of thioanisole, 100 μl of water, and 50 μl of ethanedithiol (47). After cleavage from the resin, the peptide was purified by high pressure liquid chromatography on a Zorbax 300 SB-C₈ column. The mass of fully deprotected and purified peptide was confirmed by fast atom bombardment mass spectrometry: calculated mass for Tat-(42–72) C₁₅₂H₂₆₃N₅₉O₄₃ = 3605.5, found 3606.5 (M + H); calculated mass for Tat-(48–72) C₁₂₃H₂₁₉N₅₃O₃₅ = 3000.4, found 3001.4 (M + H).

UV Irradiation of RNA-Peptide Complex

A typical cross-linking reaction mixture (15 μl) contained 0.25 μM labeled RNA, 1.9 μM Tat peptide, 80 μg/ml bovine serum albumin, 25 mM Tris (pH 7.4), and 100 mM NaCl. RNAs in H₂O were heated at 75 °C for 2 min and slowly cooled to room temperature. The reaction mixtures were incubated at 25 °C for 30 min and cooled on ice before UV irradiation. The UV irradiation was conducted on ice for 10 min (254 nm; 1280 ergs/mm²/s) in a Rayonet RPR 100 photochemical reactor (Southern New England Ultraviolet Inc.). After UV irradiation, 2 μl of 20 mg/ml yeast tRNA (to remove Tat peptide from non-cross-linked RNA-peptide complex) and 10 μl of sample loading buffer were added. The mixtures were heated at 75 °C for 1 min before being loaded on 8 M urea, 20% polyacrylamide gel. After electrophoresis, the gels were directly autoradiographed. Efficiencies of cross-linking were determined by a PhosphorImager analysis (Molecular Dynamics).

Base Hydrolysis of RNA Cross-linked with Peptide

5'-end-labeled cross-linked RNA-peptide complexes were prepared as described above. After electrophoresis, the band corresponding to the cross-linked RNA-peptide complex, identified by autoradiography, was excised and eluted into the TBE buffer at 4 °C overnight. The samples were then ethanol-precipitated and dried in a Savant Speedvac. Approximately 2 pmol of 5'-end-labeled RNA-peptide complex was partially hydrolyzed in the presence of 50 mM sodium carbonate (pH 9.2) for 20 min at 85 °C. Size was assigned to the individual band by comparison with the migration of RNA digestion products produced by RNase T1 and *Bacillus cereus* (Pharmacia Biotech Inc.).

Purification of Cross-linked Complex, Digestion, and Peptide Sequencing

Preparation of Cross-linked Complex on a Preparative Scale—The reaction mixture (10 ml) contained 4 μM 5'-end-labeled RNA, 6 μM peptide, 25 mM Tris (pH 7.4), and 50 mM NaCl. After UV irradiation, the mixtures were applied to a Mono Q HR 5/5 column (Pharmacia) equilibrated with 40 mM KCl, 25 mM Tris (pH 8.3) and eluted from the column with a salt gradient (0–8 min, 40–440 mM KCl; 8–16 min, 440–640 mM KCl; and 16–24 min, 640–1040 mM KCl) at a flow rate of 1 ml/min. RNA-peptide complex was eluted at approximately 480 mM KCl and free RNA at 550 mM KCl. The fractions containing RNA-peptide complex were identified by gel electrophoresis (SDS and urea gels) and autoradiography. The RNA-peptide complex was desalted on a Sep-Pak C₁₈ cartridge and lyophilized.

Digestion and Peptide Sequencing—The 5'-end-labeled RNA-peptide complex was dissolved in 0.15 ml of H₂O, 80 μl of 0.5 M Tris (pH 8.1), and 80 μl of 1% SDS. After heating at 70 °C for 1 min and cooling to 37 °C, 0.62 ml of H₂O, 0.067 ml of 15 mM CaCl₂, and then 0.25 μg of sequencing grade modified trypsin (Promega) were added. The samples were incubated at 37 °C for 13 h. After phenol extraction and ethanol precipitation, the samples were dissolved in 70 μl of H₂O and 70 μl of sample loading buffer and electrophoresed on 8 M urea, 20% polyacrylamide. The tryptic fragments were identified by autoradiography, excised from the gel, eluted into 2 ml of 10 mM ammonium carbonate (pH 8.1), washed exhaustively with H₂O on Sep-Pak C₁₈ cartridge, and dried in a Savant Speedvac. The concentration of cross-linked RNA-peptide complex was determined based on RNA absorption at 260 nm. Amino acid sequencing was conducted at Yale University (New Haven, CT) on an Applied Biosystems sequencer.

RESULTS

Formation of Cross-linked RNA-Peptide Complex—Before performing the UV-induced cross-linking experiment, it was necessary to establish the binding affinities of the Tat fragment (residues 42–72) used in this study. The dissociation constant

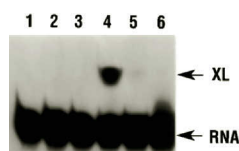


FIG. 1. Separation of covalently cross-linked Tat-(42-72)-TAR complex from TAR RNA by denaturing 8 M urea-20% acrylamide gel. Lanes 1 and 2, RNA without (lane 1) and with (lane 2) UV irradiation; lanes 3 and 4, RNA with Tat peptide without (lane 3) and with (lane 4) UV irradiation; lane 5, RNA and Tat peptide irradiated separately and then combined; lane 6, treatment of UV-irradiated products (lane 4) with Proteinase K. Reaction mixtures contained 0.25 μM of 5'- ^{32}P -end-labeled TAR RNA, 1.9 μM Tat-(42-72), and 100 mM NaCl in 25 mM Tris-HCl buffer, pH 7.4. XL, cross-linked RNA-peptide complex.

(K_d) of TAR RNA and the Tat peptide was estimated to be 0.13 μM (48). The concentrations of TAR RNA and the Tat peptide employed for cross-linking reactions were 0.25 μM and 1.9 μM , respectively. Under these conditions, electrophoretic mobility shift assays revealed (data not shown) only one slow migrating RNA-peptide complex, indicating the absence of other nonspecific RNA-peptide complex formation (44, 48). TAR RNA labeled at the 5'-end with ^{32}P was incubated with the Tat peptide for 20 min in 25 mM Tris (pH 7.4) and 100 mM NaCl and UV-irradiated with 254-nm light (see "Experimental Procedures"). Products of the photoreaction were analyzed by denaturing 8 M urea-polyacrylamide gel electrophoresis (Fig. 1). Irradiation of the RNA-peptide complex yields a new band with electrophoretic mobility less than that of TAR RNA (lane 4). Both the peptide and UV irradiation are required for the formation of a cross-linked RNA-protein complex. This is evidenced by the fact that no cross-linked products are observed when RNA is irradiated in the absence of Tat peptide (lane 2) or incubated with peptide in the dark without UV irradiation (lane 3). Further control experiments showed that no cross-linking was observed when RNA and peptide were irradiated separately and then mixed (lane 5). Digestion of the RNA-peptide cross-link with Proteinase K (5 units for 30 min at 37 $^{\circ}\text{C}$) resulted in an RNA species with mobility similar to TAR RNA (lane 6). The products of irradiation were also analyzed on a denaturing SDS-15% polyacrylamide gel. Again, a photoproduct with electrophoretic mobility less than that of TAR RNA was observed that was dependent on the presence of RNA and peptide (data not shown). The photoproduct yield is $\sim 5\%$ as determined by a PhosphorImager analysis. Since the cross-linked RNA-peptide complex is stable to alkaline pH (9.5), high temperature (85 $^{\circ}\text{C}$) and denaturing conditions (8 M urea, 2% SDS) we conclude that a covalent bond is formed between TAR RNA and the peptide during the cross-linking reaction.

Dependence of the Cross-linking Reaction on the Concentration of Peptide and Time of Irradiation—Formation of RNA-peptide photocross-link was dependent on the concentration of Tat peptide. Here, as in Fig. 1, the major cross-link product (XL1) has slightly lower electrophoretic mobility than TAR RNA. In Fig. 2, the efficiency of XL1 formation was increased as the peptide concentration was raised from 0.13 μM to 1.25 μM . At a peptide concentration higher than 1.25 μM , a second minor cross-linked product (XL2) with a lower electrophoretic mobility than that of XL1 was observed.

The photocross-linking reaction between the Tat peptide and TAR RNA was also dependent on time of irradiation. The yields of cross-linked RNA-peptide complex were increased with an increase in time of irradiation (Fig. 3). In this experiment, similar to that shown in Fig. 2, extended time of irradiation also resulted in the formation of XL2 at 30 and 40 min. This second minor photoproduct could be the result of nonspecific binding of the peptide to RNA (at higher concentrations of peptide) or nonspecific association of photodamaged RNA and

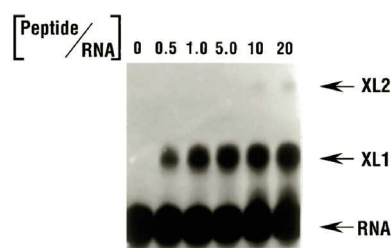


FIG. 2. Effect of Tat peptide concentration on UV-induced RNA-peptide cross-link formation. ^{32}P -5'-end-labeled TAR RNA (0.25 μM) was incubated with Tat peptide at concentrations of 0.13, 0.25, 1.25, 2.5, and 5.0 μM . The concentration ratios of peptide to RNA are shown in the figure. For details of reaction conditions see "Experimental Procedures." XL1, cross-linked RNA-peptide complex, major photoproduct; XL2, cross-linked RNA-peptide complex, minor photoproduct.



FIG. 3. Time course of cross-linking reaction of Tat-(42-72)-TAR complex. The reaction mixture contained 0.25 μM of ^{32}P -5'-end-labeled TAR RNA and 1.9 μM Tat peptide and was UV irradiated as described under "Experimental Procedures." After the indicated irradiation times, aliquots were withdrawn and analyzed on 8 M urea-polyacrylamide gel. XL1, cross-linked RNA-peptide complex, major photoproduct; XL2, cross-linked RNA-peptide complex, minor photoproduct.

peptide after longer irradiation times. Further characterization of this minor photoproduct was not carried out in this study.

Specificity of the Cross-link Formation—Specificity of the cross-linking reaction was established by competition experiments. Cross-linking reactions were performed in a 15- μl volume containing 0.25 μM of ^{32}P -5'-end-labeled TAR RNA, 1.9 μM Tat peptide, 80 $\mu\text{g/ml}$ bovine serum albumin, 25 mM Tris-HCl (pH 7.4), 100 mM NaCl, and up to 10 μM unlabeled competitor RNA. Cross-linked products were separated by 8 M urea-20% acrylamide gels and quantitated by PhosphorImager analysis. Fig. 4 shows that cross-linking was inhibited by the addition of unlabeled wild-type TAR RNA and not by a mutant TAR RNA lacking the trinucleotide bulge. Additional control experiments showed that cross-linking did not occur between a mutant TAR RNA without trinucleotide bulge and Tat-(42-72) (data not shown). Therefore, we conclude that formation of a specific RNA-protein complex between TAR RNA and Tat peptide is necessary for photocross-linking.

Guanosine 26 in TAR RNA Cross-links to Tat—Mapping of the cross-link site on TAR RNA to single nucleotide resolution was carried out by partial alkaline digestion of gel-purified RNA-protein cross-link XL1. Fragment sizes are determined by comparison with RNA oligonucleotides of defined sequence and length generated by digesting RNA with RNases T_1 and $B.$ *Cereus*. Base hydrolysis of RNA and cross-linked RNA-peptide complex generates a ladder of RNA degradation products. Bands of cross-linked RNA-peptide complexes migrate more slowly than corresponding free RNA (Fig. 5A, lanes 2 and 3). Base hydrolysis of the 5'-end-labeled cross-linked complex (Fig. 5A, lane 3) results in an RNA ladder in which all fragments up to uridine 25 are resolved. There is an obvious gap in the hydrolysis ladder after U^{25} , indicating that the fragments above uridine 25 from the 5'-end are linked to the Tat peptide (Fig. 5A, lane 3). A standard base hydrolysis ladder was observed for 5'-end-labeled TAR RNA showing sensitivity to base hydrolysis at all positions, including U^{25} in the sequence (Fig.

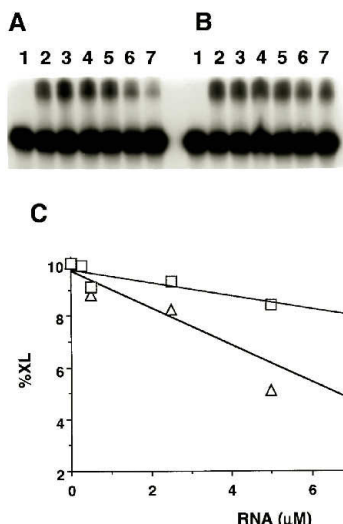


FIG. 4. **Specificity of the cross-linking reaction determined by competition assays.** Complexes were formed between $0.25 \mu\text{M}$ ^{32}P -labeled TAR RNA and $1.9 \mu\text{M}$ of Tat-(42–72) in the presence of unlabeled wild-type TAR RNA (A) or bulgeless mutant TAR RNA (B). Concentrations of the competitor RNA in lanes 2, 3, 4, 5, 6, and 7 were 0, 0.25, 0.5, 2.5, 5, and $10 \mu\text{M}$, respectively. Lane 1 was a control RNA-peptide complex without UV irradiation. C, quantitative analysis of competition experiments. The fraction of RNA in RNA-peptide cross-link was determined by PhosphorImager analysis as described under “Experimental Procedures.” Δ , wild-type TAR RNA competitor; \square , bulgeless mutant TAR RNA.

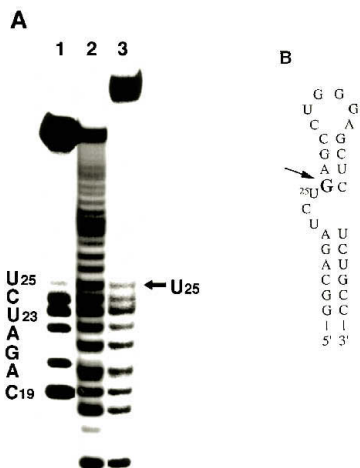


FIG. 5. **Mapping of cross-linked base in the RNA-protein cross-links by alkaline hydrolysis.** A, analysis of 5'-end-labeled TAR RNA and cross-link: *B. cereus* ladder of TAR RNA (lane 1); hydrolysis ladder of TAR RNA (lane 2); hydrolysis ladder of cross-linked RNA-peptide complex (lane 3). The sequence of TAR RNA from C¹⁹ to U²⁵ is labeled. A gap in the sequence is obvious after the U²⁵ residue, indicating that G²⁶ is the cross-linked base. B, sequence and secondary structure of wild-type TAR RNA used in this study. TAR RNA spans the minimal sequences that are required for Tat responsiveness *in vivo* (21) and for *in vitro* binding of Tat-derived peptides (38). Wild-type TAR contains two non-wild-type base pairs to increase transcription by T7 RNA polymerase. U²⁵ represents the nucleotide at which the hydrolysis of the 5'-end-labeled cross-linked RNA-peptide complex was stopped. The arrow indicates the location of guanosine 26, which is the cross-linked base in TAR RNA (shown in boldface).

5A, lane 2). Thus, we conclude that guanosine 26 of TAR RNA is the site at which cross-linking occurs (Fig. 5B).

Cross-linking Occurs at Tyr⁴⁷ of Tat—To identify the amino acid(s) of Tat that are involved in specific cross-linking with TAR RNA, the cross-linked RNA-peptide complex (XL1) was prepared on a preparative scale (see “Experimental Procedures”), purified from noncross-linked TAR RNA by ion ex-

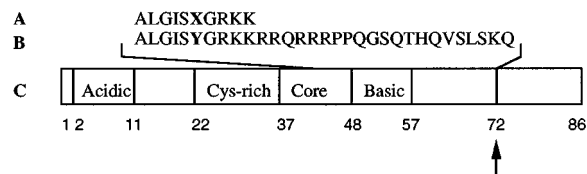


FIG. 6. **Identification of the amino acid in the Tat-(42–72) sequence that cross-links to TAR RNA.** A, amino acid sequence of the cross-linked peptide. A nonstandard amino acid (X) was identified during the 6th cycle of N-terminal sequencing. X in the sequence indicates the cross-linking site that corresponds to Y in B. B, amino acid sequence of Tat-(42–72) peptide used in the cross-linking experiments with TAR RNA. C, schematic representation of the functional domains of HIV-1 Tat protein. Various numbers refer to the amino acid positions, and the arrow represents the separation of the region contributed by adjacent exon (62).

change fast protein liquid chromatography, and digested with trypsin. The tryptic digest products were purified by 8 M urea, 20% acrylamide denaturing gels and visualized by autoradiography. We recovered ~ 100 pmol quantities of a tryptic fragment of XL1 and subjected it to N-terminal sequencing. The amino acid sequencing data showed that it had a sequence of Ala-Leu-Gly-Ile-Ser-X-Gly-Arg-Lys-Lys. This sequence corresponds to the sequence encompassing amino acids 42–51 in HIV-1 Tat protein (Fig. 6). X at the 6th position represents a nonstandard amino acid instead of tyrosine 47 of the Tat peptide. Thus, cross-linking occurs at tyrosine 47 of the Tat peptide.

DISCUSSION

Ultraviolet-induced cross-linking of RNA to proteins is a widely used technique to study *in vitro* and *in vivo* RNA-protein interactions (49–51). UV irradiation with sufficient intensity generates a highly reactive species of RNA, which reacts with protein and organic molecules involved in making direct contacts with RNA (52, 53). To identify specific RNA-protein contacts, we irradiated TAR RNA and Tat-(42–72) protein complex with UV light and observed the formation of a covalent bond between RNA and protein. Formation of this covalently cross-linked product was dependent on the concentration of Tat peptide and irradiation time (Figs. 2 and 3). Our competition and control experiments showed that a specific RNA-protein complex formation between TAR RNA and Tat fragment was necessary for photo-crosslinking reactions (Fig. 4).

To locate the cross-link sites in TAR RNA and the Tat peptide, we prepared the RNA-protein cross-link on a preparative scale, purified the cross-link, and analyzed it by RNA and protein sequencing. Alkaline hydrolysis of 5'-end-labeled cross-links indicated that a single nucleotide, G²⁶, in TAR RNA was involved in covalent interaction (Fig. 5, A and B). The absence of bands in the hydrolysis ladder after U²⁵ from the 5'-end of RNA indicates that the RNA fragments after U²⁵ are covalently linked to the Tat peptide and migrate more slowly to create a gap in the standard hydrolysis ladder. Our results clearly demonstrate that cross-linking occurs at G²⁶ in TAR RNA.

Peptide sequencing on a tryptic fragment of the cross-link complex was accomplished by Edman degradation chemistry. The sequencing data indicate that cross-linking occurred at Tyr⁴⁷ of the Tat peptide. As shown in Fig. 6, peptide sequencing identified a nonstandard amino acid X at the 6th position of the cross-linked peptide, Ala-Leu-Gly-Ile-Ser-X-Gly-Arg-Lys-Lys. This sequence corresponds to the region encompassing amino acids 42–51 in HIV-1 Tat protein (Fig. 6). The nonstandard amino acid most likely corresponds to a photomodified tyrosine. Sequencing of proteins by Edman degradation chemistry requires unmodified amino and carbonyl groups in the backbone of the peptide. Evidence that the Edman sequencing reaction

was able to continue through Tyr⁴⁷ indicates that cross-linking does not occur at these locations (54). Therefore, we conclude that the aromatic side chain or C- α atom in the peptide backbone of Tyr⁴⁷ is involved in the covalent cross-link formation with TAR RNA.

It has been shown by a number of groups that Tat-derived peptides that contain the basic arginine-rich region of Tat are able to form *in vitro* complexes with TAR RNA (36, 38–44). Recently, Churcher *et al.* (44) published a detailed comparative study arguing that Tat peptides can mimic the binding affinity and specificity of Tat protein. Results from that study showed that the addition of amino acid residues from the core region of the Tat protein to the arginine-rich domain-containing peptides increased binding specificities (44). To achieve specific RNA binding by a Tat fragment, we used a Tat peptide, Tat-(42–72), that contained an RNA-binding domain and six amino acids from the core domain of the Tat protein. In this report, our cross-linking results have established that this Tat-(42–72) peptide forms a specific covalent photocross-link to TAR RNA where Tyr⁴⁷ of the peptide contacts G²⁶ of the RNA.

What is the biological relevance of these findings? A number of studies showed that the immediate stem nucleotide base pairs flanking the bulge region of TAR RNA are required for Tat binding and *trans*-activation (44, 55, 56). During a detailed mutational analysis of TAR RNA, it was reported that a change of the G²⁶-C³⁹ base pair to C²⁶-G³⁹ base pair resulted in only 12% *trans*-activation by HIV-1 Tat (56). These reports strongly support our finding that G²⁶ is directly involved in sequence-specific recognition and *trans*-activation by HIV-1 Tat protein. However, Tat protein mutants where Tyr⁴⁷ was substituted with Ala or His were functional for *trans*-activation (57, 58). These data raise the possibility that Tyr⁴⁷ is not essential for RNA recognition and that the cross-link formation between Tyr⁴⁷ and G²⁶ could be the result of close proximity and favorable photochemistry. To address this question, we carried out cross-linking experiments with a Tat fragment lacking Tyr⁴⁷, Tat-(48–72), which binds TAR RNA with high affinities (38, 39, 44). UV irradiation of TAR RNA complexed with Tat-(48–72) did not yield any specific RNA-protein cross-link products (data not shown). These results support our model of Tat-TAR interactions where the basic recognition sequence of Tat is located in the major groove of TAR RNA, bringing Tyr⁴⁷ in close vicinity of G²⁶ (Fig. 7). The cross-link formation between G²⁶ and Tyr⁴⁷ is likely the result of close proximity, favorable orientation, and photoreactivity of tyrosine.

How does Tat interact with TAR RNA? Several lines of evidence suggest that Tat protein contacts TAR RNA in a widened major groove. In a recent study from our laboratory, we used a rhodium complex, bis(phenanthroline)(phenanthrenequinone diimine)-rhodium(III) (Rh(phen)₂phi³⁺), to probe the effect of bulge bases on the major groove width in TAR RNA (59). This metal complex does not bind double helical RNA or unstructured single-stranded regions of RNA. Instead, sites of tertiary interaction that are open in the major groove and accessible to stacking are targeted by the complex through photoactivated cleavage (60). 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 (59). 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 TAR RNA and in a one-base bulge TAR RNA. Our studies establish two important factors involved in Tat-TAR recognition. (i) There is a correlation between major groove opening and Tat binding. At least a

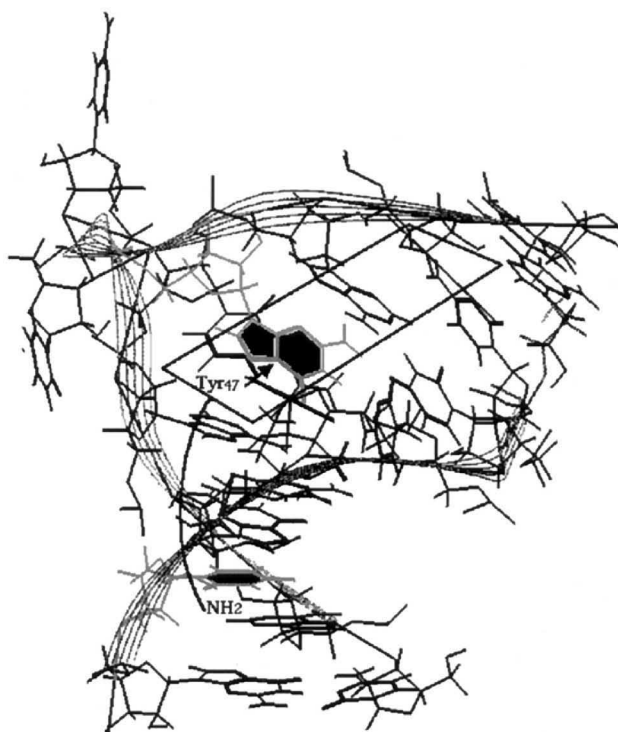


FIG. 7. A schematic illustration to show a three-dimensional model of the HIV-1 Tat binding site of TAR RNA and the location of the amino terminus of Tat-(42–72) and interactions of Tyr⁴⁷ of the peptide with G²⁶ of the RNA. Orientation of the Tat-(42–72) was determined in our laboratory by psoralen-Tat-(42–72) conjugate (48). The TAR RNA structure is based on NMR data (63). Ribbon structure of TAR RNA is shown in five dark lines. The basic region of Tat-(47–57) is represented as a barrel positioned in the wide major groove, and the N-terminal region containing Tat-(42–46) is drawn as a line. Tyrosine 47 is shown directly above the G²⁶ of TAR RNA (indicated in black) to demonstrate a close proximity between Tyr⁴⁷ and G²⁶. As determined by psoralen-Tat cross-linking experiments, the amino terminus of Tat-(42–72) contacts, or is close to, uridine 42 in the lower stem of TAR RNA (48); the amino terminus of the peptide is labeled as NH₂, and its proximal base, U⁴² of TAR RNA, is indicated in black. Structures of TAR RNA were visualized using Insight II software on an IRIS work station.

two-base bulge is required for major groove widening and other conformational changes to facilitate Tat binding. This cannot be accomplished by a single base bulge. (ii) The Tat fragment Tat-(42–72) occupies the major groove of TAR RNA and abolishes access of the rhodium complex. On the basis of chemical modification and gel mobility studies, a similar model was suggested earlier by Weeks and Crothers (55). Last, Hamy *et al.* (61) carried out site-specific modifications of functional groups on TAR RNA and showed that Tat forms multiple specific hydrogen bonds to a series of dispersed sites displayed in the major groove.

To determine the relative orientation of the nucleic acid and protein in the Tat-TAR complex, we have devised a new method based on psoralen photochemistry (48). We synthesized a 30-amino acid fragment containing the arginine-rich RNA-binding domain of Tat (residues 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 (48).

On the basis of the above studies, we suggest a model in which Tat binds to TAR RNA by inserting the basic recognition

sequence into the enlarged major groove with an orientation where lysine 41 in the core domain of Tat contacts the lower stem and Tyr⁴⁷ is close to G²⁶ of TAR RNA (Fig. 7). These findings are intriguing and suggest a possible mechanism of RNA recognition by Tat.

REFERENCES

1. Tinoco, I., Jr., Puglisi, J. D., and Wyatt, J. R. (1990) *Nucl. Acids Mol. Biol.* **4**, 205–226
2. Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989) *Bioessays* **11**, 100–106
3. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) *Science* **246**, 1135–1142
4. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R. Schimmel, P. (1991) *Science* **253**, 784–786
5. Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C., Moras, D., (1991) *Science* **252**, 1682–1689
6. Baudin, F., and Romaniuk, P. J. (1989) *Nucleic Acids Res.* **17**, 2043–2056
7. Wu, H.-N., and Uhlenbeck, O. C. (1987) *Biochemistry* **26**, 8221–8227
8. Howe, P. W. A., Nagai, K., Neuhaus, D., and Varani, G. (1994) *EMBO J.* **13**, 3873–3881
9. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994) *Nature* **372**, 432–438
10. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) *Science* **257**, 76–80
11. Puglisi, J. D., Chen, L., Frankel, A. D., and Williamson, J. R. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3680–3684
12. Jones, K. A., and Peterlin, B. M. (1994) *Annu. Rev. Biochem.* **63**, 717–43
13. Cullen, B. R. (1992) *Microbiol. Rev.* **56**, 375–394
14. Gaynor, R. (1992) *AIDS* **6**, 374–63
15. Fisher, A. G., Feinberg, M. B., Josephs, S. F., Harper, M. E., Marselle, L. M., Reyes, G., Gonda, M. A., Aldovini, A., Debouk, C., Gallo, R. C., Wong-Staal, F. (1986) *Nature* **320**, 367–371
16. Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C., and Haseltine, W. A. (1986) *Cell* **44**, 941–947
17. Jeang, K.-T., Berkhout, B., and Dropulic, B. (1993) *J. Biol. Chem.* **268**, 24940–24949
18. Rosen, C. A., Sodroski, J. G., and Haseltine, W. A. (1985) *Cell* **41**, 813–823
19. Hauber, J., and Cullen, B. R. (1988) *J. Virol.* **62**, 673–679
20. Feng, S., and Holland, E. C. (1988) *Nature* **334**, 165–168
21. Jakobovits, A., Smith, D. H., Jakobovits, E. B., and Capon, D. J. (1988) *Mol. Cell. Biol.* **8**, 2555–2561
22. Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989) *Cell* **59**, 273–282
23. Cullen, B. R. (1986) *Cell* **46**, 973–982
24. Peterlin, B. M., Luciw, P. A., Barr, P. J., and Walker, M. D. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 9734–9738
25. Muesing, M. A., Smith, D. H., and Capon, D. A. (1987) *Cell* **48**, 691–701
26. Rice, A. P., and Mathews, M. B. (1988) *Nature* **332**, 551–553
27. Laspias, M. F., Rice, A. P., and Mathews, M. B. (1989) *Cell* **59**, 283–292
28. Kao, S.-Y., Calman, A. F., Luciw, P. A., and Peterlin, B. M. (1987) *Nature* **330**, 489–493
29. Sharp, P. A., and Marciniak, R. A. (1989) *Cell* **59**, 229–230
30. Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989) *Genes & Dev.* **3**, 547–558
31. Garcia, J. A., Harrich, D., Soultanakis, E., Wu, F., Zmitsuyasu, R., and Gaynor, R. B. (1989) *EMBO J.* **8**, 765–778
32. Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., and Valerio, R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6925–6929
33. 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
34. Hauber, J., Malim, M. H., and Cullen, B. R. (1989) *J. Virol.* **63**, 1181–1187
35. Lazinski, D., Grzadziska, E., and Das, A. (1989) *Cell* **59**, 207–218
36. 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
37. Elangovan, B., Subramanian, T., and Chinnadurai, G. (1992) *J. Virol.* **66**, 2031–2036
38. 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., and Colonna, R. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8985–8989
39. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., and Crothers, D. M. (1990) *Science* **249**, 1281–1285
40. Calnan, B. J., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Genes & Dev.* **5**, 201–210
41. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Science* **252**, 1167–1171
42. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., and Sonenberg, N. (1991) *J. Virol.* **65**, 5196–5202
43. Weeks, K. M., and Crothers, D. M. (1993) *Science* **261**, 1574–1577
44. 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
45. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15**, 8783–8798
46. Kuchino, Y., and Nishimura, S. (1989) *Methods Enzymol.* **180**, 154–163
47. King, D. S., Fields, C. G., and Fields, G. B. (1990) *Int. J. Pept. Protein Res.* **36**, 255–266
48. Wang, Z., and Rana, T. M. (1995) *J. Am. Chem. Soc.* **117**, 5438–5444
49. Gott, J. M., Willis, M. C., Koch, T. H., and Uhlenbeck, O. C. (1991) *Biochemistry* **30**, 6290–6295
50. Piñol-Roma, S., Adam, S. A., Choi, Y. D., and Dreyfuss, G. (1989) *Methods Enzymol.* **180**, 410–418
51. Basilion, J. P., Rouault, T. A., Massinople, C. M., Klausner, R. D., and Burgess, W. H. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 574–578
52. Smith, K. C. (1976) in *Photochemistry and Photobiology of Nucleic Acids*, Vol. 2 (Wang, S. Y., ed) pp. 187–218, Academic Press, Inc., New York
53. Greenberg, J. R. (1979) *Nucleic Acids Res.* **6**, 715–731
54. Blatter, E. E., Ebright, Y. W., and Ebright, R. H. (1992) *Nature* **359**, 650–652
55. Weeks, K. M., and Crothers, D. M. (1991) *Cell* **66**, 577–588
56. Berkhout, B., and Jeang, K.-T. (1991) *Nucleic Acids Res.* **19**, 6169–6176
57. Kuppuswamy, M., Subramanian, T., Srinivasan, A., and Chinnadurai, G. (1989) *Nucleic Acids Res.* **17**, 3551–3561
58. Rice, A. P., and Carlotti, F. (1990) *J. Virol.* **64**, 1864–1868
59. Neenhold, H. R., and Rana, T. M. (1995) *Biochemistry* **34**, 6303–6309
60. Chow, C. S., Behlen, L. S., Uhlenbeck, O. C., and Barton, J. K. (1992) *Biochemistry* **31**, 972–982
61. 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
62. Vaishnav, Y. N., and Wong-Staal, F. (1991) *Annu. Rev. Biochem.* **60**, 577–630
63. Aboul-ela, F., Karn, J., and Varani, G. (1995) *J. Mol. Biol.* **253**, 313–332