Probing Tat Peptide−TAR RNA Interactions by Psoralen Photo-Cross-Linking[†]

Zhuying Wang, Kavita Shah, and Tariq M. Rana*

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

Received December 19, 2000; Revised Manuscript Received March 29, 2001

ABSTRACT: Replication of human immunodeficiency virus type 1 (HIV-1) requires specific interactions of Tat protein with the *trans*-activation responsive region (TAR) RNA, a 59-base stem−loop structure located at the 5′-end of all HIV mRNAs. We have used a site-specific cross-linking method based on psoralen photochemistry to determine the effect of core residues from the Tat sequence on the protein orientation in the Tat−TAR complex and on the specificity of Tat−TAR binding. We synthesized two Tat fragments, Tat(42−72) and Tat(37−72), and incorporated a psoralen-modified amino acid at position 41 during solid-phase assembly of the peptides. We used these psoralen−Tat conjugates to form specific complexes with TAR RNA. Upon near-ultraviolet irradiation (360 nm), psoralen−Asp41−Tat(37−72) 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 primer extension analysis, which revealed that Asp41 of Tat is close to U42 of the lower stem region of TAR RNA. Specificity of the RNA−peptide cross-linking reactions was determined by competition experiments. Our results show that the addition of only four residues (Cys37−Thr40) from the Tat core region significantly enhanced the specificity of the Tat peptide−TAR interactions without altering the site or chemical nature of the cross-link. These studies provide new insights into RNA−protein recognition that could be useful in designing peptidomimetics for RNA targeting. Such psoralen−peptide conjugates provide a new class of probes for sequence-specific protein−nucleic acid interactions and could be used to selectively control gene expression or to induce site-directed mutations.

The human immunodeficiency virus (HIV-1)[1] encodes a transcriptional activator protein, Tat, which is expressed early in the viral life cycle and is essential for viral gene expression, replication, and pathogenesis (for recent reviews, see refs *1*−*5*). Tat enhances the processivity of RNA Pol II elongation complexes that initiate in the HIV long terminal repeat (LTR) region. In nuclear extracts, HIV-1 Tat associates tightly with the CDK9-containing positive transcription elongation factor complex, P-TEFb (*6*−*8*). Recent studies indicate that Tat binds directly through its *trans*-activation domain to the cyclin subunit (CycT1) of the P-TEFb complex and induces loop sequence-specific binding of the P-TEFb complex to TAR RNA (*9*−*11*). Recruitment of P-TEFb to TAR has been proposed to be both necessary and sufficient for activation of transcription elongation from the HIV-1 long terminal repeat promoter (*12*). Neither CycT1 nor the P-TEFb complex binds TAR RNA in the absence of Tat, and thus the binding is highly cooperative for both Tat and P-TEFb (*9*, *11*). Tat appears to contact residues in the carboxy-terminal boundary of the CycT1 cyclin domain which are not critical for binding of cyclin T1 to CDK9 (*10*, *13*−*17*), and basic residues in CycT1 (R251, R254) further stabilize the Tat−P-TEFb−TAR RNA complex (*10*). Mutagenesis studies showed that the CycT1 sequence containing amino acids 1−303 was sufficient to form complexes with Tat−TAR and CDK9 (*10*, *13*−*17*). Recruitment of P-TEFb to TAR has been proposed to be both necessary and sufficient for activation of transcription elongation from the HIV-1 long terminal repeat promoter (*12*).

It is well established that Tat protein binds the TAR (*trans*-activation responsive) RNA element, a 59-base stem−loop structure located at the 5′-end of all nascent HIV-1 transcripts (*18*). TAR RNA was originally localized to nucleotides +1 to +80 within the viral long terminal repeat (LTR) (*19*). 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 (*20*−*22*). TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge which separates two helical stem regions (Figure 1) (*19*, *20*, *23*, *24*). The trinucleotide bulge is essential for high affinity and specific binding of the Tat protein (*25*, *26*). The loop region is required for in vivo *trans*-activation but is not involved in Tat binding (*26*−*30*).

RNA molecules can fold into extensive structures containing regions of double-stranded duplex, hairpins, internal loops, bulged bases, and pseudoknotted structures (*31*, *32*). Due to the complexity of the RNA structure, the rules governing sequence-specific RNA−protein recognition are not well understood. RNA−protein interactions are vital for

[†] This work was supported in part by National Institutes of Health Grant AI 41404 and by a Research Career Development Award to T.M.R.

* To whom correspondence should be addressed. Phone: (732) 235-4082. Fax: (732) 235-3235. E-mail: rana@umdnj.edu.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; TAR, *trans*-activation response element; LTR, long terminal repeat; P-TEFb, positive transcription elongation factor b; CycT1, human cyclin T1.

10.1021/bi0028744 CCC: $20.00
Published on Web 05/04/2001

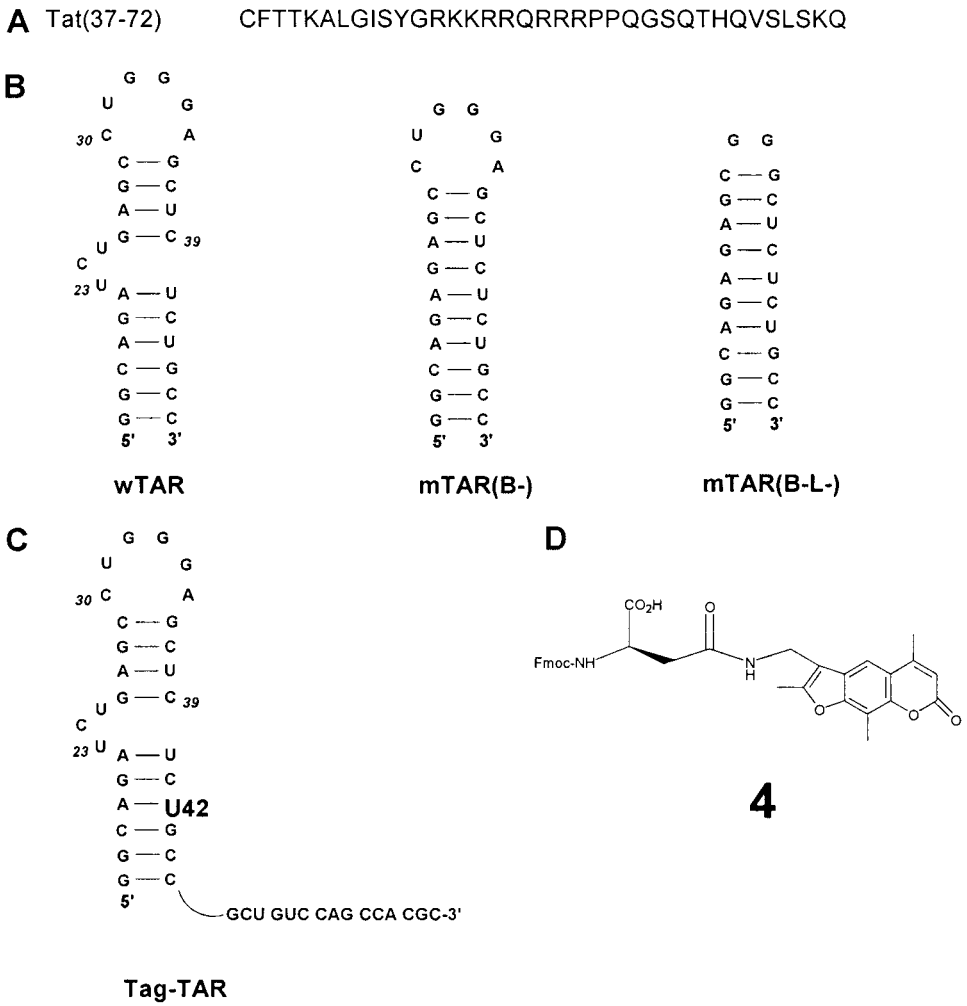

FIGURE 1: (A) The 36-residue peptide, amino acids 37−72, that contains the RNA-binding domain of Tat. (B) Secondary structures of wild-type and mutant TAR RNAs used in this study. Wild-type TAR RNA spans the minimal sequences that are required for Tat responsiveness in vivo (*20*) and for in vitro binding of Tat-derived peptides (*29*). Wild-type TAR contains two non-wild-type base pairs to increase transcription by T7 RNA polymerase. Numbering of nucleotides in the RNA corrsponds to their positions in wild-type TAR RNA. Wild-type TAR, mutant TAR lacking the bulge sequence, and a duplex TAR without the bulge and loop sequence are labeled as wTAR, mTAR(B−), and mTAR(B−L−), respectively. (C) Secondary structure of Tag TAR RNA that contains an extra 15 nucleotides at the 3′-end of TAR RNA to allow the hybridization of DNA primer for reverse transcriptase analysis of the cross-link. Uridine 42, a psoralen addition site, is highlighted. (D) A psoralen-modified amino acid, *N*-α-Fmoc-L-aspartic acid-*β*-(4′aminomethyl-4,5′,8-trimethylpsoralen) **(4)**, which was used during solid-phase assembly of Tat peptides.

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 recently devised methods based on psoralen photochemistry to identify specific contacts in RNA−protein complexes (*33*−*35*).

Attachment of psoralen to a nucleic acid binding molecule creates an efficient nucleic acid cross-linking molecule. Psoralens are bifunctional photoreagents that have been used as photoactive probes of nucleic acid structure and function (*36*, *37*). Psoralens intercalate between base pairs of double-stranded nucleic acids. Upon ultraviolet irradiation (320−400 nm), the intercalated psoralens photoreact with adjacently stacked pyrimidine bases to form pyrimidine−psoralen monoadducts (*38*, *39*). Typical photoreactions of a psoralen are shown in Figure 2. Psoralens form two kinds of monoadducts: the furan-side monoadduct, M_{Fu}, which is formed through the cycloaddition between the 4′,5′ double bond of psoralen and the 5,6 double bond of a pyrimidine base, and the pyrone-side monoadduct, M_{Py}, which is formed through the cycloaddition between the 3,4 double bond of a psoralen and the 5,6 double bond of a pyrimidine base (Figure 2). Oligonucleotides carrying psoralen at a specific site in the sequence are powerful tools for molecular biology and could provide a new class of therapeutic agents (*40*, *41*). Photochemical properties of psoralens can be exploited to generate DNA probes containing psoralen monoadducts at specific sites, and these probes can form site-specific cross-links to the complementary target sequences (*42*, *43*). Psoralen-modified oligonucleotides have been used to trap three-stranded RecA−DNA complexs and to study the repair of these cross-linked complexes (*44*, *45*). It has also been shown that psoralen-derivatized oligodeoxyribonucleotides and oligodeoxyribonucleoside methylphosphonates can specifically cross-link to double-helical DNA and viral mRNA targets, respectively (*46*−*48*). We reasoned that, as with nucleic acids, sequence-specific nucleic acid binding peptides can

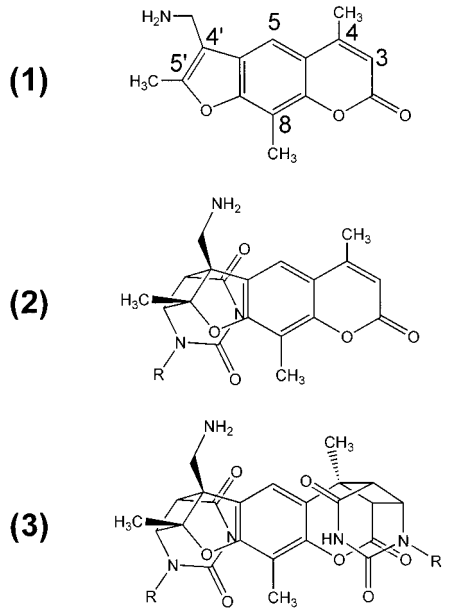

FIGURE 2: Structures of the two types of 4′-aminomethyl-4,5′,8-trimethylpsoralen (AMT)−uridine adducts formed by the photoreaction of AMT with RNA: (**1**) AMT; (**2**) furan-side monoadduct; (**3**) furan-side and pyrone-side diadduct.

be converted into sequence-specific cross-linking peptides by introducing a psoralen at a specific site within the peptide sequence. In a protein−nucleic acid complex, site-specifically placed psoralen would photoreact with adjacent pyrimidines. The protein−nucleic acid cross-linked products can be purified and characterized. Isolation and identification of the cross-linked products could provide important information about the three-dimensional structure of the protein−nucleic acid complex, revealing, for example, which bases of the nucleic acid are close to the psoralen site.

In this report, we have incorporated a psoralen-modified amino acid into Tat peptide sequences during solid-phase peptide synthesis and used these psoralen-containing peptides to investigate the effect of amino acids from the core domain of Tat for TAR recognition and on Tat peptide orientation in the Tat−TAR complex. We synthesized two fragments of the Tat protein containing the RNA-binding region and various residues from the activation domain of the Tat protein (Figure 1). A psoralen-modified amino acid was incorporated at position 41 in Tat peptides and these sequence-specific RNA-binding peptides were converted into sequence-specific RNA−cross-linking peptides by UV (360 nm) irradiation. Specificity of RNA−peptide interactions was studied by competition experiments. Our results show that the addition of residues 37−41 from the Tat sequence significantly enhanced specificity of the peptide−RNA interaction without altering the cross-link site on TAR RNA or the chemical nature of the cross-link.

EXPERIMENTAL PROCEDURES

Synthesis of Psoralen−Tat Conjugates. Synthesis of *N*-α-Fmoc-L-aspartic acid-*β*-(4′aminomethyl-4,5′,8-trimethylpsoralen) **(4)** has been previously described (*33*).

A Tat-derived peptide (from amino acids 42−72) was synthesized on an Applied Biosystems 431A peptide synthesizer using standard FastMoc protocols. All Fmoc-amino acids, piperidine, 4-(dimethylamino)pyridine, dichloromethane, *N*,*N*-dimethylformamide, 1-hydroxybenzotriazole (HOBT), 2-(1*H*-benzotriazol-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. Psoralen attachment to the N-terminus of the Tat(42−72) peptide was achieved by using compound **4** and standard FastMoc coupling reagents. At this point, a portion of the resin containing the psoralen-modified peptide was removed from the reaction vessel for cleavage and deprotection. Further addition of 37−40 amino acids of the Tat sequence was accomplished by standard FastMoc chemistry. 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 TFA, 100 μL of thioanisole, 100 μL of water, and 50 μL of ethanedithiol (*49*). After cleavage from the resin, peptide was purified by HPLC on a Zorbax 300 SB-C_8 column. The mass of fully deprotected and purified peptides was confirmed by FAB mass spectrometry.

Buffers. All buffer pH values refer to measurements at room temperature. TK buffer: 50 mM Tris-HCl (pH 7.4), 20 mM KCl, and 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.

RNA Synthesis. Wild-type and mutant TAR RNAs were prepared by in vitro transcription (*50*, *51*). All DNAs were synthesized on an Applied Biosystems ABI 392 DNA/RNA synthesizer. The template strands encode the sequences for wild-type and mutant TAR RNAs. The top strand is a short piece of DNA complementary to the 3′-end of all template DNAs having the sequence 5′TAATACGACTCACTATAG3′. The template strand of DNA was annealed to an equimolar amount of the 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 (20 μL) containing 8.0 pmol of template DNA, 40−60 units of T7 polymerase (Promega) was used. Transcription reactions were stopped by addition of an equal volume of sample loading buffer. RNA was purified on 20% acrylamide−8 M urea denaturing gels and stored in DEPC water at −20 °C.

Enzymatically transcribed 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 incubation with 16 units T4 polynucleotide kinase (New England Biolabs) in 70 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, and 5 mM DTT (*51*, *52*). 5′-End-labeled RNAs were gel purified on a denaturing gel, visualized by autoradiography, eluted out of the gels, and deslated on a reverse-phase cartridge. The sequence of RNAs was determined by base hydrolysis and nuclease digestion.

Photo-Cross-Linking Reactions. Psoralen-modified Tat-(37−72) or Tat(42−72) was used to form a complex with 5′-end-labeled TAR RNA at room temperature in TK buffer for 30 min in the dark. A typical reaction mixture contained a 10 μL volume with final concentrations of 0.25 μM labeled TAR RNA and 0.25 μM psoralen−Tat peptides. The reaction mixture was UV irradiated (360 nm) for 1−10 min in a Rayonet RR 100 photochemical reactor. After UV irradiation, 10 μL of the sample loading buffer was added, and sample was electrophoresed on a 20% polyacrylamide−8 M urea gel. Efficiencies of cross-linking were determined by phosphorimage analysis. During preparative scale cross-linking, cross-linked products were visualized by autoradiograhy and recovered from gels as described earlier (*33*).

RESULTS AND DISCUSSION

Site-Specific Incorporation of a Psoralen into Tat Peptides. We synthesized two psoralen-containing Tat peptides, psoralen−Tat(42−72) and psoralen−Tat(37−72), for our cross-linking studies. To incorporate psoralen residues at internal peptide sequences, we synthesized an amino acid analogue, *N*-α-Fmoc-L-aspartic acid-*β*-(4′aminomethyl-4,5′,8-trimethylpsoralen) (compound **4**), containing a psoralen derivative. Compound **4** was designed to be compatible with the Fmoc solid-phase peptide synthesis strategy. AMT, 4′aminomethyl-4,5′,8-trimethylpsoralen, was synthesized as described by Isaacs et al. (*53*). Detailed synthesis of compound **4** has been described elsewhere (*33*). A Tat-derived peptide (from amino acids 42 to 72) was synthesized by using standard HOBT/HBTU FastMoc protocols (*54*). Psoralen attachment to the N-terminus of the Tat(42−72) peptide was achieved by using compound **4** and standard FastMoc coupling reagents. Fmoc groups from *N*-α-Fmoc-L-aspartic acid-*β*-(4′aminomethyl-4,5′,8-trimethylpsoralen) conjugated to the Tat peptide were deprotected by piperidine treatment, and further addition of 37−40 amino acids of the Tat sequence was accomplished by standard FastMoc chemistry on an automated 431 ABI peptide synthesizer. After cleavage from the resin and deprotection, peptide was purified by reverse-phase high-performance liquid chromatography. Psoralen−Tat conjugates were purified by HPLC and characterized by mass spectrometry.

To further characterize and evaluate the binding capabilities of psoralen−peptide conjugates, we determined the dissociation constants for psoralen−peptide and compared them with those of the wild-type peptides. Psoralen−Tat-(42−72) binds TAR RNA with high affinities, and psoralen did not affect the TAR RNA binding activity of the peptide (*33*). Equilibrium dissociation constants of the Tat(37−72)−TAR complexes were measured using direct and competition electrophoretic mobility assays (*25*, *26*, *55*, *56*). We determined the relative dissociation constants (K_{rel}) by measuring the ratios of wild-type Tat(37−72) to psoralen−Tat (37−72) dissociation constants (K_d) for TAR RNA binding. The calculated value for K_{rel} was 1.25, indicating that a psoralen-derivatized Asp at position 41 of Tat(37−72) did not significantly alter the structure of the Tat(37−72), thus preserving the TAR-binding affinities of the peptide (data not shown).

Site-Specific Photo-Cross-Linking of Psoralen−Asp41−Tat(37−72) to TAR RNA. Psoralen-modified Tat(37−72) peptide was used to form a complex with 5′-^{32}P-end-labeled TAR RNA at room temperature in TK buffer and ultraviolet irradiated (360 nm) for short periods of time. Cross-linked products were separated by denaturing polyacrylamide gel electrophoresis. Results of these experiments are shown in Figure 3. Irradiation of RNA and the psoralen−peptide complex yields a new band with electrophoretic mobility less than that of TAR RNA (lane 4). These results indicate that upon irradiation this psoralen−peptide yields a single RNA−protein cross-link with high efficiency, ∼10%. Both the psoralen−peptide and UV (360 nm) irradiation are required for the formation of a cross-linked RNA−protein complex (see lanes 3 and 4). Further control experiments showed that no cross-linking was observed when RNA and the unmodified peptide were irradiated (lane 2). Since the cross-linked RNA−peptide complex is stable to alkaline pH (9.5), high temperature (85 °C), and denaturing conditions (8 M urea), we conclude that a covalent bond is formed between TAR RNA and the peptide during the cross-linking reaction.

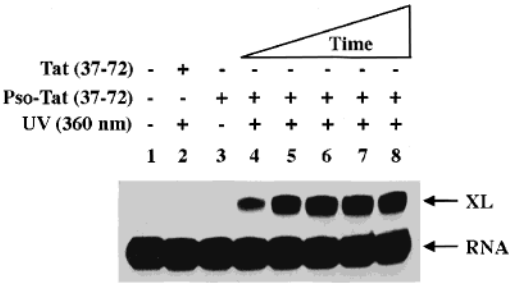

FIGURE 3: Site-specific photo-cross-linking reaction of psoralen−Asp41−Tat(37−72) and the TAR RNA complex. TAR RNA was 5′-end-labeled with ^{32}P, and cross-linked products were resolved on 20% polyacrylamide−8 M urea gels and visualized by autoradiography. RNA−peptide complexes in lanes 4−8 were irradiated for 1, 2, 5, 10, and 15 min, respectively. The psoralen−Asp41−Tat(37−72) conjugate and the RNA−protein cross-link are indicated by Pso−Tat and XL, respectively.

Cys37−Thr40 Amino Acids of Tat Enhance Specificity of the Tat−TAR Cross-Link Formation. We next determined the specificity of psoralen−Tat peptides for cross-linking reactions with TAR RNA. The specificity of the cross-linking reaction was determined by competition experiments. We synthesized three TAR RNA sequences for competition reactions: wild-type TAR RNA, bulgeless TAR RNA, and a duplex TAR RNA without the bulge and loop sequence (Figure 1). Cross-linking reactions were performed in a 15 μL volume containing 0.25 μM 5′-^{32}P-labeled TAR RNA, 0.25 μM psoralen−Tat peptide, 25 mM Tris·HCl (pH 7.4), 100 mM NaCl, and up to 4.5 μM unlabeled competitor RNA. Cross-linked products were separated by 8 M urea−20% polyacrylamide gels and visualized by phosphorimage analysis. Figure 4 shows that psoralen−Asp41−Tat(37−72)-TAR cross-linking was inhibited only by the addition of unlabeled wild-type TAR RNA and not by mutant TAR RNAs lacking the trinucleotide bulge or duplex TAR without the bulge and loop residues. As shown in Figure 5, the psoralen−Asp41−Tat(42−72)−TAR cross-linking reaction was not highly specific because mutant RNA was also able to compete for peptide binding and inhibited the cross-link formation. It is interesting that, during competition experiments for psoralen−Asp41−Tat(37−72)−TAR cross-linking, we observed that the addition of competitor RNA first caused an increase of the RNA−peptide cross-linked products. It is quite

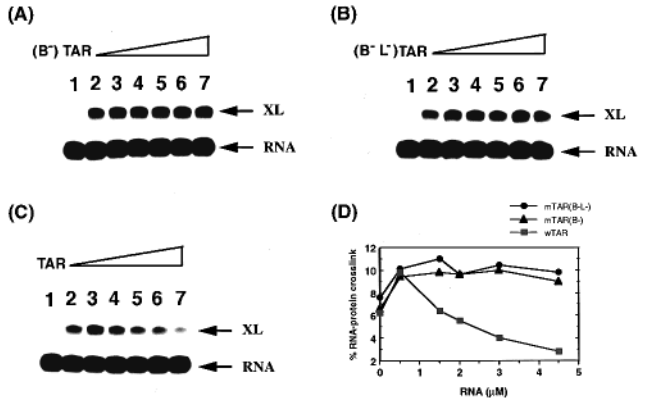

FIGURE 4: Specificity of the psoralen−Asp41−Tat(37−72)-TAR cross-linking reaction determined by competition assays. RNA−protein complexes were formed between 0.25 μM ^{32}P-5′-end-labeled TAR RNA and 0.25 μM psoralen−Asp41−Tat(37−72) in the presence of unlabeled bulgeless TAR (A), duplex TAR without the bulge and loop sequence (B), or wild-type TAR RNA (C). Lane 1 was a control RNA−peptide complex without UV irradiation. The RNA−protein cross-link is shown as XL. Concentrations of the competitor RNA in lanes 2, 3, 4, 5, 6, and 7 were 0, 0.5, 1.5, 2.0, 3.0, and 4.5 μM, respectively. (D) Quantitative analysis of competition experiments. The fraction of RNA in the RNA−peptide complex was determined by phosphorimage analysis. Wild-type TAR, bulgeless TAR, and duplex TAR without the bulge and loop residues are labeled as wTAR, mTAR(B−), and mTAR(B−L−), respectively.

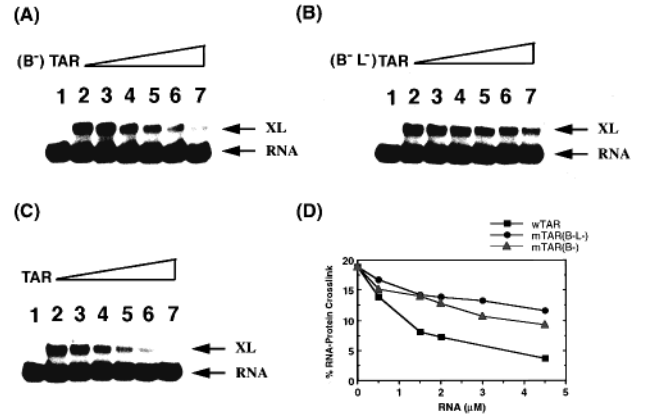

FIGURE 5: Specificity of the psoralen−Asp41−Tat(42−72)-TAR cross-linking reaction determined by competition assays. RNA−protein complexes were formed between 0.25 μM 5′-^{32}P-end-labeled TAR RNA and 0.25 μM psoralen−Asp41−Tat(42−72) in the presence of unlabeled bulgeless TAR (A), duplex TAR without the bulge and loop sequence (B), or wild-type TAR RNA (C). Lane 1 was a control RNA−peptide complex without UV irradiation. The RNA−protein cross-link is shown as XL. Concentrations of the competitor RNA in lanes 2, 3, 4, 5, 6, and 7 were 0, 0.5, 1.5, 2.0, 3.0, and 4.5 μM, respectively. (D) Quantitative analysis of competition experiments. The fraction of RNA in the RNA−peptide complex was determined by phosphorimage analysis. Wild-type TAR, bulgeless TAR, and duplex TAR without the bulge and loop residues are labeled as wTAR, mTAR(B−), and mTAR(B−L−), respectively.

possible that addition of competitor RNA neutralizes the charge or interacts with the core region of Tat to unmask the basic RNA-binding region. However, the effect of competitor RNA on psoralen−nucleic acid interaction cannot be ruled out by these experiments. Taken together, these results indicate that the Cys37−Thr40 amino acids of the Tat core region enhance the specificity of the Tat−TAR cross-link formation.

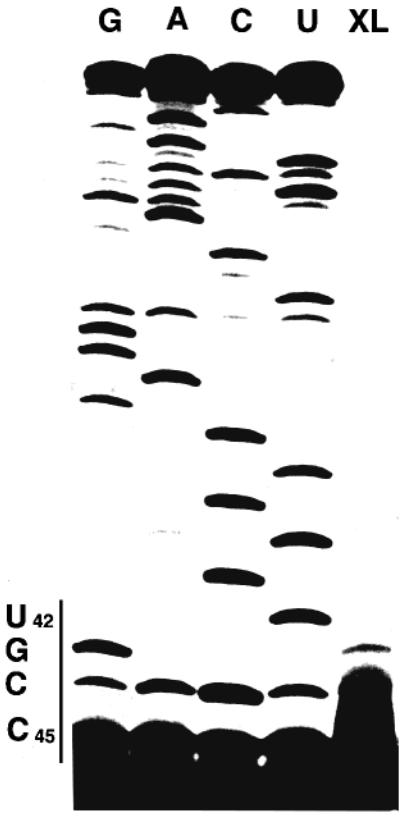

FIGURE 6: Mapping the exact position of the psoralen cross-link to TAR RNA by primer extension analysis. Gel-purified cross-linked Tag TAR RNA, which contains 15 nucleotides at its 3′-end (Figure 1C and ref *33*), was primer extended by hybridizing an oligonucleotide complementary to the 15 nucleotides at the 3′-end (lane XL). Lanes of sequencing reactions are labeled as G, A, C, and U. Sequencing lanes are presented as their representative complementary RNA position rather than the cDNA itself. The sequence of TAR RNA in the cross-linked region is shown on the left.

Uridine 42 in TAR RNA Cross-Links to Psoralen−Asp41−Tat(37−72). To map the exact site of the cross-link on TAR RNA, we carried out primer extension analysis of the gel-purified RNA−protein cross-link (*57*, *58*). Reverse transcriptase synthesizes cDNA copies from an RNA template, and this enzyme stops at a point where RNA has been modified or cross-linked (*57*−*62*). The stops are mapped by comparing their cDNA length with standard chain termination sequencing. During primer extension of psoralen cross-linked RNA, AMV reverse transcriptase displays a characteristic pattern by pausing at one nucleotide 3′ of the cross-link site. However, there are some instances where the transcriptase will advance to the adducted site (*57*−*59*, *62*−*64*). In this case, a characteristic doublet is produced, corresponding to stops at one nucleotide 3′ of the cross-link and up to the cross-link site. We synthesized TAR RNA with an extra 15 nucleotides at its 3′-end, Tag TAR (Figure 1 and ref *33*). The 15-nucleotide extension in Tag TAR was used to hybridize the DNA primer for reverse transcriptase analysis of the cross-link. Results of this analysis are shown in Figure 6. In our experiments, only one stop was observed at G43, corresponding to an adduct on the 3′ face of the cross-linked nucleotide (*59*). Reverse transcriptase analysis on wild-type TAR RNA showed that there was no pause site at G43 (data not shown). Mapping of the cross-link site on TAR RNA to single nucleotide resolution was also carried

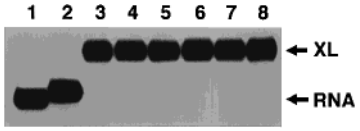

FIGURE 7: Determination of the chemical nature of TAR RNA and the psoralen−Tat(37−72) cross-link. Protease K digestion: lane 1, 5′-^{32}P-end-labeled TAR RNA; lane 2, 5′-^{32}P-end-labeled TAR−psoralen−Tat cross-link complex after protease K digest. Photoreversal of the gel-purified 5′-^{32}P-end-labeled cross-link at 365 nm, light intensity 5×10^{16} photons/s: lane 3, the dark control; lanes 4−8, samples exposed to 6×10^{18}, 1.2×10^{19}, 1.8×10^{19}, 2.4×10^{19}, and 3×10^{19} photons, respectively.

out by partial RNase digestion and alkaline hydrolysis of the gel-purified RNA−protein cross-link as described previously (*33*, *34*), which showed that the cross-linking occurred at U42 (data not shown). Thus, we conclude that the psoralen addition site is U42 in TAR RNA.

Primer extension analysis and alkaline hydrolysis of the TAR−psoralen−Tat(42−72) cross-link also showed that the psoralen cross-link site was at U42 (*33*). It is interesting that incorporation of four residues, Cys37−Thr40, from the Tat core region enhanced the specificity for TAR RNA recognition without altering the psoralen addition site on the RNA. There are two possible explanations for the enhanced specificity of the Tat(37−72) peptide. (a) The basic region of Tat initiates contacts with TAR RNA in the major groove, causing RNA to undergo a structural change, exposing additional sites for interactions with Tat core region. (b) After initial contacts with TAR RNA, the Tat conformation is changed and the core residues stabilize this Tat structure that is required for specific TAR recognition. In the context of the full-length Tat protein, CycT1 interacts with the Tat core region and could stabilize a Tat structure involved in specific TAR RNA binding in human cells.

The Pyrone Ring of Psoralen−Asp41−Tat(37−72) Photoreacts with TAR RNA. The photoreaction between psoralens and RNA can be divided into at least three distinct steps: (1) formation of a noncovalent complex with RNA via intercalation of the psoralen between adjacent base pairs; (2) photoreaction between the psoralen and a pyrimidine base to yield a monoadduct; (3) absorption of a second photon to yield an interstrand cross-link. The psoralen−Tat(37−72) conjugate binds TAR RNA and brings psoralen close to the double-helical region of lower stem of the RNA. After intercalation and upon UV irradiation, psoralen can form a monoadduct with U42 or it can photoreact with two bases to create an interstrand cross-link. This situation raises two questions regarding the chemical nature of psoralen−peptide and RNA cross-link: (a) Is it a monoadduct or an interstrand cross-link? (b) If it is a monoadduct, which side of the psoralen is added to the RNA, furan or pyrone? To answer the first question, we performed protease K digestion of the psoralen−Tat(37−72) and TAR RNA cross-link and electrophoresed next to TAR RNA. After complete digestion of the peptide, a linear RNA was recovered that migrated one nucleotide slower than the wild-type TAR RNA (Figure 7, lane 2). This indicates that psoralen formed a monoadduct cross-link with U42 in TAR RNA rather than a diadduct between two bases, which would result in a hairpin structure and migrate much slower than a linear RNA.

The second question was answered by carrying out photoreversal experiments on the cross-linked complexes. Shi and Hearst reported a detailed study describing wavelength dependence for the photoreversal of monoadducts (*65*). In the case of the pyrone-side monoadducts, two absorption bands contribute to the photoreversal with a quantum yield of 2×10^{-2} at wavelengths below 250 nm and 7×10^{-3} at wavelengths from 287 to 314 nm (*65*). For the furan-side monoadducts, at least three absorption bands contribute to the photoreversal with a varying quantum yield from 5×10^{-2} at wavelengths below 250 nm to 7×10^{-4} at wavelengths between 295 and 365 nm (*65*). After purification of the 5′-end-labeled cross-linked complexes, we conducted a photoreversal experiment by UV irradiating the samples at 365 nm with varying light intensities. As shown in Figure 7 (lanes 4−8), there was no detectable photoreversal of the cross-linked complex. These results show that the monoaddition to U42 on TAR RNA is at the pyrone side of the psoralen.

CONCLUSIONS

We have used a site-specific cross-linking method to determine the effect of core residues from the Tat sequence on the protein orientation in the Tat−TAR complex and on the specificity of Tat−TAR binding. Our results show that Lys41 of Tat(37−72) is close to uridine 42 in the lower stem of TAR RNA. Our results also show that the addition of only four residues (Cys37−Thr40) from the Tat core region significantly enhanced the specificity of Tat peptide−TAR interactions. An important test of the biological significance of RNA−peptide complexes is their specificity, especially in the context of a vast excess of nonspecific RNA in the cell. It is well established that arginine-rich peptides bind TAR RNA with high affinities. However, these arginine-rich peptides are not highly specific for TAR RNA recognition. On the basis of our findings, it can be envisioned to design peptides that can bind TAR RNA with high affinities and specificities by incorporating the four residues (Cys37−Thr40) from the Tat core region to the arginine-rich peptide sequences.

Incorporation of a psoralen in a nucleic acid binding peptide sequence creates a bifunctional peptide which can recognize and cross-link specific sites on nucleic acids. Other multifunctional peptides can be designed which are capable of recognizing specific substrates and their chemical modifications. Since psoralen photo-cross-linking does not require any previous structural knowledge of the protein or protein−nucleic acid complex, it should provide a general approach to study other RNA- and DNA-binding proteins. Such psoralen−peptide conjugates provide a new class of probes for sequence-specific protein−nucleic acid interactions and could be used to selectively control gene expression or to induce site-directed mutations.

ACKNOWLEDGMENT

We thank Yueh-Hsin Ping for assistance in preparation of the illustrations.

REFERENCES

1. Cullen, B. R. (1998) *Cell 93*, 685−692.
2. Emerman, M., and Malim, M. (1998) *Science 280*, 1880−1884.

3. Jeang, K.-T., Xiao, H., and Rich, E. A. (1999) *J. Biol. Chem. 274*, 28837−28840.
4. Taube, R., Fujinaga, K., Wimmer, J., Barboric, M., and Peterlin, B. M. (1999) *Virology 264*, 245−253.
5. Karn, J. (1999) *J. Mol. Biol. 293*, 235−254.
6. Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., and Price, D. H. (1997) *Genes Dev. 11*, 2622−2632.
7. Mancebo, H. S. Y., Lee, G., Flygare, J., Tomassini, J., Luu, P., Zhu, Y., Peng, J., Blau, C., Hazuda, D., Price, D., and Flores, O. (1997) *Genes Dev. 11*, 2633−2644.
8. Gold, M., Yang, X., Herrmann, C., and Rice, A. (1998) *J. Virol. 72*, 4448−4453.
9. Wei, P., Garber, M. E., Fang, S.-M., Fischer, W. H., and Jones, K. A. (1998) *Cell 92*, 451−462.
10. Garber, M. E., Wei, P., KewalRamani, V. N., Mayall, T. P., Herrmann, C. H., Rice, A. P., Littman, D. R., and Jones, K. A. (1998) *Genes Dev. 12*, 3512−3527.
11. Garber, M. E., Wei, P., and Jones, K. A. (1998) *Cold Spring Harbor Sump. Quant. Biol. 63*, 371−380.
12. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7791−7796.
13. Bieniasz, P. D., Grdina, T. A., Bogerd, H. P., and Cullen, B. R. (1998) *EMBO J. 17*, 7056−7065.
14. Fujinaga, K., Cujec, T., Peng, J., Garriga, J., Price, D., Grana, X., and Peterlin, B. (1998) *J. Virol. 72*, 7154−7159.
15. Zhou, Q., Chen, D., Pierstorff, E., and Luo, K. (1998) *EMBO J. 17*, 3681−3691.
16. Wimmer, J., Fujinaga, K., Taube, R., Cujec, T., Zhu, Y., Peng, J., Price, D., and Peterlin, B. (1999) *Virology 255*, 182−189.
17. Ivanov, D., Kwak, Y. T., Nee, E., Guo, J., Garcia-Martinez, L. F., and Gaynor, R. B. (1999) *J. Mol. Biol. 288*, 41−56.
18. Rana, T. M., and Jeang, K.-T. (1999) *Arch. Biochem. Biophys. 365*, 175−185.
19. Rosen, C. A., Sodroski, J. G., and Haseltine, W. A. (1985) *Cell 41*, 813−823.
20. Jakobovits, A., Smith, D. H., Jakobovits, E. B., and Capon, D. J. (1988) *Mol. Cell. Biol. 8*, 2555−2561.
21. Selby, M. J., Bain, E. S., Luciw, P. A., and Peterlin, B. M. (1989) *Genes Dev. 3*, 547−558.
22. Garcia, J. A., Harrich, D., Soultanakis, E., Wu, F., Zmitsuyasu, R., and Gaynor, R. B. (1989) *EMBO J. 8*, 765−778.
23. Muesing, M. A., Smith, D. H., and Capon, D. A. (1987) *Cell 48*, 691−701.
24. Berkhout, B., Silverman, R. H., and Jeang, K. T. (1989) *Cell 59*, 273−82.
25. 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.
26. 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.
27. Feng, S., and Holland, E. C. (1988) *Nature 334*, 165−167.
28. Berkhout, B., and Jeang, K.-T. (1989) *J. Virol. 63*, 5501−5504.
29. 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 Colonno, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8985−8989.
30. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., and Sonenberg, N. (1991) *J. Virol. 65*, 5196−5202.
31. Tinoco, I., Jr., Puglisi, J. D., and Wyatt, J. R. (1990) *Nucleic Acids Mol. Biol. 4*, 205−226.
32. Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989) *BioEssays 11*, 100−106.
33. Wang, Z., and Rana, T. M. (1995) *J. Am. Chem. Soc. 117*, 5438−5444.
34. Wang, Z., Wang, X., and Rana, T. M. (1996) *J. Biol. Chem. 271*, 16995−16998.
35. Wang, Z., and Rana, T. M. (1999) in *RNA−protein interaction protocols* (Haynes, S. R., Ed.) pp 49−62, Humana Press, Totowa, NJ.
36. Hearst, J. E. (1981) *Annu. Rev. Biophys. Bioeng. 10*, 69−86.
37. Cimino, G. D., Gamper, H. B., Isaacs, S. T., and Hearst, J. E. (1985) *Annu. Rev. Biochem. 54*, 1151−1193.
38. Kanne, D., Straub, K., Hearst, J. E., and Rapoport, H. (1982) *J. Am. Chem. Soc. 104*, 6754−6764.
39. Straub, K., Kanne, D., Hearst, J. E., and Rapoport, H. (1981) *J. Am. Chem. Soc. 103*, 2347−2355.
40. Hearst, J. E. (1988) *Annu. Rev. Phys. Chem. 39*, 291−315.
41. Helene, C., Thuong, N. T., and Harel-Bellan, A. (1992) *Ann. N.Y. Acad. Sci. 660*, 27−36.
42. Gamper, H. B., Cimino, G. D., Isaacs, S. T., Ferguson, M., and Hearst, J. E. (1986) *Nucleic Acids Res. 14*, 9943−9954.
43. Van Houten, B., Gamper, H. B., Hearst, J. E., and Sancar, A. (1986) *J. Biol. Chem. 261*, 14135−14141.
44. Cheng, S., Van Houten, B., Gamper, H. B., Sancar, A., and Hearst, J. E. (1988) *J. Biol. Chem. 263*, 15110−15117.
45. Cheng, S., Sancar, A., and Hearst, J. E. (1991) *Nucleic Acids Res. 19*, 657−663.
46. Takasugi, M., Guendouz, A., Chassingnol, M., Decout, J. L., Lhomme, J., Thuong, N. T., and Helene, C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5602−5606.
47. Lee, B. L., Murakami, A., Blake, K. R., Lin, S.-B., and Miller, P. S. (1988) *Biochemistry 27*, 3197−3203.
48. Kean, J. M., and Miller, P. S. (1994) *Biochemistry 33*, 9178−9186.
49. King, D. S., Fields, C. G., and Fields, G. B. (1990) *Int. J. Pept. Protein Res. 36*, 255−266.
50. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
51. Wang, Z., and Rana, T. M. (1996) *Biochemistry 35*, 6491−6499.
52. Neenhold, H. R., and Rana, T. M. (1995) *Biochemistry 34*, 6303−6309.
53. Isaacs, S. T., Shen, C.-k., J., Hearst, J. E., and Rapoport, H. (1977) *Biochemistry 16*, 1058−1064.
54. Knorr, R., Trzeciak, A., Bannwarth, W., and Gillessen, D. (1989) *Tetrahedron Lett. 30*, 1927−1930.
55. Fried, M., and Crothers, D. M. (1981) *Nucleic Acids Res. 9*, 6505−6525.
56. 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.
57. Youvan, D. C., and Hearst, J. E. (1981) *Nucleic Acids Res. 9*, 1723−1741.
58. Youvan, D. C., and Hearst, J. E. (1982) *Anal. Biochem. 119*, 86−89.
59. Ericson, G., and Wollenzien, P. (1988) *Anal. Biochem. 174*, 215−223.
60. Burgin, A. B., and Pace, N. R. (1990) *EMBO J. 9*, 4111−4118.
61. Harris, M. E., Nolan, J. M., Malhotra, A., Brown, J. W., Harvey, S. C., and Pace, N. R. (1994) *EMBO J. 13*, 3953−3963.
62. Nolan, J. M., Burke, D. H., and Pace, N. R. (1993) *Science 261*, 762−765.
63. Ericson, G., and Wollenzien, P. (1989) *J. Biol. Chem. 264*, 540−545.
64. Watkins, K. P., Dungan, J. M., and Agabian, N. (1994) *Cell 76*, 171−182.
65. Shi, Y.-b., and Hearst, J. E. (1987) *Biochemistry 26*, 3786−3792.

BI0028744