Chemical Conversion of a *trans*-Activation Responsive RNA-Binding Fragment of HIV-1 Tat Protein into a Site-Specific Cross-Linking Agent

Zhuying Wang and Tariq M. Rana*

Contribution from the Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, and Biochemistry & Molecular Biology Graduate Program at Rutgers State University, 675 Hoes Lane, Piscataway, New Jersey 08854

Received November 10, 1994[⊗]

Abstract: Replication of human immunodeficiency virus type 1 (HIV-1) requires specific interactions of Tat protein with the *trans*-activation responsive (TAR) region of RNA, a 59-base stem−loop structure located at the 5′-end of all mRNAs. We have devised a new method based on psoralen photochemistry to identify a specific contact between a fragment of Tat protein (residues 42−72) and TAR RNA. We synthesized a 30-amino acid fragment containing the arginine-rich RNA-binding domain of Tat(42−72) and chemically attached a psoralen at the amino terminus. Upon near ultraviolet irradiation (360 nm), this synthetic psoralen−peptide cross-linked to a single site in the TAR RNA sequence. The RNA−protein complex was purified, and the cross-link site on TAR RNA was determined by chemical and primer extension analyses. Our results show that the amino terminus of Tat(42−72) contacts, or is close to, uridine 42 in the lower stem of TAR RNA. 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.

Introduction

Human immunodeficiency virus type 1 (HIV-1), like other lentiviruses, encodes a *trans*-activating regulatory protein, Tat, that is essential for *trans*-activation of viral gene expression.[1−4] In the absence of Tat, most of the viral transcripts terminate prematurely, producing short RNA molecules ranging in size from 60 to 80 nucleotides. 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′-end of all nascent HIV-1 transcripts.[5−9] Upon binding to the TAR RNA sequence, Tat causes a substantial increase in transcript levels.[10−14] The increased efficiency in transcription may result from preventing premature termination of the transcriptional elongation complex[15] or directly at the level of initiation of transcription.[16] TAR was originally localized to nucleotides +1 to +80 within the viral long terminal repeat (LTR).[5] Subsequent studies have further mapped its 3′-boundary to +44.[8] Nucleotides spanning positions +19 to +42 are sufficient for Tat responsiveness *in vivo*.[8] The TAR RNA contains a six-nucleotide loop and a three-nucleotide pyrimidine bulge which separates two helical stem regions.[5,8,9,12] The trinucleotide bulge is essential for the high affinity and specific binding of the Tat protein.[17,18]

The Tat protein is a small, cysteine-rich nuclear protein containing 86 amino acids and is comprised of three important functional domains. The Tat protein is absolutely essential for viral replication, making it an important potential target for the development of new drug therapies against HIV infections. Understanding the principles of Tat−TAR interactions is a crucial step for drug design. However, there is little structural information available for this RNA−protein complex. In the absence of high-resolution crystallographic and nuclear magnetic resonance data, new methods are needed to determine the topology of RNA−protein complexes under physiological conditions. We have devised a new method based on psoralen photochemistry to identify a specific contact between a fragment of Tat protein (residues 42−72) and TAR RNA.

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.[19,20] Psoralens intercalate between base pairs of double-stranded nucleic acids. Upon ultraviolet irradiation (320−400 nm), the

* To whom correspondence should be addressed at the University of Medicine and Dentistry of New Jersey. Phone: (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) Cullen, B. R. *Microbiol. Rev.* **1992**, *56*, 375−394.
(2) Gaynor, R. *AIDS* **1992**, *6*, 374−63.
(3) 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. *Nature* **1986**, *320*, 367−371.
(4) Dayton, A. I.; Sodroski, J. G.; Rosen, C. A.; Goh, W. C.; Haseltine, W. A. *Cell* **1986**, *44*, 941−947.
(5) Rosen, C. A.; Sodroski, J. G.; Haseltine, W. A. *Cell* **1985**, *41*, 813−823.
(6) Hauber, J.; Cullen, B. R. *J. Virol.* **1988**, *62*, 673−679.
(7) Feng, S.; Holland, E. C. *Nature* **1988**, *334*, 165−168.
(8) Jakobovits, A.; Smith, D. H.; Jakobovits, E. B.; Capon, D. J. *Mol. Cell. Biol.* **1988**, *8*, 2555−2561.
(9) Berkhout, B.; Silverman, R. H.; Jeang, K. T. *Cell* **1989**, *59*, 273−82.
(10) Cullen, B. R. *Cell* **1986**, *46*, 973−982.
(11) Peterlin, B. M.; Luciw, P. A.; Barr, P. J.; Walker, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 9734−9738.
(12) Muesing, M. A.; Smith, D. H.; Capon, D. A. *Cell* **1987**, *48*, 691−701.
(13) Rice, A. P.; Mathews, M. B. *Nature* **1988**, *332*, 551−553.
(14) Laspia, M. F.; Rice, A. P.; Mathews, M. B. *Cell* **1989**, *59*, 283−292.
(15) Kao, S.-Y.; Calman, A. F.; Luciw, P. A.; Peterlin, B. M. *Nature* **1987**, *330*, 489- 493.
(16) Sharp, P. A.; Marciniak, R. A. *Cell* **1989**, *59*, 229−230.
(17) Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Heaphy, S.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A.; Valerio, R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 6925−6929.
(18) Dingwall, C.; Ernberg, I.; Gait, M. J.; Green, S. M.; Heaphy, S.; Karn, J.; Lowe, A. D.; Singh, M.; Skinner, M. A. *EMBO J.* **1990**, *9*, 4145−4153.
(19) Hearst, J. E. *Annu. Rev. Biophys. Bioeng.* **1981**, *10*, 69−86.
(20) Cimino, G. D.; Gamper, H. B.; Isaacs, S. T.; Hearst, J. E. *Annu. Rev. Biochem.* **1985**, *54*, 1151−93.

0002-7863/95/1517-5438$09.00/0

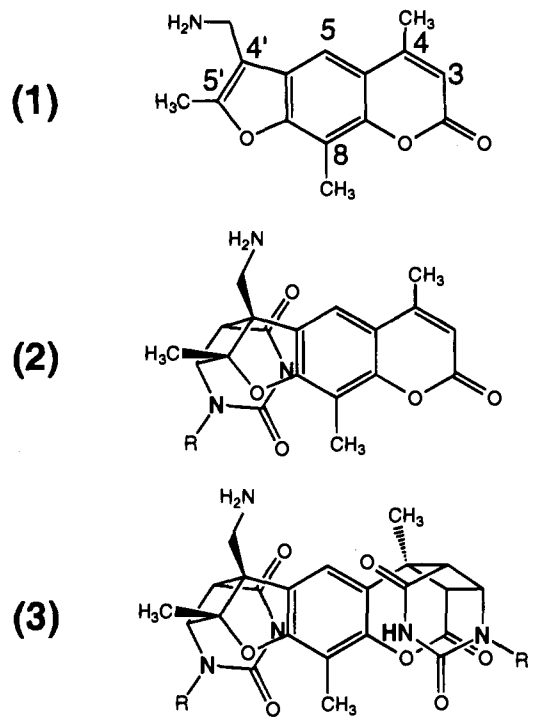

Figure 1. Structures of the two types of 4′-(aminomethyl)-4,5′,8-trimethylpsoralen (AMT)−uridine adducts formed by the photoreaction of AMT with RNA: (1) AMT; (2) furan-side monoadduct; (3) furan-side and pyrone-side diadduct.

intercalated psoralens photoreact with adjacently stacked pyrimidine bases to form a pyrimidine−psoralen monoadduct.[21,22] Typical photoreactions of a psoralen are shown in Figure 1. 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 1). 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.[23-25] 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.[26,27] Psoralen-modified oligonucleotides have been used to trap three-stranded RecA−DNA complexes and to study the repair of these cross-linked complexes.[28,29] It has also been shown that psoralen-derivatized oligodeoxyribonucleotides and oligoribonucleoside methylphosphonates can specifically cross-link to double-helical DNA and viral mRNA targets, respectively.[30-32] We reasoned that, as with nucleic acids, sequence-specific nucleic acid binding peptides can 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.

Before such experiments can be successfully developed, a number of fundamental questions must be addressed. Psoralen-containing protected amino acids that are suitable for solid-phase peptide synthesis must be synthesized. Conditions must be devised or identified under which psoralen molecules remain intact after the harsh treatment of peptide synthesis, cleavage from the resin, and deprotection. Chemical linkage of psoralen to the amino acid should not affect its photochemical cross-linking efficiencies. If information about the three-dimensional structure of a nucleic acid−protein complex is sought, it should be possible to carry out the photochemistry under the physiological conditions of pH, temperature, and ionic strength. Generally, it will be preferred that the cross-linking efficiencies are high so that the cross-link material can easily be isolated and characterized.

Here we report the development of this new methodology to study nucleic acid−protein interactions in solution. We synthesized a 30-residue peptide containing the basic region of Tat protein, and converted this sequence-specific RNA-binding peptide into a sequence-specific RNA-cross-linking peptide by covalent attachment of a photo-cross-linking reagent, psoralen, at the NH_2-terminus of the peptide. Upon ultraviolet irradiation (365 nm), psoralen-modified Tat peptide cross-linked to TAR RNA, revealing the base contact and orientation of the NH_2-terminus of the peptide when bound to RNA.

Experimental Section

Proton nuclear magnetic resonance spectra were recorded at 200 MHz on a Gemini spectrometer (Varian) using tetramethylsilane as an internal reference. Chemical shifts are reported in parts per million (ppm) downfield from tetramethylsilane. Mass spectra were recorded using the fast atom bombardment (FAB) technique at the Midwest Center for Mass Spectrometry at the University of Nebraska. Infrared spectra were recorded on a Perkin-Elmer FT 1600 series FTIR. Reversed-phase HPLC purification of peptides was carried out on a Beckman 344 with a variable wavelength detector (Beckman Model 165) by using a C_8 reversed-phase column (Zorbax 300 SB, 4.6 mm × 25 cm). Thin layer chromatography was performed with precoated 0.2 μm silica gel 60 F-254 TLC plates (EM Reagents, Darmstadt, FRG). Plates were visualized under long wave UV light and with iodine vapors. Column chromatography was performed using silica gel (70−230 mesh, 60 Å) purchased from Aldrich. Reagent grade chemicals were used as such unless noted. All the reactions were carried out in the dark.

***N*-α-Fmoc-L-Aspartic Acid−β-(4′-(Aminomethyl-4,5′,8-trimethylpsoralen) α-*tert*-Butyl Ester (3).** To an ice-cooled solution of *N*-α-Fmoc-aspartic acid α-*tert*-butyl ester (**1**) (206 mg, 0.5 mmol) in dry dichloromethane (2 mL) was added *N*-hydroxybenzotriazole (HOBt) (68 mg, 0.5 mmol) dissolved in dry DMF (2 mL), followed by DCC (103 mg, 0.5 mmol) in dry dichloromethane (2 mL). The reaction mixture was stirred at 0 °C for 1 h and at room temperature for 2 h.

(21) Kanne, D.; Straub, K.; Hearst, J. E.; Rapoport, H. *J. Am. Chem. Soc.* **1982**, *104*, 6754−6764.
(22) Straub, K.; Kanne, D.; Hearst, J. E.; Rapoport, H. *J. Am. Chem. Soc.* **1982**, *103*, 2347−2355.
(23) Hearst, J. E. *Annu. Rev. Phys. Chem.* **1988**, *39*, 291−315.
(24) Sastry, S. S.; Spielmann, H. P.; Hearst, J. E. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1992**, *66*, 85−147.
(25) Hélène, C.; Toulmé, J.-J. *Biochim. Biophys. Acta* **1990**, *1049*, 99−125.
(26) Gamper, H. B.; Cimino, G. D.; Isaacs, S. T.; Ferguson, M.; Hearst, J. E. *Nucleic Acids Res.* **1986**, *14*, 9943−9954.
(27) Van Houten, B.; Gamper, H. B.; Hearst, J. E.; Sancar, A. *J. Biol. Chem.* **1986**, *261*, 14135−14141.
(28) Cheng, S.; Van Houten, B.; Gamper, H. B.; Sancar, A.; Hearst, J. E. *J. Biol. Chem.* **1988**, *263*, 15110−15117.
(29) Cheng, S.; Sancar, A.; Hearst, J. E. *Nucleic Acids Res.* **1991**, *19*, 657−663.
(30) Takasugi, M.; Guendouz, A.; Chassingnol, M.; Decout, J. L.; Lhomme, J.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5602−5606.
(31) Lee, B. L.; Murakami, A.; Blake, K. R.; Lin, S.-B.; Miller, P. S. *Biochemistry* **1988**, *27*, 3197−3203.
(32) Kean, J. M.; Miller, P. S. *Biochemistry* **1994**, *33*, 9178−9186.

The reaction mixture was filtered, the filtrate was evaporated to dryness, and the activated ester was redissolved in dry dichloromethane (4 mL). To form a free base of psoralen, a suspension of 4′-(aminomethyl)-4,5′,8-trimethylpsoralen (AMT) hydrochloride (**2**) (117 mg, 0.4 mmol) in dry DMF (5 mL) was mixed with *N,N*-diisopropylethylamine (0.07 mL, 0.4 mmol), and stirred at room temperature for 2 min. The coupling reaction was started by adding the free base of AMT to the activated ester solution. The final pH of the coupling reaction was adjusted to 8.0 by the addition of diisopropylethylamine (50 μL). The reaction mixture was stirred for 15 min, and coupling was complete as monitored by TLC. The reaction mixture was concentrated to dryness under reduced pressure. The residue was taken up in ethyl acetate (15 mL) and washed with water (5 × 2 mL). The organic phase was dried over molecular sieves, and the solvent was removed under reduced pressure. The crude product thus obtained was purified by recrystallization using methanol to provide compound **3** as a white solid (208 mg, 0.32 mmol; 80%). ^{1}H NMR (Me_2SO-d_6): δ 7.90−7.24 (9H, m, aromatic), 6.30 (1H, s, olefinic), 4.40−4.32 (3H, m, CH_2 and CH), 4.23−4.17 (2H, m, CH_2), 4.09−4.06 (1H, m, CH), 2.51 (3H, s, CH_3), 2.49 (3H, s, CH_3), 2.48 (3H, s, CH_3), 2.42−2.39 (2H, m, β CH_2), 1.27 (9H, s, CH_3). IR: 1728 and1702 cm^{-1} (C═O, s). HRMS (FAB) for $C_{38}H_{38}N_2O_8$: 673.2529 (M + Na).

***N*-α-Fmoc-L-Aspartic Acid−β-(4′-(Aminomethyl)-4,5′,8-trimethylpsoralen) (4).** Compound **3** (200 mg, 0.3 mmol.) was dissolved in dry dichloromethane (2 mL) and cooled to 0 °C. To this solution was added trifluoroacetic acid (2 mL), and stirring was continued at 0 °C for 2 h, followed by overnight stirring at room temperature. The reaction mixture was concentrated in vacuo without heating to dryness. Silica gel TLC confirmed that compound **4** was pure. Yield: 172 mg, 0.29 mmol; 99%. ^{1}H NMR: δ 7.90−7.25 (9H, m, aromatic), 6.28 (1H, s, olefinic), 4.50−4.35 (1H,m, CH), 4.30−3.90 (5H, m, CH and $2CH_2$), 2.51 (3H, s, CH_3), 2.50 (3H, s, CH_3), 2.48 (3H, s, CH_3), 2.36 (2H, m, β CH_2). IR: 1728 and 1701 cm^{-1} (C═O, s). MS (FAB) for $C_{34}H_{31}N_2O_8$: 595.2 (M + H).

Peptide Synthesis. 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), *N,N*-diisopropylethylamine, and HMP-linked polystyrene resin were obtained from Applied Biosystems Division, Perkin-Elmer. Trifluoroacetic acid, 1,2-ethanedithiol, phenol, and thioanisol 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. Psoralen attachment to the N-terminus of Tat(42−72) peptide was achieved by using compound **4** and standard FastMoc coupling reagents. 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 1,2-ethanedithiol.[33] After cleavage from the resin, peptide was purified by HPLC on a Zorbax 300 SB-C_8 column. The masses of fully deprotected and purified peptides were confirmed by FAB mass spectrometry: calculated mass for Tat(42−72) $C_{152}H_{263}N_{59}O_{43}$ 3605.5, found 3606.5 (M + H); calculated mass for Asp−psoralen−Tat(42−72) $C_{171}H_{283}N_{61}O_{48}$ 3957.5, found 3958.5 (M + H).

Buffers. All buffer pH values refer to measurements at room temperature: (TK buffer) 50 mM Tris−HCl (pH 7.4), 20 mM KCl, 0.1% Triton X-100; (transcription buffer) 40 mM Tris−HCl (pH 8.1), 1 mM spermidine, 0.01% Triton X-100, 5 mM DTT; (TBE buffer) 45 mM Tris−borate, pH, 8.0, 1 mM EDTA; (sample loading buffer) 9 M urea, 1 mM EDTA and 0.1% bromophenol blue in 1X TBE buffer; (binding buffer) 25 mM Tris−HCl (pH 7.5), 100 mM NaCl, 1 mM $MgCl_2$, 0.1% Triton X-100; (RT buffer) 34 mM Tris−HCl (pH 8.3), 50 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT; hydrolysis buffer) 50 mM Na_2CO_3−$NaHCO_3$, pH 9.2.

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 Tag TAR RNA (Figure 2). The top strand is a short piece of DNA complementary to the 3′-end of all template DNAs having the sequence

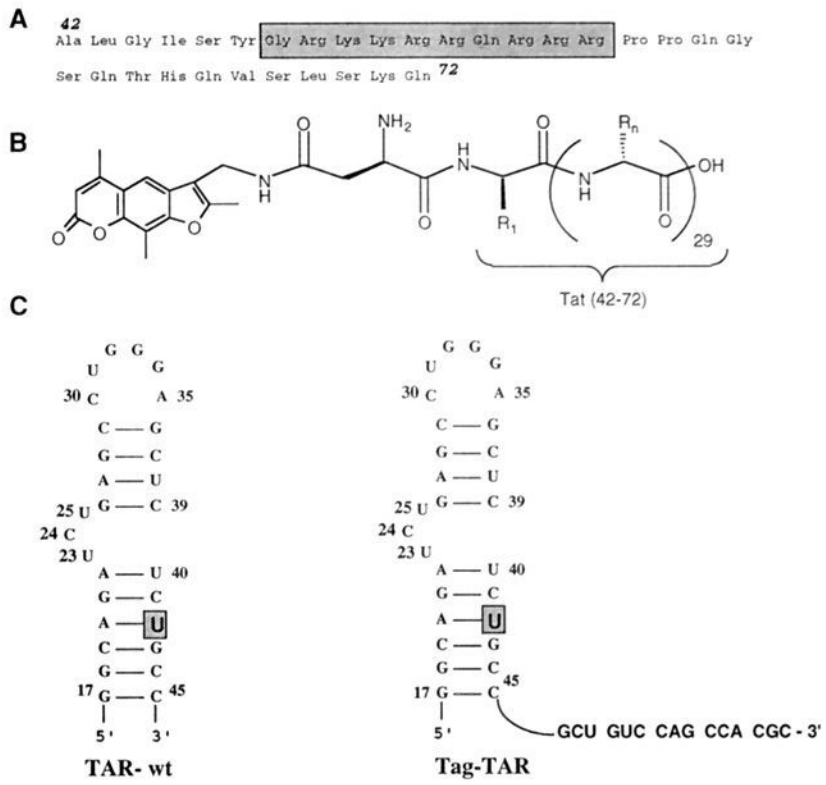

Figure 2. (A) 30-residue peptide, amino acids 42−72, containing the RNA-binding domain of Tat. The arginine rich region of Tat, corresponding to the Tat 48−57 peptide, is highlighted. (B) Structure of the synthetic cross-linking peptide, psoralen−Tat(42−72). (C) Secondary structures of wild-type and Tag TAR RNAs. Wild-type TAR RNA spans the minimal sequences that are required for Tat responsiveness *in vivo*,[8] and for *in vitro* binding of Tat-derived peptides.[55] Wild-type TAR contains two non-wild-type base pairs to increase transcription by T7 RNA polymerase. Tag TAR RNA 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 in both the TAR and Tag TAR RNAs.

5′-TAATACGACTCACTATAG-3′. DNA was deprotected in NH_4-OH at 55 °C for 8 h and then dried in a Savant Speedvac. The samples were resuspended in sample loading buffer and were purified on 20% acrylamide−8 M urea denaturing gels, 50 cm × 0.8 mm. Gels were run for 3 h at 30 W until 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 concentrations of DNAs were determined by measuring absorbance at 260 nm in a Shimadzu UV spectrophotometer. Samples were stored at −20 °C.

RNAs. All RNAs were prepared by *in vitro* transcription using the method of Milligan *et al.*[34] 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 (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. RNAs were stored in DEPC water at −20 °C.

TAR RNAs were 5′-dephosphorylated by incubation with calf intestinal alkaline phophatase (Promega) for 1 h at 37 °C in 50 mM Tris−Cl, pH 9.0, 1 mM $MgCl_2$, 0.1 mM $ZnCL_2$, and 1 mM spermidine. The RNAs were purified by multiple extractions with Tris-saturated phenol and one extraction with 24:1 chloroform−isoamyl alcohol followed by ethanol precipitation. The RNAs were 5′-end-labeled with 0.5 μM [γ-^{32}P]ATP (6000 Ci/mmol) (ICN) per 100 pmol of RNA by incubating with 16 units of T4 polynucleotide kinase (New England Biolabs) in 70 mM Tris−HCl, pH 7.5, 10 mM $MgCl_2$, and 5 mM DTT.[35] RNAs were gel purified on a denaturing gel, visualized by autoradiography, and recovered from gels as described above.

(33) King, D. S.; Fields, C. G.; Fields, G. B. *Int. J. Peptide Protein Res.* **1990**, *36*, 255−266.

(34) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783−8798.

(35) Kuchino, Y.; Nishimura, S. *Methods Enzymol.* **1989**, *180*, 154−163.

Scheme 1

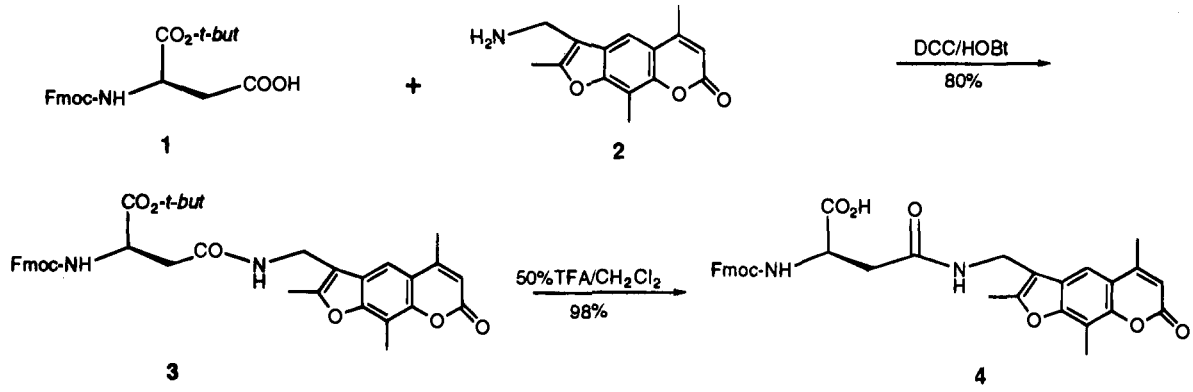

Photo-Cross-Linking Reactions. Psoralen-modified Tat(42−72) or wild-type Tat(42−72) was used to form a complex with 5′-end-labeled TAR RNA at room temperature in TK buffer for 30 min in the dark. A typical reaction mixture contained 20 μL of volume with final concentration of 0.5 μM labeled TAR RNA and 2.0 μM Psoralen−Tat(42−72). The reaction mixture was UV irradiated (360 nm) for 20 min in a Rayonet RR 100 photochemical reactor (Southern New England Ultraviolet). After UV irradiation, 20 μL of sample loading buffer was added, and the sample was electrophoresed on a 20% acrylamide−8 M urea gel. Efficiencies of cross-linking were determined by phosphor image analysis. During preparative scale cross-linking, cross-linked products were visualized by autoradiograhy and recovered from gels as described above.

Primer Extension Analysis. AMV reverse transcriptase was used to map the cross-link site on TAR RNA.[36−39] We synthesized TAR RNA with 15 extra nucleotides at its 3′-end, Tag TAR (Figure 2). The 15 nucleotide extension in Tag TAR was used to hybridize DNA primer for reverse transcriptase analysis of the cross-link. Large scale cross-linking reactions were performed with unlabeled Tag TAR and the psoralen−peptide complex. Cross-linked products were purified and digested with protease K to hydrolyze the Tat peptide. The cross-linked Tag TAR and 5′-^{32}P-end-labeled cDNA primer were mixed in RT buffer. After heating at 90 °C, the samples were allowed to cool slowly to 50 °C. The extension was initiated by addition of AMV reverse transcriptase and deoxynucleotides. Marker lanes were prepared by adding the appropriate amounts of dideoxynucleotides.[39]

Sequencing Reactions. Alkaline hydrolysis of RNAs was carried out in hydrolysis buffer for 5−10 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 phosphor image analysis.

Results and Discussion

Synthesis of a Psoralen-Labeled Tat(42−72). The synthetic route for psoralen-modified amino acid is outlined in Scheme 1. *N*-α-*tert*-Butyloxycarbonyl (Boc) protected amino acids are commonly used for Merrifield solid phase peptide synthesis.[40] There are two major concerns about this synthetic strategy: (a) repetitive TFA acidolysis in Boc-group deprotection could lead to acid-catalyzed side reactions; (b) cleavage and deprotection of peptides requires HF and a specific laboratory setup which is not available to many researchers. Due to these concerns, Fmoc (9-fluorenylmethyl carbamate) solid phase peptide synthesis was developed which employs *N*-α-Fmoc-amino acids.[41−43] In this strategy, the Fmoc group is deprotected with piperidine and TFA is required only for the final cleavage and deprotection step. 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.*[44] Compound **3** was prepared in two steps: (i) an active ester of *N*-α-Fmoc-L-aspartic acid α-*tert*-butyl ester (**1**) was prepared by using 1-hydroxybenzotriazole (HOBt) and dicyclohexylcarbodiimide; (ii) coupling of the active ester to 4′-(aminomethyl)-4,5′,8-trimethylpsoralen. The coupling reaction is very facile, and the product was isolated in high yield. Treatment of 50% TFA in dry dichloromethane gave the final product **4**. A Tat-derived peptide (from amino acids 42−72) was synthesized by using standard HOBT/HBTU FastMoc protocols.[45] Psoralen attachment to the N-terminus of Tat(42−72) peptide was achieved by using compound **4** and standard FastMoc coupling reagents. After cleavage from the resin and deprotection, the peptide was purified by reversed-phase high-performance liquid chromatography. The mass of the peptide was confirmed by fast atom bombardment (FAB) mass spectrometry. The structure of psoralen-labeled Tat peptide is shown in Figure 2B.

Determination of Dissociation Constants. To further characterize and evaluate the binding capabilities of the psoralen−peptide conjugate, we determined the dissociation constants for psoralen−peptide and compared them with those of the wild-type peptide (Tat(42−72)). Equilibrium dissociation constants of the Tat(42−72)−TAR complexes were measured using direct and competition electrophoretic mobility assays.[17,18,46,47] For direct mobility shift assays, the fractional saturation of 0.5 μM 5′-^{32}P-end-labeled TAR RNA was measured as a function of wild-type and psoralen-modified Tat-(42−72). A typical gel of these experiments is shown in Figure 3. Binding constants were calculated from three sets of experiments. Using the wild-type Tat(42−72), 60% of the TAR RNA was bound by 0.5 μM peptide ($K_d \approx 0.13$ μM). The psoralen−peptide showed a little lower affinity with $K_d \approx 0.24$ μM; 51% of the TAR RNA was bound by 0.5 μM psoralen−

(36) Youvan, D. C.; Hearst, J. E. *Nucleic Acids Res.* **1981**, *9*, 1723−1741.
(37) Youvan, D. C.; Hearst, J. E. *Anal. Biochem.* **1982**, *119*, 86−89.
(38) Wollenzien, P. *Methods Enzymol.* **1988**, *164*, 319−329.
(39) Ericson, G.; Wollenzien, P. *Anal. Biochem.* **1988**, *174*, 215−223.
(40) Merrifield, R. B. *Adv. Enzymol.* **1969**, *32*, 221−298.
(41) Fields, G. B.; Noble, R. L. *Int. J. Peptide Protein Res.* **1990**, *35*, 161−214.
(42) Carpino, L. A.; Han, G. Y. *J. Org. Chem.* **1972**, *37*, 3404−3409.
(43) Hazel Fox, E. A.; Harkiss, D.; Logan, C. G.; Sheppard, R. C.; Williams, B. J. *J. Chem. Soc., Chem. Commun.* **1978**, *537*−539,
(44) Isaacs, S. T.; Shen, C.-k., J.; Hearst, J. E.; Rapoport, H. *Biochemistry* **1977**, *16*, 1058−1064.
(45) Knorr, R.; Trzeciak, A.; Bannwarth, W.; Gillessen, D. *Tetrahedron Lett.* **1989**, *30*, 1927−1930.
(46) Fried, M.; Crothers, D. M. *Nucleic Acids Res.* **1981**, *9*, 6505.
(47) Churcher, M. J.; Lamont, C.; Hamy, F.; Dingwall, C.; Green, S. M.; Lowe, A. D.; Butler, P. J. C.; Gait, M. J.; Karn, J. *J. Mol. Biol.* **1993**, *230*, 90−110.

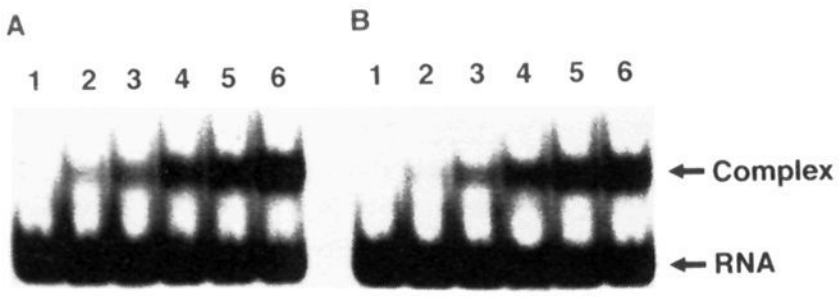

Figure 3. Binding of TAR RNA to the Tat(42−72) peptide (A) or psoralen−Tat conjugate (B). Binding reactions contained 0.5 μM 5′-^{32}P-end-labeled TAR RNA and 0.1, 0.2, 0.3, 0.4, and 0.5 μM Tat(42−72) or psoralen−Tat conjugate, lanes 2−6, respectively. Lane 1 was a control lane without the peptide. Complex formation was performed in the binding buffer and incubation at room temperature for 10 min. Complexes were separated from unbound RNA by electrophoresis in nondenaturing 8% polyacrylamide gels containing 0.1% Triton X-100. Gels were run in a cold room at 260 V for 2 h. The relative amounts of free and bound RNA were determined by phosphor imaging.

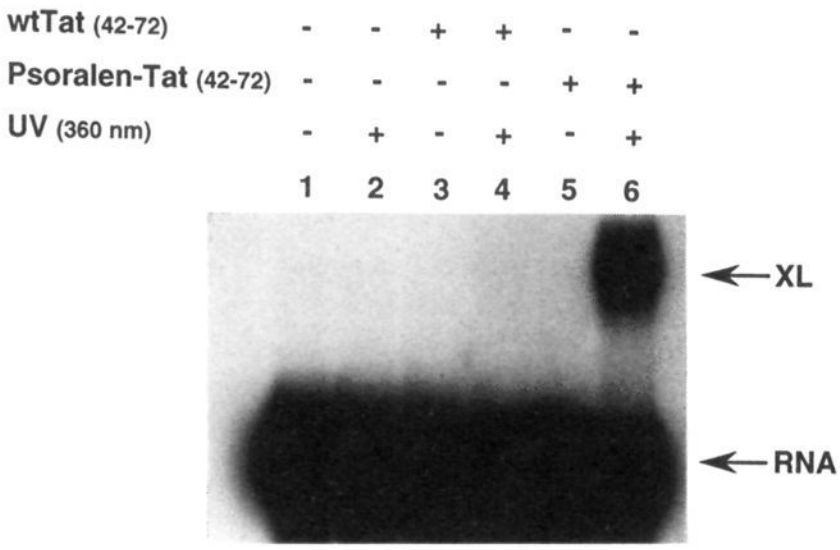

Figure 4. Cross-linking reaction of the Psoralen−Tat(42−72) and TAR RNA complex. Cross-linked products were resolved on 20% polyacrylamide−8 M urea gels and visualized by autoradiography: (lane 1) RNA without peptide or irradiation, (lane 2) RNA irridiated in the absence of peptide, (lane 3) RNA with unmodified Tat(42−72) and no irridiation, (lane 4) RNA irradiated with unmodified Tat(42−72), (lane 5) RNA with psoralen−Tat(42−72) and no irradiation, (lane 6) RNA irradiated in the presence of psoralen−Tat(42−72). The RNA−protein cross-link is indicated by XL.

peptide. From the results of gel shift experiments, we conclude that attachment of a psoralen-derivatized aspartic acid to the amino terminus of Tat(42−72) did not significantly alter the structure of Tat(42−72), thus preserving the TAR RNA-binding affinities of the peptide.

Site-Specific Photo-Cross-Linking of Psoralen−Tat to TAR RNA. Psoralen-modified Tat(42−72) 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 20 min. Cross-linked products were separated by denaturing polyacrylamide gel electrophoresis. Results of this experiment are shown in Figure 4. These results indicate that upon irradiation this psoralen−peptide yields a single RNA−peptide cross-link with high efficiency, ~20%. Control experiments showed that TAR RNA neither alone nor with unmodified peptide gave a cross-link under similar conditions. Further control experiments established that a specific RNA−protein complex between TAR RNA and psoralen−peptide was necessary for photo-cross-linking. Cross-linking was inhibited by the addition of unlabeled wild-type TAR RNA and not by a mutant TAR RNA without a trinucleotide bulge. Additional control experiments showed that cross-linking did not occur with yeast tRNA^{Phe} and psoralen−peptide (data not shown).

Identification of the Cross-Link Site on TAR RNA. To map the exact site of cross-link on TAR RNA, we carried out primer extension analysis of the gel-purified RNA−protein cross-link.[36,37] Reverse transcriptase synthesizes cDNA copies from an RNA template, and this enzyme stops at a point where RNA has been modified or cross-linked.[36,37,39,48−50] 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 from the 3′-end of the cross-link site. However, there are some instances where the transcriptase will advance to the adducted site.[36,37,39,50−52] In this case, a characteristic doublet is produced, corresponding to stops at one nucleotide from the 3′-end of the cross-link and up to the cross-link site. We synthesized TAR RNA with 15 extra nucleotides at its 3′-end, Tag TAR (Figure 2C). 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 5A. In our experiments, only one stop was observed at G43, corresponding to an adduct on the 3′-face of the cross-linked nucleotide.[39] Thus, we conclude that the psoralen addition site is U42 in TAR RNA. Reverse transcriptase analysis on wild-type TAR RNA showed that there was no pause site at G43 (data not shown).

Reverse transcriptase provides a well-characterized method to map psoralen cross-link sites on RNA. However, the enzyme cannot bypass the adduct site; therefore, any modification downstream from the first adduct cannot be identified. To overcome this problem of reverse transcriptase and to ensure that psoralen−Tat was cross-linked to a single site on TAR RNA, we performed limited alkaline hydrolysis on 5′-^{32}P-end-labeled TAR RNA and the cross-linked complex. In this experiment, psoralen−Tat(42−72) was cross-linked to wild-type TAR RNA instead of Tag TAR. After purification by gel electrophoresis and desalting by the reversed-phase cartridge, the cross-linked RNA was subjected to alkaline hydrolysis. Results of this experiment are shown in Figure 5B. Base hydrolysis generated a ladder of RNA degradation products. Fragment sizes were determined by comparison with TAR RNA hydrolysis and sequencing. Alkaline hydrolysis of the 5′-end-labeled cross-linked complexes resulted in an RNA ladder in which all fragments up to 25 nucleotides in length were resolved, while bands corresponding to higher molecular mass precipitously disappeared. Apparent insensitivity to base hydrolysis appeared at position 25 in the sequence of TAR RNA (Figure 5B, lane 2). A band corresponding to G21 is horseshoe shaped and not absent in alkaline hydrolysis of the cross-linked RNA (Figure 5B, lane 2). Repeat gels of the same experiment showed a standard hydrolysis ladder in this region; therefore, a minor band shift at G21 is most likely due to different gel conditions (salts, dyes, etc.). A standard base hydrolysis ladder was observed for 5′-end-labeled TAR RNA, showing no insensitivity to base hydrolysis at position 25 in the sequence (Figure 5B, lane 1). These observations clearly support two conclusions: (a) there is a single cross-link site in the complex; (b) the cross-link site is at U42 as evidenced by primer extension analysis. These results also indicate that the psoralen−peptide binds specifically to TAR RNA in such a fashion that psoralen photoreacts with only a single uridine in the TAR RNA sequence.

(48) Burgin, A. B.; Pace, N. R. *EMBO J.* **1990**, *9*, 4111−8.
(49) Harris, M. E.; Nolan, J. M.; Malhotra, A.; Brown, J. W.; Harvey, S. C.; Pace, N. R. *EMBO J.* **1994**, *13*, 3953−63.
(50) Nolan, J. M.; Burke, D. H.; Pace, N. R. *Science* **1993**, *261*, 762−765.
(51) Ericson, G.; Wollenzien, P. *J. Biol. Chem.* **1989**, *264*, 540−545.
(52) Watkins, K. P.; Dungan, J. M.; Agabian, N. *Cell* **1994**, *76*, 171−182.

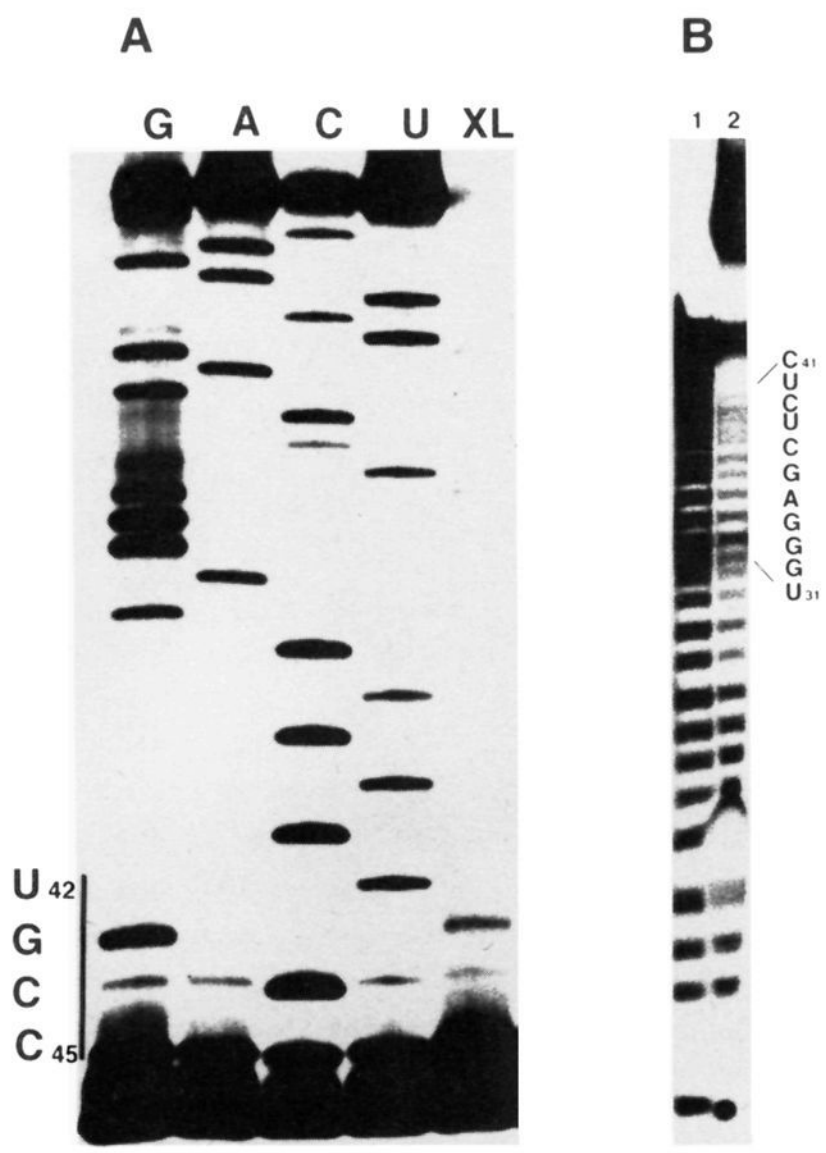

Figure 5. (A) Mapping of the exact position of the psoralen cross-link to TAR RNA by primer extension analysis. Gel-purified cross-linked Tag TAR RNA was the primer extended by hybridizing an oligonucleotide complementary to the 15 nucleotides at the 3′-end (lane XL). Lanes of sequencing reactions are 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. (B) Base hydrolysis analysis of TAR RNA cross-linked to psoralen−Tat(42−72): (lane 1) alkaline hydrolysis of 5′-^{32}P-end-labeled TAR RNA, (lane 2) alkaline hydrolysis of the 5′-^{32}P-end-labeled TAR−psoralen−Tat cross-link complex. The sequence of 11 nucleotides in the region of RNA which showed decreased sensitivity to base hydrolysis is shown on the right; bands with molecular mass higher than C41 disappeared.

Chemical Nature of the Cross-Link. 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(42−72) conjugate binds TAR RNA and brings psoralen close to the double helical region of the lower stem of 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 the psoralen−peptide and RNA cross-link: (a) Is it a monoadduct or 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(42−72) 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 6, 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

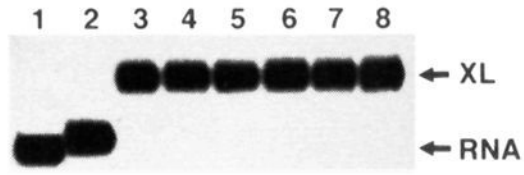

Figure 6. Determination of the chemical nature of TAR RNA and the psoralen−Tat(42−72) cross-link. Protease K digestion: (lane 1) 5′-^{32}P-end-labeled TAR RNA, (lane 2) 5′-^{32}P-end-labeled TAR−psoralen−Tat cross-link complex after protease K digest. Photoreversal of the gel-purified 5′-^{32}P-end-labeled 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.

result in a hairpin structure and migrate much slower than a linear RNA. Additional evidence to support monoaddition rather than a cross-link comes from alkaline hydrolysis of 5′-end-labeled cross-linked complexes. In the case of an interstrand cross-link, a stop after four or five bases would have been prominent (Figure 5B). The second question was answered by carrying out photoreversal experiments on the cross-link complexes. Shi and Hearst reported a detailed study describing wavelength dependence for the photoreversal of monoadducts.[53] With 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.[53] 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.[53] 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 6 (lanes 4−8), there was no detectable photoreversal of the cross-link complex. These results show that the monoaddition to U42 on TAR RNA is at the pyrone side of the psoralen.

Applications to RNA−Protein Structures. The photo-cross-linking results have demonstrated where the NH_2-terminal residue of psoralen−peptide is located relative to the RNA-binding site on TAR RNA. Peptides containing the basic domain (residues 48−57) of Tat protein can bind TAR RNA with high affinities.[47,54−58] In a recent study, circular dichroism (CD) and two-dimensional NMR were used to determine the peptide backbone conformation of a 25-amino acid hybrid peptide that contained the basic RNA-binding domain of HIV-1 Tat fused to a 15-residue activation domain from the Tat protein of equine infectious anemia virus.[59] This study revealed that the basic domain of HIV-1 Tat(47−57) forms a stable α-helical structure in solution. It has been proposed by chemical footprinting studies that a trinucleotide bulge introduces distortion into the geometry of TAR RNA to enlarge the major groove for protein recognition.[58] In another recent report, Hamy *et al.*[60]

(53) Shi, Y.-b.; Hearst, J. E. *Biochemistry* **1987**, *26*, 3786−3792.
(54) Weeks, K. M.; Ampe, C.; Schultz, S. C.; Steitz, T. A.; Crothers, D. M. *Science* **1990**, *249*, 1281−1285.
(55) 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. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8985−8989.
(56) Calnan, B. J.; Tidor, B.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Science* **1991**, *252*, 1167−1171.
(57) Calnan, B. J.; Biancalana, S.; Hudson, D.; Frankel, A. D. *Genes Dev.* **1991**, *5*, 201−210.
(58) Weeks, K. M.; Crothers, D. M. *Cell* **1991**, *66*, 577−588.
(59) Mujeeb, A.; Bishop, K.; Peterlin, B. M.; Turk, C.; Parslow, T. G.; James, T. L. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 8248−8252.
(60) Hamy, F.; Asseline, U.; Grasby, J.; Iwai, S.; Pritchard, C.; Slim, G.; Butler, P. J. G.; Karn, J.; Gait, M. J. *J. Mol. Biol.* **1993**, *230*, 111−123.

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. This suggests a model in which Tat binds to TAR RNA by inserting the basic recognition sequence, a putative α-helix, into the major groove with an orientation where lysine 41 in the core domain of Tat contacts the lower stem of TAR RNA (Figure 7). The knowledge of the orientation of Tat in the Tat–TAR complex has significant implications for understanding gene regulation in HIV-1.

Summary. We have synthesized a 30-amino acid Tat fragment and covalently attached a psoralen at the NH_2-terminus. Upon near ultraviolet irradiation (365 nm), psoralen–Tat(42–72) cross-linked to a pyrimidine of TAR RNA at the Tat binding site, revealing the location of the NH_2-terminus of Tat(42–72). The attachment of a psoralen to a nucleic acid binding peptide 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. This new method based on psoralen photo-cross-linking does not require any previous structural knowledge of the protein or protein–nucleic acid complex and 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 Gabriele Varani for coordinates for the RNA model, Kavita Shah for preliminary synthesis and purification of peptide conjugates, Marc Gartenberg for helpful discussions, and Heather Neenhold for her assistance in phosphor image analysis, and illustrations. This work was supported by Grant AI 34785 from the National Institutes of Health.

Supplementary Material Available: Figures showing data from a competition experiment with unlabeled wild-type TAR and mutant TAR RNA and primer extension analysis of wild-type TAR (2 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA943658J

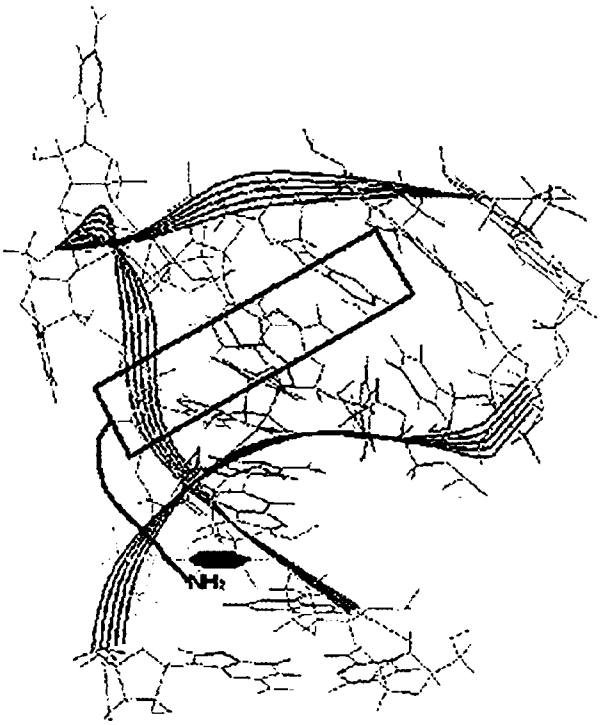

Figure 7. A schematic illustration to show a three-dimensional model of the HIV-1 Tat binding site of TAR RNA and the location of the amino terminus of Tat(42–72) mapped by psoralen–Tat(42–72). The TAR RNA structure is based on NMR data, and is similar to that proposed by Loret *et al.*,[61] Week and Crothers,[58] and Puglisi *et al.*[62] The ribbon structure of TAR RNA is shown in five dark lines. As proposed by Mujeeb *et al.*[59] for a hybrid Tat, the α-helical structure of the Tat(47–57) peptide backbone is shown in a barrel, and the N-terminal region containing Tat(42–46) is drawn as a line to represent a random coil. The α-helical region of the Tat(47–57) was positioned in the wide major groove of TAR RNA. The amino terminus of the peptide is labeled as NH_2, and the cross-linked uridine, U42, of TAR RNA is indicated in black. Structures of TAR RNA were visualized using Insight II software on an IRIS workstation.

(61) Loret, E. P.; Georgel, P.; Johnson, W. C.; Ho, S. P. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9734–9738.

(62) Puglisi, J. D.; Tan, R.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76–80.