Site-Specific Cross-Linking of Amino Acids in the Basic Region of Human Immunodeficiency Virus Type 1 Tat Peptide to Chemically Modified TAR RNA Duplexes

Mark A. Farrow,[‡] Fareed Aboul-ela,[‡] David Owen,[‡] Alexander Karpeisky,[§] Leonid Beigelman,[§] and Michael J. Gait*,[‡]

Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 2QH, U.K., and Department of Chemistry and Biochemistry, Ribozyme Pharmacueticals Inc., 2950 Wilderness Place, Boulder, Colorado 80301

Received October 30, 1997; Revised Manuscript Received December 26, 1997

ABSTRACT: The Human Immunodeficiency Virus type 1 Tat protein interacts specifically with a U-rich bulge within an RNA stem−loop known as the *trans*-activation responsive region (TAR) that occurs in all viral transcripts. We have photochemically cross-linked to Tat peptide (37−72), a model TAR duplex substituted at U_{23} in the bulge by 4-thioU. We have identified the cross-linked amino acid as Arg^{55} in the basic region of the Tat peptide by use of a combination of proteolytic digestions and MALDI−TOF mass spectrometric analysis. The identification also required use of a synthetic Tat peptide containing a site-specific, uniformly ^{13}C and ^{15}N isotopically labeled arginine. We also describe a new chemical procedure for obtaining site-specific cross-links to Tat via the use of 2′-*β*-alanyl U-substituted TAR and the amino-specific reagent dithiobis(succinimidyl propionate). U_{23}-2′-functionalized TAR was shown to cross-link uniquely to Lys^{51} in the basic region of Tat, whereas other sites in the upper and lower stems of TAR (U_{35}, U_{38}, and U_{42}) showed cross-linking only to the N-terminus of Tat peptide (37−72). U_{40} cross-linked to both Lys^{51} and the N-terminus of the peptide. The results help to establish a preliminary model of the binding of Tat peptide to the major groove of TAR RNA.

How proteins recognize RNAs is still poorly understood (for reviews see refs 1 and 2). One important class of RNA−protein interaction is that which occurs between highly basic domains of proteins and RNA duplexes containing bulged residues, as exemplified by the interaction of the Human Immunodeficiency Virus type 1 (HIV-1) *trans*-activator protein Tat with its RNA recognition sequence, the *trans*-activation responsive region TAR.[1] This interaction triggers a *trans*-activation process involving host cellular factors which results in alteration of the elongation properties of the nascent RNA polymerase transcription complex. Whereas in the absence of Tat HIV transcripts are prematurely terminated, they become full length when Tat levels rise (*3*).

The RNA recognition sequence TAR is a 59-residue RNA stem-loop which is found at the 5′-end of all HIV transcripts. Tat binds to TAR in vitro with high affinity (*4*−*7*). HIV-1 Tat is an 86−102 amino acid protein with varying C-terminus depending on the type of isolate. Only the first 72 residues are required for biological activity, and these can be subdivided into five regions (Figure 1a): an N-terminal section (I), a cysteine-rich region (II), a sequence-conserved hydrophobic core (III), a highly basic region that also contains a nuclear localization signal (IV), and a glutamine-rich sequence (V). The three-dimensional structure of HIV-1 Tat is not known, but preliminary NMR (*8*), molecular modeling (*9*), and spectroscopic studies (*10*) have suggested that the basic region has a flexible extended structure whereas the flanking core and glutamine-rich regions may form more pronounced structural domains that influence the overall folding of Tat.

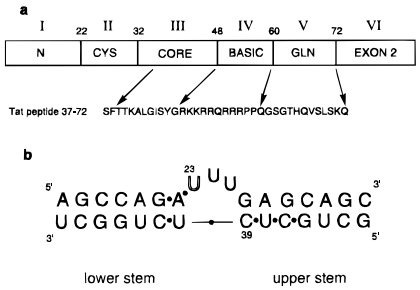


FIGURE 1: (a) Regions of the HIV-1 Tat protein and sequence of the Tat peptide (37−72). (b) Structure of the HIV-1 TAR synthetic model duplex containing the U-rich bulge. Filled circles denote phosphates in TAR which are seen to be protected in ethylation interference analysis or where substitution by methanephosphonate interferes with Tat binding (*20*−*22*). The outlined residue U_{23} is a site crucial to Tat recognition (*19, 20*) and which is substituted by 4-thioU (*36 and this work*).

Because of difficulties in obtaining Tat in properly folded form, most in vitro studies have centered on Tat peptides.

* To whom correspondence should be addressed. Email: mgait@mrc-lmb.cam.ac.uk.

‡ Medical Research Council.

§ Ribozyme Pharmaceuticals Inc.

[1] Abbreviations: EEDQ2, ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline; Fmoc, 9-fluorenylmethoxycarbonyl; HOBt, hydroxybenzotriazole; HPLC, high-performance liquid chromatography; MALDI−TOF, matrix-assisted laser desorption ionization/time-of-flight; PAGE, polyacrylamide gel electophoresis; TAR, *trans*-activation responsive element; TEA, triethylamine.

S0006-2960(97)02695-0 CCC: $15.00
Published on Web 02/14/1998

Peptides spanning the basic region of Tat (*11, 12*), or the basic and glutamine-rich regions (*13*−*15*), bind TAR with high affinity. However, to obtain the full specificity of the Tat−TAR interaction, a larger region of Tat is required that includes the hydrophobic core. For example, Tat peptide (37−72) includes most of the core and glutamine-rich regions in addition to the basic region (Figure 1a) and mimics in competition binding experiments more closely the ability of Tat protein to distinguish mutant TARs from wild-type than those peptides containing the basic region alone (*16, 17*).

The region of TAR recognized by Tat is near the apex and contains a three-residue U-rich bulge (*5, 6*). Mutation of U_{23} in the bulge, or either of the two base pairs above the bulge, leads to a drastic reduction in Tat binding affinity (*17, 18*). Since the apical loop is not required for Tat binding, it can be dispensed with in TAR models (*19, 20*). Such model TAR duplexes (e.g. Figure 1b) containing singly modified base residues have been used by us and others to identify a number of base functional groups essential to Tat binding. These are the N^7-positions of G_{26} and A_{27} in the two base pairs above the bulge and the O^4- and N^3−H-positions of U_{23} in the bulge (*19, 20*). In addition, ethylation interference (*20, 21*) and methylphosphonate substitution analysis (*22*) have led to the identification of a number of phosphate residues important to Tat binding, two on one strand (P21−22 and P22−23) and five on the other strand (P36−37 to P40−41) (Figure 1b). Some of these phosphates are likely candidates for direct interactions with Tat, possibly with basic residues such as arginine or lysine, which might account for the large electrostatic contribution to binding energy (*23*).

The major groove of TAR in the region of the U-rich bulge is much wider than for normal A-type RNA helixes, as shown by recent NMR studies of free TAR (*24*). In the presence of Tat peptide (37−72) there is a dramatic conformational change resulting in considerable narrowing of the major groove (*25, 26*). The base and phosphate functional groups on TAR identified by chemical substitution experiments to be important to Tat binding become clustered close to the surface of the duplex on the major groove side and are available for interaction with Tat. However, the lack of a defined structural element in free Tat peptide (37−72) and a number of other difficulties have meant that it has not been possible so far to produce an unequivocal structural model of the complex from the NMR data alone, even with the use of isotopic labeling of RNA and peptide (Aboul-ela and Varani, personal communication).

To obtain information about the proximities of specific amino acids in Tat to particular RNA residues in TAR, there have been a number of recent reports where Tat peptides modified at unique positions have been cross-linked to TAR and the locations of cross-linking on the RNA determined. An N-terminal psoralen conjugate of Tat peptide 42−72 was found to bind specifically to TAR and the psoralen photochemically cross-linked to U_{42} in the lower stem (*27*). Another psoralen conjugate of Tat (38−72) to a unique Cys^{57} residue, introduced synthetically to replace Arg^{57}, cross-linked to the TAR apical loop at position U_{31} (*28*). In addition, Liu et al. showed that Tyr^{47} in the core region of Tat peptide 42−72 could be specifically UV-cross-linked to G_{26} (*29*). These types of experiment have suggested that the orientation of the basic region of Tat lies from N to C in the direction from lower TAR stem to upper TAR stem and loop.

A complementary and potentially more powerful approach is to introduce a specific modification on the TAR RNA and to cross-link to an unmodified Tat peptide. Advances in RNA chemistry are now such that it is possible for a range of RNA positions to be functionalized and cross-linkable groups attached to see which RNA modifications are tolerated by the peptide without significant loss of RNA binding. Recently, we developed a new procedure for the incorporation of trisubstituted pyrophosphate (tsp) analogues to replace individual phosphate residues in TAR RNA model duplexes and showed that such substitution did not interfere significantly with Tat−TAR recognition (*30*). Only phosphates in TAR expected to be in close proximity to Tat were sites where tsps reacted readily with Tat peptide (37−72). We found that TAR containing a tsp at site P38−39 in the upper stem reacted covalently with the epsilon amino group of a lysine residue in the basic region of Tat and identified this residue as Lys^{51} by MALDI−TOF mass spectrometry and enzymatic digestion methods. This was the first example of identification of a cross-linked amino acid in Tat to a TAR analogue.

A better known RNA−protein cross-linking strategy involves the introduction of 4-thioU and other thiolated bases into RNA (*31*). Such bases may be photochemically cross-linked to other RNA sites (for example within ribozymes (*32, 33*)) or to proteins (*34, 35*). In the case of the Tat−TAR interaction, a TAR duplex containing 4-thioU at position U_{23} has been photochemically cross-linked to Tat peptide (38−72) (*36*). The authors were unable to identify the precise position of cross-linking on the peptide by conventional peptide sequencing techniques, although there were indications that the cross-link might be localized to the basic domain of Tat.

We now report that we have been able to identify a unique amino acid (Arg^{55}) in the basic region of Tat that is cross-linked to TAR substituted at U_{23} by 4-thioU. The identification was made by use of a combination of proteolytic digestions and MALDI−TOF mass spectrometric analysis of a cross-linked Tat peptide as well as of a Tat peptide containing a uniformly ^{13}C and ^{15}N isotopically labeled arginine. In addition, we have developed a new chemical procedure for obtaining site-specific cross-links to Tat via the use of 2′-*β*-alanyl U-substituted TAR and the amine-specific reagent dithiobis(succinimidyl propionate). U_{23}-2′-functionalized TAR was shown to cross-link uniquely to Lys^{51} in the basic region of Tat, whereas other sites (U_{35}, U_{38}, and U_{42}) showed cross-linking only to the N-terminus of Tat peptide (37−72). U_{40} cross-linked to both Lys^{51} and the N-terminus of the peptide. The results help to establish a preliminary model of the binding of Tat peptide to the major groove of TAR.

MATERIALS AND METHODS

Anhydrous solvents and reagents were from Aldrich unless otherwise stated and were used without purification. NMR spectra were acquired on a Varian Gemini-400 spectrometer equipped with a Varian 5 mm multinuclear probe. ^1H NMR spectra were referenced to the internal $CHCl_3$ signal in the samples $\delta = 7.24$ ppm. ^{31}P NMR chemical shifts were referenced to an external 85% H_3PO_4 standard, $\delta = 0$ ppm.

Synthesis of Tat Peptide (37−72). The peptide sequence SFTTKALGISYGRKKRRQRRRPPQSGTHQVSLSKQ (Cys[37] was replaced by Ser to avoid problems with disulfide formation) was prepared by continuous-flow Fmoc-polyamide solid-phase synthesis using *N*-(2-hydroxy-4-methoxybenzyl) (Hmb) backbone protection for the addition of Gly[44] (*37*−*39*).

Synthesis of Fmoc U[13]C6, U[15]N4 Arginine. H-L-U[13]C6 (98%) U[15]N4 (97%) arginine·HCl (Cambridge Isotope Laboratories; 249 mg, 1.13 mmol) was dissolved with stirring in 10% aqueous (w/v) sodium carbonate solution (4.8 mL) and dioxane (2.4 mL). After cooling on ice, 9-fluorenylmethylchloroformate (310 mg, 1.2 mmol) dissolved in dioxane (5 mL) was added in portions over 30 min. The mixture was stirred at room temperature for 2 h. Water (20 mL) was added dropwise and the mixture cooled to 4 °C overnight. The precipitated product was isolated by filtration and washing with cold water until the washings had a pH of 5−6. The product was dried over potassium hydroxide pellets. Yield: 302 mg (65.8%). Reversed phase HPLC analysis (Aquapore 300 C8, monitoring at 215 nm, gradient of 10−90% buffer B, 25 min: buffer A is 0.1% TFA solution; buffer B is 0.1% TFA/acetonitrile (1:9)) showed approximately 80% in the main peak.

Synthesis of U[13]C6, U[15]N4-Labeled Arg[52] Tat Peptide (37−72). This was prepared as for the unlabeled Tat peptide except that Arg[52] was coupled off column for 16 h using a 5-fold excess of Fmoc U[13]C6, U[15]N4 arginine as its hydroxybenzotriazolyl (HOBt) ester generated in situ by reaction with equimolar amounts of PyBOP and diisopropylethylamine and 2 equiv of HOBt in *N*,*N*-dimethylformamide (DMF). All subsequent couplings were carried out for 1 h using a 20-fold excess of Fmoc amino acid Pfp ester (Novabiochem) in the presence of 2 equiv of HOBt in DMF except Gly[44] which was coupled for 3 h using a 20-fold excess of bisFmoc(Hmb)−Gly OPfp (Novabiochem) and 2 equiv of HOBt in DMF and Leu[43], which was double coupled for 2 h using a 20-fold excess of Fmoc Leu−OPfp and 2 equiv of HOBt in DMF.

*2′-Deoxy-2′-(N−Fmoc-β-alanyl)aminouridine (Compound **2**).* EEDQ (4.2 g, 17 mmol) was added to a solution of 2′-amino-2′-deoxyuridine (Dalton Chemical Laboratories, Inc.; 4 g, 16.45 mmol) and *N*-Fmoc-β-alanine (5.1 g, 16.45 mmol) in methanol, and the reaction mixture was boiled for 2 h. Subsequent flash chromatography on silica using a linear gradient of methanol (5−10%) in dichloromethane afforded 6 g (77%) of compound **2**. ^{1}H NMR ($CDCl_3$−DMSO-d_6): δ 10.40 (s, 1H, N3−H), 6.98−7.63 (m, 6-H, Fmoc), 6.16 (t, 1H, NHFmoc), 5.73 (d, 1H, 1′-H, $J_{1',2'}$ = 8.4), 5.34 (d, 1H, 5-H), 4.22 (m, 1H, 2′-H), 3.98 (dd, 2H, CH_2 from Fmoc), 3.88 (m, 1H, 3′-H), 3.77 (br s, 1H, 4′-H), 3.43 (m, 2H, 5′-CH_2), 3.1 (m, 2H, CH_2NHFmoc), 2.07 (t, 2H, CH_2CO).

*5′-O-Dimethoxytrityl-2′-deoxy-2′-(N-Fmoc-β-alanyl)aminouridine (Compound **3**).* To a solution of compound **2** (1.5 g, 2.8 mmol) in anhydrous pyridine was added dimethoxytrityl chloride (1.23 g, 3.6 mmol), and the reaction mixture was stirred overnight at room temperature. Pyridine was removed in vacuo, and the residue was dissolved in dichloromethane and washed with saturated aqueous sodium bicarbonate solution and then brine. The organic layer was dried over sodium sulfate, filtered, and evaporated to dryness. Flash silica gel chromatography of the residue in 2% methanol in dichloromethane afforded 1.73 g (74%) of compound **3**. ^{1}H NMR (DMSO-d_6): δ 11.40 (d, 1H, N3−H), 8.01 (d, 1H, 2′-N−H, $J_{2'-NH}$ 8.3), 6.96−7.97 (m, Fmoc, DMT, NHFmoc), 7.71 (d, 1H, 6-H, $J_{6,5}$ = 8.12), 5.96 (d, 1H, 1′-H, $J_{1',2'}$ = 7.8), 5.76 (bs, 1H, 3′-OH), 5.49 (dd, 1H, 5-H, J_{5-NH} 2.08, $J_{5,6}$ = 8.12), 4.69 (m, 1H, 2′-H), 4.26−4.36 (m, 3H, 3′-H, CH_2 from Fmoc), 4.10 (m, 1H, 4′-H), 3.81 (s, 6H, 2OCH_3), 3.25−3.41 (m, 4H, 5′-CH_2, CH_2NHFmoc), 2.43 (m, 2H, CH_2CO).

*5′-(O-Dimethoxytrityl)-2′-deoxy-2′-(N-Fmoc-β-alanyl)aminouridine 3′-O-(2-Cyanoethyl-N,N,-diisopropyl)phosphoramidite (β-AlaU amidite) (Compound **4**).* Compound **3** (1.4 g, 1.67 mmol) was dissolved in dry dichloromethane under dry argon. The solution was ice-cooled and *N*,*N*-diisopropylethylamine (0.63 mL, 3.62 mmol) and 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoroamidite (0.61 mL, 2.02 mmol) were simultaneously added dropwise to a stirred reaction solution. Stirring was continued for 5 h at room temperature. The reaction mixture was again ice-cooled and quenched with dry methanol (12 mL). After 5 min of stirring, the mixture was concentrated in vacuo (40 °C) and the product purified by flash chromatography on silica gel using hexane−acetone (3:2) with 1% triethylamine mixture as an eluent.

Appropriate fractions were collected and evaporated in vacuo to give 1.2 g (69%) of phosphoramidite (**4**) as a white foam. ^{31}P NMR ($CDCl_3$): δ 151.80, 150.23.

Synthesis of Oligoribonucleotides. Oligoribonucleotides were synthesized on a 1 μmol scale using a controlled glass support (CPG) by standard phosphoramidite synthesis with a PE Applied Biosystems 394 DNA/RNA Synthesizer. 5′-O-dimethoxytrityl-2′-*O*-(*tert*-butyldimethylsilyl)nucleoside 3′-*O*-(2-cyanoethyl-*N*,*N*,-diisopropyl)phosphoramidites of U, N^6-phenoxyacetyl (PAC) A, N^2-isopropylPAC G, and N^4-benzoyl C (Glen Research via Cambio) were used in the coupling reactions. Site-specifically modified oligoribonucleotides were prepared using a 2′-*O*-Fpmp protected 4-thiouridine phosphoramidite obtained from Cruachem or the β-AlaU amidite described above. Deprotection and workup of oligoribonucleotides were essentially as described (*30*, *40*) except for those containing 4-thiouridine, which were treated with 0.3 M 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in acetonitrile for 1 h prior to ammonia deprotection, similarly to the method previously described (*35*). Then, following silyl deprotection using TEA·3HF−DMF (3:1) (*40*−*42*), oligoribonucleotides were precipitated with butan-1-ol and the single 2′-*O*-Fpmp group removed by addition of TEA·3HF−water (1:9 (v/v)) for 1 h (*42*). Integrities and purities of oligoribonucleotides were confirmed by both base composition analysis and MALDI−TOF mass spectroscopy (*42*).

Preparation of ^{32}P-Labeled TAR RNA Duplex. Approximately 500 pmol of the oligoribonucleotide TAR strand containing the modification for cross-linking was 5′-^{32}P end-labeled by reaction with γ-^{32}P-ATP and T4 polynucleotide kinase as previously described (*43*). The reaction mixture was then heated to 90 °C for 2 min, and a 1.2-fold molar excess of the second (unmodified) TAR RNA strand was added. The solution was allowed to cool to 4 °C over 30 min, and the remaining nucleotides were removed by gel filtration through a Sephadex NAP 10 column (Pharmacia) preequilibrated with ice-cold water.

Filter Binding Assays. Binding reactions (*42*) (250 μL) contained 2 nM ^{32}P-labeled TAR RNA duplex, 0.4 μg mL^{-1} tRNA, 1 μg mL^{-1}calf thymus DNA in 1 × TK buffer (50 mM Tris−HCl, pH 7.4, 20 mM KCl), and 0−5000 nM Tat peptide (37−72). Reactions were incubated on ice for 60 min in individual wells of a prewetted 96 well Multiscreen HA mixed cellulose ester plate (Millipore) and filtered simultaneously using a vacuum manifold (Millipore). Each filter well was washed with 200 μL of ice-cold 1 × TK buffer and dried and the filter punched into scintillation vials with the amount of radioactivity retained on each filter being determined by liquid scintillation counting.

Gel Mobility Shift Assays. Varying concentrations (0−160 nM) of Tat peptide (37−72) were incubated with TAR RNA duplex (8 nM) in 20 μL of 1 × TK, 0.01% Trition X-100. The solution was incubated on ice for 30 min prior to the addition of 5 μL of 40% sucrose, 0.1% bromophenol blue. The mixtures were then applied to a preequilibrated (4 °C) 8% native polyacrylamide gel. Electrophoresis was for 1 h at 4 °C in 0.5 × TB (50 mM TrisHCl, 50 mM boric acid, pH 8.3), 0.01% Triton X-100 at 25 mA. The gel was dried down under vacuum and the bands visualized by autoradiography.

4-Thiouridine Cross-Linking Reactions. A 100 μL aliquot of a solution containing 15 μg of 17-mer RNA containing 4-thiouridine at position 23 ($^{4S}U_{23}$) was annealed to a 1.2-fold molar excess of 14-mer RNA by heating to 70 °C for 2 min in 1 × TK and allowed to cool slowly to room temperature. The solution was then chilled on ice prior to the addition of a 2-fold molar excess of Tat peptide (37−72). The solution was incubated for a further 20 min before being transferred to a microtiter plate (on ice) and irradiated for 5 min at 366 nm using a hand held UV lamp. The product was transferred to a microfuge tube, trifluroacetic acid (TFA) added to 1% and injected onto an HPLC column (Brownlee SPHERI-5 RP 8 220 × 4.6 mm) which was eluted with a linear gradient of 20−50% B over 45 min where buffer A was 0.06% (v/v) TFA in water and buffer B was 0.04% (v/v) TFA in acetonitrile/water (9:1 (v/v)). The flow rate was 0.7 mL min^{-1}. Detection of peaks was by means of a dual wavelength detector (Gilson) at 218 and 260 nm. Products were collected and evaporated immediately using a SpeedVac (Savant) and the dried material redissolved in 0.1 M ammonium hydrogen carbonate solution prior to enzymatic digestion or mass spectral analysis.

4-Thiouridine Cross-Linking Reactions to Amino Acids. A 2.5 μg sample of $^{4S}U_{23}$ RNA duplex was prepared in 10 μL of 0.5 × TK buffer. Lysine, glutamine, or argininamide was added to a final concentration of 5 mM and the solution incubated for 20 min at 4 °C. The solution was transferred to a microtiter plate (on ice) and irradiated for 10 min prior to desalting by drop dialysis and mass spectral analysis.

Small-Scale Cross-Linking Reactions for β-AlaU-Containing TAR Duplexes. A 50 pmol (25 μL) amount of ^{32}P-labeled oligoribonucleotide containing β-alaU was added to 25 μL of 0.2 M sodium phosphate solution (pH 8.0). A 2.5 μL aliquot of a freshly prepared solution of dithiobis(succinimidyl propionate) (DSP, Pierce, approximately 10 mg mL^{-1} in *N*,*N*-dimethylformamide) was added and the reaction incubated at room temperature for 10 min. The reaction was applied directly to a Sephadex NAP10 column preequilibrated with ice-cold water. A 250 μL fraction containing the radiolabeled RNA was collected, potassium chloride was added to a final concentration of 20 mM and the RNA annealed to a 1.2 molar excess of the second (unmodified) TAR strand by incubating at room temperature for 10 min. The solution was then cooled on ice for 5 min and HEPES (pH 7.5) added to a final concentration of 50 mM; 20 μL aliquots of the solution were taken and Tat peptide (37−72) added to a final concentration of 320 nM. The cross-linking reaction was allowed to proceed for 10 min before the addition of 10 μL of loading buffer (7 M urea in 1 × TBE). The solution was then applied to a thin (0.5 mm) 20% denaturing polyacrylamide (19:1 acrylamide−bisacrylamide) gel and electrophoresis carried out for 1.5 h at 18 mA. The gel was dried down under vacuum and the bands visualized by autoradiography. Quantification of the bands was by phosphorimager (Molecular Dynamics).

Large-Scale Cross-Linking Reactions for β-AlaU-Containing TAR Duplexes. A 50 μL solution containing 15 μg of RNA in 0.1 M sodium phosphate (pH 8.0) was prepared. To this 2.5 μL of DSP (10 mg/mL in DMF) was added and the reaction incubated at room temperature for 10 min. Excess cross-linking reagent was removed by gel filtration using a Sephadex NAP10 column preequilibrated with 2 mM sodium citrate (pH 5.5). A 500 μL fraction was collected and concentrated using a centrifugal concentrator (Nanosep, Pall Filtron). The functionalized RNA strand was annealed with a 1.2-fold molar excess of the second (unmodified) TAR RNA strand for 10 min at room temperature in 20 mM KCl. The solution was then cooled to 4 °C and HEPES (pH 7.5) added to a final concentration of 50 mM. A 10 μg sample of Tat peptide (37−72) was added and the solution incubated for a further 10 min. Trifluroacetic acid (TFA) was added to a final concentration of 1% (v/v), and the cross-linked species was immediately loaded onto a reversed phase HPLC column. Elution and workup were identical to those described above for 4-thioU cross-linking.

Enzyme Digestions of Cross-Linked Products. Chymotrypsin, trypsin, endoproteinase LysC (sequencing grades), and Pronase were obtained from Boehringer Mannheim. Digestions were carried out in 0.1 M ammonium hydrogen carbonate for between 0 and 18 h. Pronase was used at a final concentration of 0.1 mg mL^{-1}, and other proteases were used at a final concentration of 0.02 mg mL^{-1}.

MALDI−TOF Mass Spectrometry. Samples were prepared 1:1 with a solution of 3-hydroxypicolinic acid (50 mg mL^{-1}) in 60:40 acetonitrile−water containing diammonium hydrogen citrate (8 mg mL^{-1}). We have found these matrix conditions to be particularly suitable for oligonucleotides and oligonucleotide−peptide fragments, but in many cases free peptides released by degrading enzymes are not seen. Where necessary, samples were desalted by drop dialysis against water using Millipore VS 0.025 μm filters. Calibration was by means of either an internal RNA standard or external peptide standards (PerSeptive Biosystems). Spectra were recorded using a Voyager DE mass spectrometer (PerSeptive Biosystems) in positive ion mode using an accelerating voltage of 25 000 V and a pulse delay time of 450 ns.

RESULTS

Synthesis of 4-thioU-Containing TAR Analogues and Cross-Linking to Tat Peptide (37−72). Wang and Rana were

able to incorporate 4-thioU into a TAR model duplex (*36*), where 4-thioU-containing oligoribonucleotides were prepared either by transcription techniques or by chemical synthesis using a convertible nucleoside approach (*44*). They showed that a TAR model duplex containing 4-thioU at position 23 became cross-linked to U_{40} when irradiated at 360 nm in the absence of Tat, but in the presence of Tat peptide (38−72) there were only RNA−peptide cross-links and no RNA−RNA cross-links. Although there were indications of localization to the basic domain, the precise site(s) of cross-linking on Tat could not be identified, however. Since position 23 in TAR is crucial to Tat recognition (*17, 18*) and since methylation at either N^3 (*19*) or O^4 (*20*) drastically interferes with Tat binding, we thought that identification of the site(s) in Tat to which the 4-thioU_{23} analogue cross-linked would be particularly revealing from a structural standpoint. We therefore decided to investigate alternative techniques for such identification.

We first prepared a TAR duplex containing 4-thioU at position 23 of TAR (Figure 1b). The sequence of the TAR duplex is identical to that previously used by us for functional group substitution analysis (*20*) and for our previous RNA−protein cross-linking studies (*30*). Our procedure for synthesis of 17-mer oligoribonucleotide containing 4-thioU differs only slightly from previously published methods (*32*, *35*) in that we have used a commercially available 2′-Fpmp-protected 4-thioU amidite together with standard 2′-silyl-protected phosphoramidites for the oligoribonucleotide assembly (*42*), which we found to result in good yields and reliable incorporation of the 4-thioU residue.

We then confirmed by mobility shift analysis that ^{32}P-labeled 4-thioU_{23}-containing TAR duplex bound Tat peptide (37−72) equally as well as unmodified TAR duplex did (data not shown). This result was expected following the previous report of unaltered binding characteristics for an almost identical 4-thioU-containing TAR duplex to Tat peptide (38−72) (*36*). Upon UV irradiation at 366 nm, an RNA−protein cross-linked species was found as a single delayed radioactive band by denaturing polyacrylamide gel electophoresis (PAGE) in >95% yield (data not shown), again very similarly to that reported by Wang and Rana (*36*). Since our experience with other Tat/TAR cross-linking methods had shown that cross-linked peptides were poorly resolved from free Tat peptide by denaturing PAGE (*30*), we used reversed-phase HPLC to purify the cross-linked TAR duplex, which was well-resolved from free Tat−peptide (37−72) and un-cross-linked TAR RNA duplex (Figure 2). We characterized the cross-linked product by MALDI−TOF mass spectrometry using an external mass standard (Figure 3a). Of the two major peaks, one corresponded to an unmodified 14-mer TAR strand (*m*/*z*: found, 4376.4; calculated, 4376.8); the second had a molecular mass of 9600.1 compared to a calculated value of 9636.3 for the sum of the molecular masses of Tat peptide (37−72) and the 4-thioU-containing 17-mer TAR RNA strand. It was thus assumed that during photochemical cross-linking approximately 36 mass units had been lost in the reaction.

Identification of Site of Cross-Linking on Tat Peptide (37−72). We were previously able to identify a site of Tat peptide cross-linking to a tsp-substituted TAR duplex using proteolytic digestion of the product of cross-linking followed by mass spectral analysis (*30*). We therefore subjected the

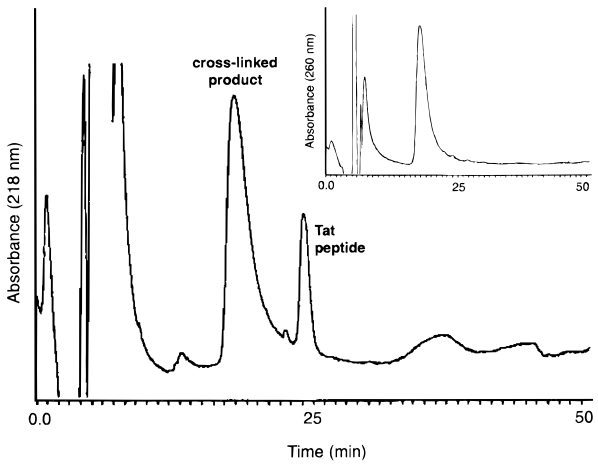


FIGURE 2: Reversed phase HPLC chromatogram recorded at 218 nm (inset: 260 nm) showing the purification of Tat peptide (37−72) cross-linked to TAR RNA substituted at position U_{23} by 4-thioU.

cross-linked and HPLC-purified 4-thioU TAR−Tat peptide (37−72) to digestion with trypsin, which was expected to cleave after lysine and arginine residues, and analyzed the products by mass spectrometry after various time intervals of digestion. In addition to the unmodified 14-mer TAR strand (which was used for this and subsequent analyses as an internal mass standard), two major peaks were found (Figure 3b and Table 1). The peak at *m*/*z* 6303.7 was unequivocally assigned as the 4-thioU 17-mer linked to Tat peptide 50−55 (KKRRQR). The second peak at *m*/*z* 6047.7 corresponds to 4-thioU 17-mer linked to peptide 52−55 (RRQR), 53−56 (RQRR), or 54−57 (QRRR). There was no significant alteration in the pattern following more prolonged trypsin digestion. When Pronase, which can potentially cleave after any residue, was added to the trypsin digestion, both peaks at 6303.7 and 6047.7 were lost and there was a new peak at *m*/*z* 5889.2 (Table 1). This could correspond to any or all of the fragments of 4-thioU 17-mer linked to 51−53 (KRR), 52−54 (RRQ), 53−55 (RQR), or 54−56 (QRR) which could be derived from the two larger mass peaks found by trypsin digestion alone. Digestion of the cross-linked product with Pronase alone produced a pronounced peak at 6176.0 which could be unequivocally assigned as 4-thioU 17-mer linked to Tat peptide 51−55 (KRRQR) (Table 1). Other peaks were seen at *m*/*z* 5764.7 (4-thioU 17-mer linked to 55−56 or 52−53 (RR)) and 5734.9 (4-thioU 17-mer linked to 54−55 (QR) or 53−54 (RQ) or 51−52 (KR)).

Because of the identical theoretical molecular masses of several of these proteolytic digestion fragments, it was clear that additional information would be needed to locate precisely the cross-linked amino acid within fragment 51−55 (KRRQR). We therefore chemically synthesized a Tat peptide (37−72) where one of the arginine residues (Arg^{52}) was uniformly ^{13}C and ^{15}N isotopically labeled. Apart from an otherwise standard Fmoc solid-phase peptide synthesis involving 2-hydroxymethylbenzoyl backbone protection of Gly^{44} (*37−39*), the coupling of $U^{13}C6$, $U^{15}N4$ Arg^{52} was carried out by preparation of labeled Fmoc Arg and coupling where the side chain was protonated but otherwise unprotected (*45, 46*). Cross-linking to the 4-thioU_{23}-containing TAR duplex and HPLC purification of the cross-linked

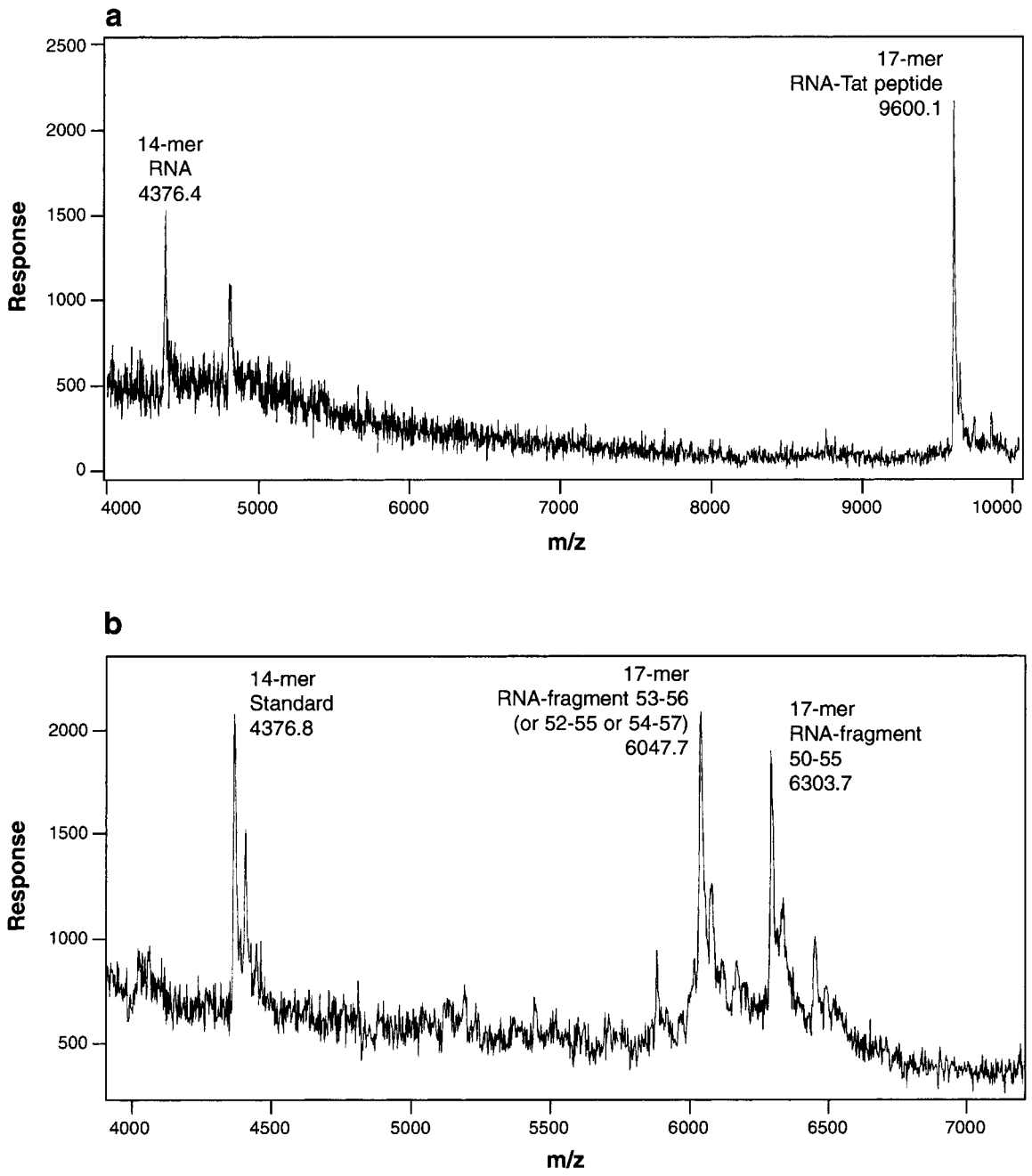


FIGURE 3: MALDI−TOF mass spectra: (a) HPLC-purified Tat peptide (37−72) cross-linked to TAR RNA substituted at U_{23} with 4-thioU; (b) trypsin digestion product of cross-linked species.

product was carried out as before. MALDI−TOF mass spectometry showed a peak at *m/z* 9609.4 corresponding to 4-thioU 17-mer cross-linked to the labeled Tat peptide (Table 1) and was approximately 9 mass units greater than the unlabeled cross-linked product, within experimental error the same as expected (10 units). The mass spectrum of the tryptic digestion of the cross-linked product showed three peaks. The peak at 6312.8 was assigned as 4-thioU 17-mer linked to labeled peptide 50−55 (9−10 mass units heavier than the corresponding unlabeled fragment). The second peak at 6047.2 was unchanged in mass from the corresponding unlabeled digestion and therefore could only involve fragment 53−56 (RQRR) or 54−57(QRRR). Coupled with previous information, these results localized the cross-link to residue, R^{53}, Q^{54}, or R^{55}.

Addition of Pronase to the trypsin digestion led to loss of both fragments at 6312.8 and 6047.2, and a third peak at *m/z* 5890.0 was assignable to either fragments involving 53−55 (RQR) or 54−56 (QRR) (Table 1). Pronase digestion alone gave rise to 4-thioU 17-mer-linked fragments at *m/z* 6182.5 (51−55, K*RRQR), 5762.0 (RR unlabeled, assignable to residues 55−56 only), and 5733.3 (QR unlabeled, assignable to residues 54−55 only). These results can only be rationalized if the cross-link is to residue Arg^{55}.

To confirm cross-linking reactivity of amino acid side chains, we subjected the 4-thioU TAR to UV irradiation at 366 nm in the presence of 5 mM glutamine, 5 mM lysine, or 5 mM argininamide. There was no cross-linked product with either glutamine or lysine. By contrast argininamide formed two products with *m/z* 5642.6 and 5605.5 (data not

Table 1: Mass Values for 4-thioU TAR Oligonucleotides Cross-Linked to Tat Peptide (37−72) or *R52 Tat Peptide (37−72) and Their Proteolytic Digestion Products

	Tat peptide (37−72)		*R52 Tat peptide (37−72)	
	mass	assignment	mass	assignment
undigested	9600.7	37−72	9609.4	37−72
Trypsin	6303.7	50−55 KKRRQR	6312.8	50−55
	(6301.9)[a]		(6311.9)	KK*RRQR
	6047.7	53−56, RQRR	6047.2	
	(6045.6)	(52−55, RRQR; 54−57, QRRR)[b]	(6045.6)	53−56, RQRR (54−57, QRRR)
Trypsin followed by Pronase	5889.2	53−55, RQR	5890.0	53−55, RQR
	(5889.4)	(51−53, KRR; 52−54, RRQ; 54−56, QRR)	(5889.4)	(54−56, QRR)
Pronase alone	6176.0	51−55, KRRQR	6182.5	51−55, K*RRQR
	(6173.7)		(6183.7)	
	5764.7	55−56, RR	5762.0	55−56, RR
	(5761.2)	(52−53, RR)	(5761.2)	
	5734.9	54−55, QR	5733.3	54−55, QR
	(5733.2)	(53−54, RQ; 51−52, KR)	(5733.2)	

[a] The expected mass is given in parentheses. The calculation is based upon internally calibrated *m*/*z* of 9600.7 being the mass for the Tat peptide (37−72) cross-linked to 4-thioU 17-mer RNA, being 36 units less than the sum of the components before cross-linking. [b] Assignments in parentheses are formal possibilities which become eliminated as a result of subsequent digestion data.

shown). The first peak is consistent with an adduct of argininamide with 4-thioU TAR 17-mer (*m*/*z* 5639.7), whereas the second peak is 37 mass units lighter. Since this value is within experimental error the same as that found for loss of mass on cross-linking of Tat peptide (37−72) to 4-thioU TAR (36 units), it is very likely that the same photochemical cross-linking reaction took place with the side chain of Arg^{55} in Tat peptide (37−72) as with argininamide.

We also prepared TAR duplexes containing 4-thioU in place of U_{38} or U_{40} and irradiated them at 366 nm in the presence of Tat peptide (37−72). However, we found that the yields of peptide−RNA cross-linked product were rather low and complicated by the presence of RNA−RNA cross-links, as had been found previously (*36*). We have so far been unable to isolate sufficient cross-linked material to be able to identify the sites of cross-linking on the Tat peptide in these cases.

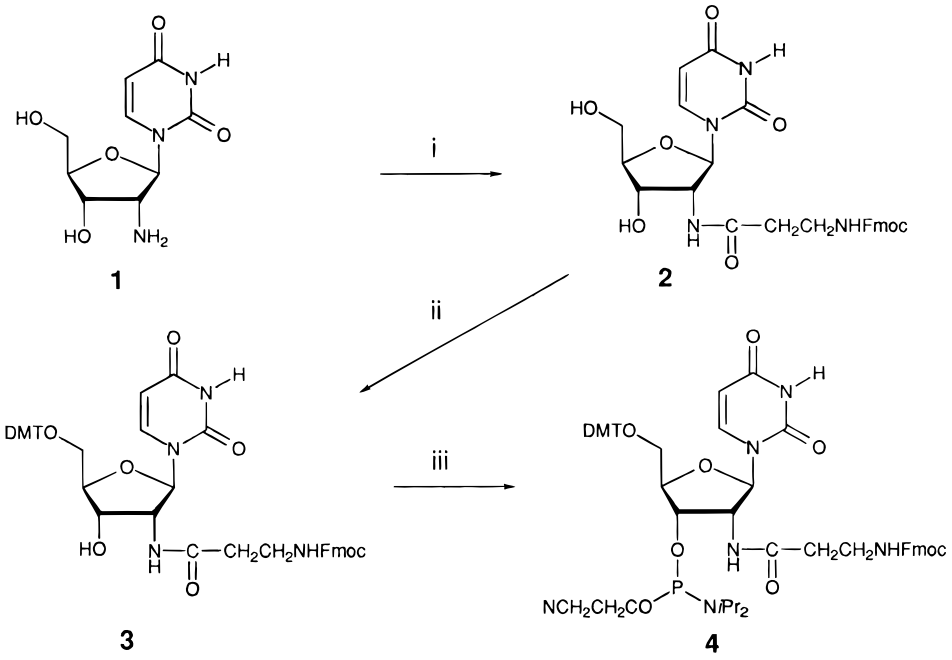
Synthesis and Tat Binding Studies of TAR Duplexes Containing 2′-β-alaU Residues. Only certain types of base analogues in RNA, particularly halo or thio derivatives of U or G, are amenable to high-wavelength photochemical cross-linking to proteins and peptides. Further, our experience described above was that good cross-linking yields are restricted to favorable cases where a conformationally and chemically amenable amino acid residue is able to easily approach the modified base. We wished to establish a more general procedure for obtaining cross-links between Tat and a variety of locations within TAR under mild, nonphotochemical conditions. In addition, we wished to use a longer and more flexible tether on the RNA to explore a wider spatial interface between RNA and protein.

The 2′-position of ribose moieties is a very suitable location for such tethering. To date very few 2′-positions of RNAs have been found to be crucial for recognition by a protein. Moreover, in regular RNA duplexes, the 2′-positions of ribose units are found on the outside of the helix and therefore it should be possible to use a tether of sufficient length to reach into both major or minor grooves from any particular 2′-position. Finally, protein−RNA cross-linking via any residue in a synthetic RNA could be obtained through incorporation of an appropriately 2′-functionalized nucleoside derivative. In the case of the Tat−TAR interaction it has been shown that replacement of an individual ribose unit within the bulge region of TAR by a 2′-deoxy residue in chemical substitution experiments results in little or no detriment to Tat binding (*20*, *22*). Further it was shown recently that an N3′ → P5′ phosphoramidate DNA analogue of a TAR duplex, expected to have an "RNA-like" conformation, was able to bind a Tat peptide with high affinity even though all 2′-hydroxyl groups were absent (*47*).

We therefore chemically synthesized a new 2′-deoxy-2′-β-alanylaminouridine (β-alaU) phosphoramidite derivative that has a flexible 2′-amino-containing arm and which is suitable for use in standard solid-phase oligonucleotide synthesis. The β-alanyl functionality is N-protected by the fluorenylmethoxycarbonyl (Fmoc) which is removed at the end of synthesis during ammonia deprotection. Briefly, the derivative was prepared in three steps from 2′-amino-2′-deoxyuridine (*48*) by coupling with Fmoc−β-alanine in the presence of EEDQ, similarly to the preparation of other 2′-amino acid functionalized uridines (*49*), followed by 5′-*O*-dimethoxytritylation and 3′-*O*-phosphitylation under standard conditions (Scheme 1).

We chemically synthesized oligoribonucleotides corresponding to the TAR RNA 14-mer and 17-mer (Figure 1b) where individual position U_{23}, U_{35}, U_{38}, U_{40}, or U_{42} had been replaced by the β-alaU derivative. We then tested the ability of such substituted TAR duplexes to bind Tat peptide (37−72). In a filter binding assay (Figure 4a), where increasing concentrations of Tat peptide are added to TAR duplex in the presence of competitor tRNA and DNA, all of the TAR duplexes containing the β-alaU analogue were bound by Tat peptide with apparent K_d within a 2-fold range of unmodified TAR duplex. This result suggests that a β-alaU modification does not affect Tat recognition significantly. In mobility shift analysis in the absence of competitors, all TAR analogues were able to shift the RNA duplex under the same conditions as those of the umodified TAR duplex except for the U_{40} β-ala TAR derivative (Figure 4b). We speculate that perhaps the dissociation rate of the complex is unusually high in this case. This would not be evident in filter binding since the complex is rapidly immobilized on a filter at 0 °C, whereas with mobility shift analysis there is a much longer period

Scheme 1



Reagents & conditions: i) N-Fmoc-β-alanine, EEDQ/ MeOH; ii) DMT-Cl/Py; iii) phosphitylation

during gel loading and during electrophoresis when the complex might dissociate.

Chemical Cross-Linking of 2′-Functionalized TAR Duplexes to Tat Peptide (37−72). Since there are several lysine residues in Tat peptide (37−72), two of which are within the basic domain and one of which has already been cross-linked to a tsp analogue of TAR (*30*), we investigated a number of linear, amino-specific homobifunctional cross-linking reagents. Of these we found that dithiobis(succinimidyl propionate) (DSP) (*50*) gave the best results and has the advantage that cross-linked products are cleavable by reduction with dithiothreitol. We obtained good cross-linking by use of DSP in a two-step process (Scheme 2). Thus, DSP reacted smoothly at pH 8 with each ^{32}P-labeled β-AlaU-substituted 14-mer or 17-mer oligoribonucleotide analogue at positions 23, 35, 38, 40, and 42. After purification by gel filtration, each product was annealed to the second unmodified TAR strand and incubated with a 10-fold excess of Tat peptide (37−72) at pH 7.5 at 4 °C for 10 min. Analysis of the products by denaturing PAGE (Figure 5) showed that in each case a product with delayed mobility was formed (lanes 4, 6, 8, 10, and 12). There was no cross-linked product (e.g. RNA−RNA) formed in the absence of Tat peptide (lanes 3, 5, 7, 9, and 11). There were also no cross-linked products formed in control reactions with unmodified TAR treated with DSP in the presence or absence of Tat peptide (37−72) (lanes 1 and 2). Yields of cross-linked products were about 50% in the case of the β-AlaU$_{23}$ TAR derivative and about 10% in the other four cases. When one-step cross-linking was attempted by addition of bifunctional cross-linker to β-ala TAR duplex in the presence of Tat peptide (37−72), very poor cross-linking yields were obtained (data not shown). This may be due to rapid DSP reaction with lysines in Tat peptide as well as the modified TAR duplex preventing cross-linking from taking place.

To characterize the cross-linked products, larger scale cross-linking reactions were carried out and the products in each case purified by reversed-phase HPLC. For positions 23, 35, 38, and 42, a single product peak was seen eluting earlier than the Tat peptide alone (e.g. position 42, Figure 6a). By contrast, in the case of position 40, two product peaks were obtained (Figure 6b). The molecular masses of each cross-linked product were determined by mass spectrometry. In each case, in addition to the peak corresponding to the unmodified TAR RNA strand, a second peak was found with *m*/*z* agreeing closely with either the expected TAR 17-mer−β-ala-dipropionyl disulfide−Tat peptide (37−72) (position 23) or TAR 14-mer−β-ala-dipropionyl disulfide−Tat peptide (37−72) (positions 35, 38, 42, and 40, peaks 1 and 2) (Table 2). Treatment of cross-linked products with DTT followed by mass spectrometry showed the expected 2′-deoxy-2′-(2-thiopropionyl-β-alanylamino)-U-containing TAR fragments (data not shown).

Each HPLC-purified cross-linked product was also subjected to enzymatic degradation by endoproteinase Lys C, which is expected to cleave uniquely after lysine residues, and the products characterized by mass spectrometry. In the case of the cross-link to position 23, two peaks were obtained corresponding to TAR 17-mer linked to peptides 42−51 and 51−72, respectively (Table 2). These fragments can only be obtained if the cross-link is to position Lys^{51}. The same peptide fragments (but linked to TAR 14-mer) were obtained for the endoproteinase Lys C digestion of the β-ala-U$_{40}$ peak 1 product, confirming that the cross-link is also to Lys^{51}. By contrast in the cases of the other cross-linked products (positions 35, 38, 42, and 40, peak 2), endoproteinase Lys C gave just one fragment corresponding to TAR 14-mer linked to the N-terminal peptide 37−41 (SFTTK). Since there are two possible amino groups in this fragment, Lys^{41} and the N-terminus, we subjected the HPLC-purified cross-

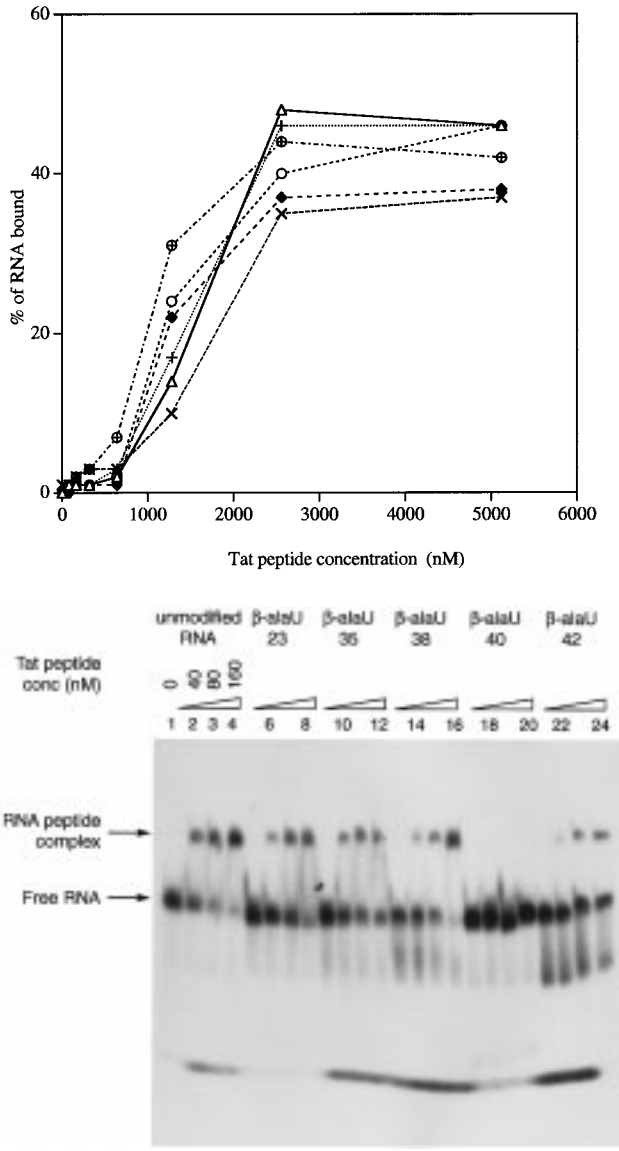


FIGURE 4: (a, top) Filter binding assay showing the amount of complex formation of wild-type TAR (−⊕−), β-alaU$_{23}$ (−×−), β-alaU$_{35}$ (−○−), β-alaU$_{38}$ (−+−), β-alaU$_{40}$ (−♦−), and β-alaU$_{42}$ (−△−) with increasing concentrations of Tat peptide (37−72): (b, bottom) autoradiograph of a polyacrylamide gel mobility shift assay showing complex formation of β-ala-modified TAR RNAs with increasing concentrations of Tat peptide (37−72).

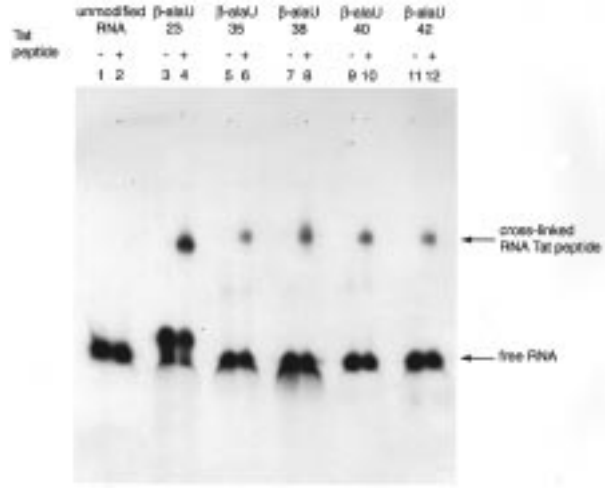


FIGURE 5: Autoradiograph of a denaturing polyacrylamide gel electrophoresis showing cross-linking between Tat peptide (37−72) and TAR RNA duplexes treated with DSP.

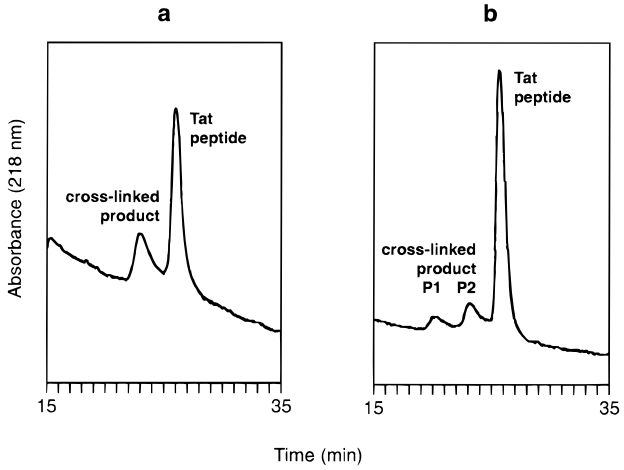


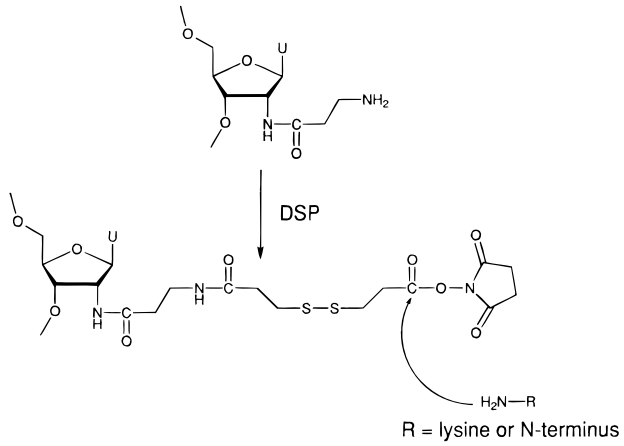
FIGURE 6: Chromatograms showing the reversed-phase HPLC purification of Tat peptide (37−72) cross-linked to DSP-treated TAR RNA substituted by β-alaU at (a) position U$_{42}$ or (b) position U$_{40}$.

linked products to chymotrypsin digestion, which is expected to cleave after aromatic residues, and products were characterized by mass spectrometry. In each of these four cases, a single peak was found corresponding to TAR 14-mer linked to the N-terminal dipeptide SF (Table 2). This indicates that in each case cross-linking has occurred uniquely to the N-terminal amino group of the Tat peptide.

DISCUSSION

We have applied the techniques of proteolytic digestion and MALDI−TOF mass spectrometry to determine the precise amino acid in a Tat peptide that had been cross-linked to TAR RNA containing a base modification (4-thioU) and also a number of TAR RNAs containing a novel sugar modification (β-alaU). In the case of 4-thioU$_{23}$-modified TAR, we have shown that photochemical cross-linking

Scheme 2



occurred uniquely at Arg^{55} within the basic region of Tat. The chemical structure of the photochemical product of 4-thioU cross-linking to arginine is not known, and its determination is beyond the scope of the present investigation. However, it is probable that reaction occurred by attack

Table 2: Mass Values for β-alaU TAR Oligonucleotides Cross-Linked to Tat Peptide (37−72) and Their Proteolytic Digestion Products

	TAR 17-mer + Tat peptide (37−72)		TAR 14-mer + Tat peptide (37−72)		
	β-alaU$_{23}$	β-alaU$_{35}$	β-alaU$_{38}$	β-alaU$_{40}$	β-alaU$_{42}$
undigested	9867.7 (9865.4)[a]	8790.1 (8791.8)	8789.3 (8791.8)	P1 8788.9 P2 8790.0 (8791.8)	8789.3 (8791.8)
LysC	6786.3 (6784.7) 42−51 ALGISYGRKK 8351.9 (8354.2) 51−72 KRRQRRRPP QGSQTHQVS LSKQ	5203.1 (5204.3) 37−41 SFTTK	5204.0 (5204.3) 37−41 SFTTK	P1 5713.0 (5713.5) 42−51 ALGISYGRKK P1 7281.1 (7280.4) 51−72 KRRQRRRPP QGSQTHQVS LSKQ P2 5203.1 (5204.3) 37−41 SFTTK	5203.8 (5204.3) 37−41 SFTTK
Chymotrypsin		4873.6 (4874.1) 37−38 SF	4873.2 (4874.1) 37−38 SF	P2 4873.5 (4874.1) 37−38 SF	4873.3 (4874.1) 37−38 SF

[a] The expected mass is given in parentheses.

of a photochemically generated thio radical on the carbon atom within the guanidinium group of arginine, followed by a rearrangement. The observed outcome is a product that has lost about 36 or 37 mass units. Although the photochemistry of RNA−RNA cross-linking products involving 4-thiouracil and 4-thiothymine has been studied in detail (*33*), to our knowledge this is not the case with 4-thioU RNA cross-linking to protein, even though the cross-linking technique has been used for some years (*31*, *34*). Nevertheless, it may be possible to use peptide digestion and mass spectral analysis techniques to determine precise cross-linking sites within larger RNA−protein complexes cross-linked through 4-thioU photochemistry. In this respect we note that contact sites between rRNA and ribosomal proteins within *E. coli* 30S subunits in terms of iminothiolane/UV cross-linked RNA−peptide fragments have recently been identified by MALDI−TOF mass spectral analysis of the oligoribonucleotide part and N-terminal microsequencing of the peptide part (*51*).

A limitation of 4-thioU cross-linking to Tat peptide we found was that cross-linking yields were extremely dependent on the site of introduction of 4-thio U into TAR RNA. RNA−RNA cross-linking is also a complicating problem (*36*). Low yields of RNA−protein cross-linking were also reported recently for photochemical cross-linking of 6-thioG-substituted RRE RNA to HIV-1 Rev protein (*52*). Thus, the ability to determine precise sites on peptides and proteins using such photochemical cross-linking techniques may be limited to favorable cases where sufficient cross-linked material can be isolated.

Our new cross-linking strategy was aimed at obtaining a more generally applicable method that would be suitable for any RNA fragment accessible by chemical synthesis and that would not be accompanied by significant RNA−RNA cross-linking. We chose the 2′-position of the RNA as a potentially good anchoring point on the exterior of TAR duplex RNA and a β-ala functionality to provide an accessible amino group to which a variety of cross-linkers might in principle be attached. We showed that introduction of the β-ala group into TAR RNA was not harmful to Tat peptide binding, thus confirming the suitability of 2′-functionalization.

In principle it should be possible to attach to the β-alaU residue a range of cross-linkers, such as a thiol-specific reagent or a photochemically activated azidophenacyl group. However, since there are two lysine residues within the important basic domain of Tat and since one of these had already been shown in our previous studies to be cross-linkable to a trisubstituted pyrophosphate analogue of TAR (*30*), we chose the amino-specific homobifunctional reagent DSP to chemically cross-link the β-ala TAR to Tat. After reaction with DSP, the TAR duplex contains a reactive succinimido group attached to a flexible linker which could span a distance of up to 16 Å when fully extended. Our results showed that cross-linking took place nevertheless with remarkable specificity within the bound Tat peptide complex. Of the five potentially reactive amino groups on Tat only two, Lys^{51} and the N-terminus, were found cross-linked in each of the five TAR sites tested. Further, four of the five TAR sites cross-linked uniquely and in only one case (U$_{40}$) were two cross-linked products obtained. The results show that the presence of a long and flexible arm extending from the 2′-position of a sugar does not cause a significant loss in cross-linking specificity as compared to armless cross-linking via a base or phosphate residues.

Implications of Cross-Linking Specificity on Tat−TAR Recognition. The ability of a particular amino group in Tat peptide to be cross-linked might be expected to be dependent primarily on two factors. First, the cross-linking arm must be able to reach to the amino group, and second, the amino group must be able to flex sufficiently from its preferred position in the bound structure. Presumably, each of these processes must be able to take place under conditions that do not disturb the bound state of the RNA−protein complex significantly. The high degree of specificity obtained indicates that Lys^{51} and the N-terminus in Tat peptide (37−72) must meet these criteria, whereas none of the other three

lysines apparently do so, at least for the five TAR sites selected.

In the case of 4-thio U_{23} cross-linking to Arg^{55}, similar principles must apply. However, since there is no long arm on the RNA to search for a susceptible amino acid, we propose that here both the arginine side chain and the nucleoside base itself must be able to move just enough to allow the close contact required for photochemical reaction, again without disturbing the bound state of the complex. Keeping in mind these principles, it is possible to consider to what extent the cross-linking data can be reconciled with models of Tat−TAR recognition.

Comparison of the NMR structures of free HIV-1 TAR and in the presence of Tat peptide (37−72) showed that a major conformational change in TAR is triggered by Tat binding in the region of the U-rich bulge (*24, 26*). The major groove is narrowed considerably and appears to be too narrow to accommodate common protein structural elements such as an α-helix or β-sheet. Since a similarly large conformational change is obtained in the case of binding by Tat peptide (37−72) and an arginine derivative (argininamide) (*25, 26, 53*), it has been proposed that a single arginine residue in the basic region of Tat is key to the conformational change in TAR (*25*). TAR binding and CD studies of small basic peptides have indicated that the critical arginine in Tat which triggers the conformational change might be Arg^{52} (*54*). However, other transactivation and TAR binding studies have shown the importance of both Arg^{52} and Arg^{53} (*12, 21, 23*). Arginines in positions 55 and 56 were shown to be less important to TAR binding (*15*). An early model of argininamide binding to HIV-1 TAR suggested insertion into the major groove and recognition of N^7 and O^6 positions of G_{26} (*25*). It has also been proposed that an arginine recognizes two adjacent phosphates (P21−22 and P22−23) in TAR (*21, 25*).

There have been two recent NMR structures of bovine immunodeficiency virus (BIV) TAR complexed to a BIV Tat peptide (*55, 56*) in which it was found that a single arginine (position 73) carries out both roles of base and phosphate recognition simultaneously. However, in contrast to BIV, there are two adjacent arginines (Arg^{52} and Arg^{53}) in HIV-1 Tat in the analogous location, and there is a strong possibility that here base and phosphate recognition is divided between two arginines. There are also other structural differences between the BIV and HIV TAR NMR structures. For example, the proposed BIV TAR base-triple occurring between U_{23}:A_{27}:U_{38} in the bound complex (*56*), has not been evidenced so far in the case of the NMR data for the HIV-1 Tat peptide−TAR complex (*26*). This suggests that the folding of the HIV-1 Tat peptide within the TAR structure may be different from the BIV case where the peptide has a β-turn conformation.

NMR measurements on the HIV-1 Tat peptide−TAR complex have so far led to the identification of a number of intermolecular NOE resonances, for example between arginines and the bases of G_{26} and U_{23} (*59*), but it has been difficult so far to assign these unambiguously and to obtain an unequivocal structure. We have examined preliminary models of the basic region of the HIV-1 Tat peptide docked within the major groove of the TAR RNA bulge region that are consistent with the NMR data. Our cross-linking results are best accommodated by one of these models where the basic region of the Tat peptide has an extended structure (Figure 7). The orientation of the basic region in the groove is N-terminal toward the lower stem and C-terminal toward the upper stem, which is consistent with cross-linking data of Wang and Rana (*27, 28*). The side chain of Arg^{52} points into the major groove toward the bases of G_{26} and A_{27} while Arg^{53} is located close to the phosphates P21−22 and P22−23. Lys^{51} is directed toward the phosphate backbone on the opposite strand, consistent with the finding of cross-linking to tsp 38−39 (*30*).

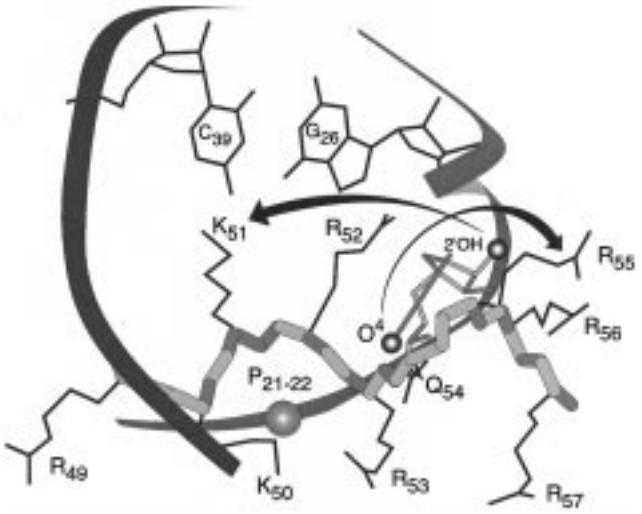


FIGURE 7: Schematic model of the basic region (residues 49−57) of the HIV-1 Tat peptide bound to TAR RNA in the region of the U-rich bulge, which most readily accommodates the cross-linking data. The model is viewed by looking down into the major groove of the RNA from above. TAR RNA is shown in ribbon form except for the G_{26}•C_{39} base pair, which is almost in the plane of the paper, and residue U_{23} which is perpendicular to this plane. The O^4 and 2′-OH groups of U_{23} are shown as filled spheres. This view emphasizes the extended nature of the peptide backbone, which is shown as a thick line with the side chains distributed on both sides. Arg^{52} is seen pointing in the direction of G_{26}, whereas Arg^{53} points toward phosphate P21−22 (larger sphere). The proposed flexing of the side chain of Arg^{55} toward the O^4-position of U_{23} (S^4 in 4-thioU) is marked by a curved arrow. A second curved arrow delineates the relatively clear pathway toward the epsilon amino group of Lys^{51} which could be followed by a flexible arm of up to 16 Å.

Of the six arginines in the basic region of Tat in the model, the two closest to the O^4-position of U_{23} (the site of cross-linking in 4-thioU) are Arg^{52} and Arg^{53} (each about 6 Å), but each have their guanidino groups pointing away from the base of U_{23}. Arg^{55} is about 9 Å, but this residue does not apparently form a close contact with the RNA. Examination of the positioning of Arg^{55} (Figure 7) shows that rotation and flexing of the side chain alone could bring it sufficiently close to the O^4-position of U_{23} for photochemical reaction. Since U_{23} is not base-paired, this nucleoside residue also may be somewhat flexible. By contrast, movement of either Arg^{52} or Arg^{53} side chains would result in breakage of strong interactions with the RNA which would presumably not be favored. Arg^{56} and Arg^{57} are each pointing away from residue U_{23} in this model, and movement of the peptide backbone would be necessary as well as the side chain in order to get close to residue U_{23}. An alternative model, where Arg^{53} makes contact with G_{26} and Arg^{52} makes phosphate contacts, is less satisfactory because Arg^{55} is in a

less favorable orientation and further away from the U_{23} O^4-position (data not shown).

In our model the 2′-hydroxyl group of U_{23} is predicted to be about 11 Å from Lys^{51} (Figure 7). A 16 Å flexible extension from U_{23} O2′ (a β-ala-propionyl disulfide linker) would be able to pass easily over the top of the peptide without significant hindrance. Although pointing toward the phosphate backbone, Lys^{51} must be able to flex its side chain in order to react. By contrast Lys^{50} (14 Å distant) is pointing in the opposite direction. A cross-link to the 2′-position of U_{23} would need to pass through the region of the peptide itself as well as nucleoside U_{23}. This is structurally unlikely and explains why Lys^{50} does not become cross-linked. A similar situation arises when considering distances of U_{40} 2′-OH to Lys^{51} (13 Å), which is much closer than Lys^{50} (20 Å) (not shown). Thus, the model nicely explains the U_{23} and U_{40} sugar cross-linking data.

As might be expected, cross-linking from the 2′-positions of U_{35}, U_{38} in the upper stem and U_{42} in the lower stem did not find any lysine residue in the basic region of the Tat peptide because these are too distant or presumably in an unfavorable orientation. Interestingly cross-linking was seen to the N-terminus in each case rather than to either of the residues Lys^{41} or Lys^{71}. We can rationalize this finding if the N-terminal section of the peptide is highly flexible and can reach both to the lower stem as well as to the upper stem. Tat−peptide (37−72) does not contain the complete core region of Tat, and the N-terminal residue would normally be Cys (replaced by Ser here). Thus, in the full Tat protein this region is likely to be involved in a defined structure involving the seven conserved cysteine residues, whereas the N-terminal few residues in the Tat peptide are not so constrained. This result is consistent with a recent report that the N-terminus of a similar Tat peptide modified by an affinity cleaving group can cleave residues in both the TAR bulge and apical loop regions (*57*). By contrast, Lys^{41} and Lys^{71} may be within more structured elements involved closely in interaction with the RNA or with other parts of the peptide. In this respect we have recently shown that the C-terminal region of Tat peptide is indeed orientated toward the upper stem of TAR, since a trisubstituted pyrophosphate analogue of phosphate P36−37 is able to cross-link to Lys^{71} (*60*).

Through the techniques described in this and previous studies (*30*), we have been able to identify a number of sites on Tat peptide (37−72) susceptible to specific cross-linking to chemically modified TAR duplexes. These data have been shown to be most consistent with one of a number of preliminary models of the interaction of the basic region of Tat with TAR obtained through NMR studies. Further NMR and cross-linking data will clearly be necessary to verify and further refine Tat−TAR interaction models. The use of specifically labeled arginine peptides may prove crucial in this respect. A fully verified model of the Tat−TAR interaction would be of significant value in aiding the design of peptide-like inhibitors of the Tat−TAR interaction (*58*).

ACKNOWLEDGMENT

We thank Mohinder Singh for invaluable technical assistance and Terry Smith, Jan Fogg, and Richard Grenfell for the synthesis and purification of oligoribonucleotides. We thank Jonathan Karn, Gabriele Varani, and Andres Ramos for encouragement and advice at various stages of this work.

REFERENCES

1. Nagai, K., and Mattaj, I. W., Ed. (1994) *RNA-protein interactions*, Frontiers in Molecular Biology, Oxford University Press, Oxford.
2. Draper, D. E. (1995) *Annu. Rev. Biochem. 64*, 593−620.
3. Karn, J., Churcher, M. J., Rittner, K., Keen, N., and Gait, M. J. (1996) in *Eukaryotic Gene Transcription* (Goodbourn, S., Ed.) pp 254−286, Oxford University Press, Oxford.
4. 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.
5. 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.
6. Roy, S., Delling, U., Chen, C.-H., Rosen, C. A., and Sonenberg, N. (1990) *Genes Dev. 4*, 1365−1373.
7. Müller, W. E. G., Okamoto, T., Reuter, P., Ugarkovic, D., and Schröder, H. C. (1990) *J. Biol. Chem. 265*, 3803−3808.
8. Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., and Rösch, P. (1995) *J. Mol. Biol. 247*, 529−535.
9. Gregoire, C. J., and Loret, E. P. (1996) *J. Biol. Chem. 271*, 22641−22646.
10. Metzger, A. U., Schindler, T., Willbold, D., Kraft, M., Steegborn, C., Volkmann, A., Frank, R. W., and Rösch, P. (1996) *FEBS Lett. 384*, 255−259.
11. Cordingley, M. G., LaFemina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., LeGrow, K., Gotlib, L., Schlabach, A. J., and Colonno, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8985−8989.
12. Calnan, B. J., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Genes Devel. 5*, 201−210.
13. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., and Crothers, D. M. (1990) *Science 249*, 1281−1285.
14. Weeks, K. M., and Crothers, D. M. (1991) *Cell 66*, 577−588.
15. Long, K. S., and Crothers, D. M. (1995) *Biochemistry 34*, 8885−8895.
16. Kamine, J., Loewenstein, P., and Green, M. (1991) *Virology 182*, 570−577.
17. Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. G., Gait, M. J., and Karn, J. (1993) *J. Mol. Biol. 230*, 90−110.
18. Delling, U., Reid, L. S., Barnett, R. W., Ma, M. X. Y., Climie, S., Sumner-Smith, M., and Sonenberg, N. (1992) *J. Virol. 66*, 3018−3025.
19. Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U., and Sonnenberg, N. (1991) *J. Virol. 65*, 5196−5201.
20. Hamy, F., Asseline, U., Grasby, J. A., Iwai, S., Pritchard, C. E., Slim, G., Butler, P. J. G., Karn, J., and Gait, M. J. (1993) *J. Mol. Biol. 230*, 111−123.
21. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Science 252*, 1167−1171.
22. Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J., and Gait, M. J. (1994) *Nucleic Acids Res. 22*, 2592−2600.
23. Tao, J., and Frankel, A. D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 1571−1575.
24. Aboul-ela, F., Karn, J., and Varani, G. (1996) *Nucleic Acids Res. 24*, 3974−3982.
25. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) *Science 257*, 76−80.
26. Aboul-ela, F., Karn, J., and Varani, G. (1995) *J. Mol. Biol. 253*, 313−332.
27. Wang, Z., and Rana, T. M. (1995) *J. Am. Chem. Soc. 117*, 5438−5444.
28. Wang, Z., Wang, X., and Rana, T. (1996) *J. Biol. Chem. 271*, 16995−16998.

29. Liu, Y., Wang, Z., and Rana, T. M. (1996) *J. Biol. Chem. 271*, 10391−10396.
30. Naryshkin, N. A., Farrow, M. A., Ivanovskaya, M. G., Orestkaya, T. S., Shabarova, Z. A., and Gait, M. J. (1997) *Biochemistry 36*, 3496−3505.
31. Favre, A., Bezerra, R., Hajnsdorf, E., Dubreuil, L., and Expert-Bezançon, A. (1986) *Eur. J. Biochem. 160*, 441−449.
32. Adams, C. J., Murray, J. B., Arnold, J. R. P., and Stockley, P. G. (1994) *Tetrahedron Lett. 35*, 765−768.
33. Favre, A., and Fourrey, J.-L. (1995) *Acc. Chem. Res. 28*, 375−382.
34. Wollenzien, P., Expert-Bezançon, A., and Favre, A. (1991) *Biochemistry 30*, 1788−1795.
35. McGregor, A., Rao, M. V., Duckworth, G., Stockley, P. G., and Connolly, B. A. (1996) *Nucleic Acids Res. 24*, 3173−3180.
36. Wang, Z., and Rana, T. M. (1996) *Biochemistry 35*, 6491−6499.
37. Johnson, T., Quibell, M., Owen, D., and Sheppard, R. C. (1993) *J. Chem. Soc., Chem. Commun.* 369−372.
38. Quibell, M., Turnell, W. G., and Johnson, T. (1994) *J. Org. Chem. 59*, 1745−1750.
39. Quibell, M., Packman, L. C., and Johnson, T. (1996) *J. Chem. Soc., Perkin Trans. 1*, 1227−1234.
40. Schmidt, S., Beigelman, L., Karpeisky, A., Usman, N., Sørensen, U. S., and Gait, M. J. (1996) *Nucleic Acids Res. 24*, 573−581.
41. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) *Nucleic Acids Res. 23*, 2677−2684.
42. Gait, M. J., Earnshaw, D. J., Farrow, M. A., Fogg, J. H., Grenfell, R. L., Naryshkin, N. A., and Smith, T. V. (1997) in *RNA-Protein Interactions: A Practical Approach* (Smith, C., Ed.) in press, OUP, Oxford, U.K.
43. Slim, G., and Gait, M. J. (1991) *Nucleic Acids Res. 19*, 1183−1188.
44. Shah, K., Wu, H., and Rana, T. M. (1994) *Bioconj. Chem. 5*, 508−512.
45. Atherton, E., Cammish, L. E., Goddard, P., Richards, J. D., and Sheppard, R. C. (1984) in *Peptides 1984. Proceedings of the 18th European Peptide Symposium, Djurönäset, Sweden, June 10−15, 1984* (Ragnarsson, U., Ed.) pp 153−157, Almqvist & Wiskell, Stockholm, Sweden.
46. Cammish, L. E. (1987) Ph.D. Thesis, University of Cambridge, U.K.
47. Rigl, C. T., Lloyd, D. H., Tsou, D. S., Gryaznov, S. M., and Wilson, W. D. (1997) *Biochemistry 36*, 650−659.
48. Pieken, W. A., Olsen, D. B., Benseler, F., Aurup, H., and Eckstein, F. (1991) *Science 253*, 314−317.
49. Matulic-Adamic, J., Beigelman, L., Dudycz, L. W., Gonzalez, C., and Usman, N. (1995) *Bioorg. Med. Chem. Lett. 5*, 2721−2724.
50. Joshi, S., and Burrows, R. (1990) *J. Biol. Chem. 265*, 14518−14525.
51. Urlaub, H., Thiede, B., Müller, E.-C., Brimacombe, R., and Wittmann-Leibold, B. (1997) *J. Biol. Chem. 272*, 14547−14555.
52. Ping, Y.-H., Liu, Y., Wang, X., Neenhold, H. R., and Rana, T. M. (1997) *RNA 3*, 850−860.
53. Puglisi, J. D., Chen, L., Frankel, A. D., and Williamson, J. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3680−3684.
54. Tan, R., and Frankel, A. D. (1992) *Biochemistry 31*, 10288−10294.
55. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) *Science 270*, 1200−1203.
56. Ye, X., Kumar, R. A., and Patel, D. J. (1995) *Chem. Biol. 2*, 827−840.
57. Huq, I., and Rana, T. M. (1997) *Biochemistry 36*, 12592−12599.
58. Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., Aboul-ela, F., Varani, G., Karn, J., and Klimkait, T. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3548−3553.
59. Aboul-ela, F., and Varani, G. Unpublished results.
60. Farrow, M., and Gait, M. Unpublished results.

BI972695V