

# Molecular recognition in the bovine immunodeficiency virus Tat peptide–TAR RNA complex

Xiaomei Ye, R Ajay Kumar<sup>†</sup> and Dinshaw J Patel\*

Cellular Biochemistry & Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA

**Background:** In lentiviruses such as human immunodeficiency virus (HIV) and bovine immunodeficiency virus (BIV), the Tat (*trans*-activating) protein enhances transcription of the viral RNA by complexing to the 5'-end of the transcribed mRNA, at a region known as TAR (the *trans*-activation response element). Identification of the determinants that account for specific molecular recognition requires a high resolution structure of the Tat peptide–TAR RNA complex.

**Results:** We report here on the structural characterization of a complex of the recognition domains of BIV Tat and TAR in aqueous solution using a combination of NMR and molecular dynamics. The 17-mer Tat peptide recognition domain folds into a  $\beta$ -hairpin and penetrates in an edge-on orientation deep into a widened major groove of the 28-mer TAR RNA recognition domain in the complex. The RNA fold is defined, in part, by two uracil bulged bases; U12 has a looped-out conformation

that widens the major groove and U10 forms a U•AU base triple that buttresses the RNA helix. Together, these bulged bases induce a ~40° bend between the two helical stems of the TAR RNA in the complex. A set of specific intermolecular hydrogen bonds between arginine side chains and the major-groove edge of guanine residues contributes to sequence specificity. These peptide–RNA contacts are complemented by other intermolecular hydrogen bonds and intermolecular hydrophobic packing contacts involving glycine and isoleucine side chains.

**Conclusions:** We have identified a new structural motif for protein–RNA recognition, a  $\beta$ -hairpin peptide that interacts with the RNA major groove. Specificity is associated with formation of a novel RNA structural motif, a U•AU base triple, which facilitates hydrogen bonding of an arginine residue to a guanine and to a backbone phosphate. These results should facilitate the design of inhibitors that can disrupt HIV Tat–TAR association.

**Chemistry & Biology** December 1995, 2:827–840

Key words: arginine–guanine interactions, buttressing U•AU base triple, glycine and isoleucine packing, peptide–RNA recognition, RNA bending

## Introduction

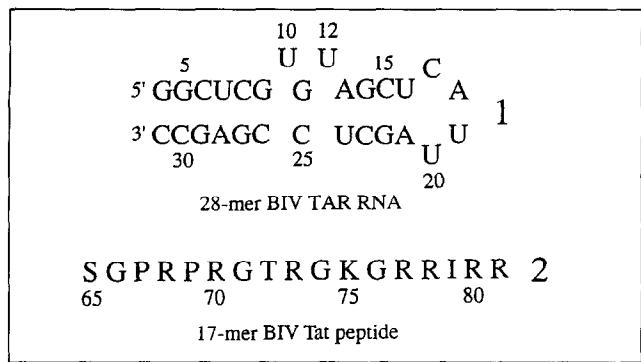
Recent crystallographic studies of protein–RNA complexes (reviewed in [1,2]) have revealed several important principles for the sequence-specific recognition of RNA secondary and tertiary structure. Our understanding of the basis of protein–RNA recognition is limited, as the structures of very few protein–RNA complexes have been determined. The loop, bulge, bubble and pseudoknot motifs embedded within helical segments of RNA may be important targets for molecular recognition by peptides.

There are several protein–RNA interactions that are fundamentally important in the life cycle of immunodeficiency viruses. The Tat (*trans*-activating) protein of lentiviruses enhances viral RNA transcription by binding to the 5'-end of the transcribed mRNA, at a region known as TAR (*trans*-activation response element). The Tat–TAR interaction has been studied in detail in both the human (HIV) and bovine (BIV) immunodeficiency viruses (reviewed in [3,4]). Short stretches of Tat and TAR, the recognition domains, are sufficient to examine the basis for sequence-specific complex formation. It has been shown for both HIV and BIV that single or multiple base bulges in the stem-loop TAR RNA are crucial for the recognition of this site by Tat; the region of Tat

responsible for recognition is basic and rich in arginines. The recognition sites have been mapped out in detail from footprinting studies combined with the effects of mutations in peptide and RNA on binding affinity [3,4].

NMR studies have been applied previously to characterize the HIV Tat–TAR interaction in solution. These studies have been hindered by the inability to monitor the peptide protons in the complexes. The previous research therefore addressed the simplest possible system, namely, the binding of arginineamide to the bulge-containing stem-loop TAR site [5,6]. A key insight that emerged from these studies was that the arginine side chain shows a bifurcated interaction with the major-groove edge of guanine and with a backbone phosphate. This arginine–guanine–phosphate interaction was postulated to contribute to the specificity of molecular recognition in this system. Further, a novel U•AU base triple was proposed to form upon arginineamide–HIV TAR RNA complex formation through pairing of a bulged uracil with a stem A•U base pair [6]. The evidence for the U•AU base triple formation was primarily based on sequence covariation experiments; the arginineamide still bound when the sequence was changed to generate the potential C•GC triple in the TAR RNA [7]. More recent detailed NMR studies on the same arginineamide–HIV

\*Corresponding author. <sup>†</sup>This author was responsible for the computational aspects of the project.



**Fig. 1.** Nucleotide sequence of the 28-nucleotide BIV TAR RNA 1 showing base-paired regions, and amino-acid sequence (in single letter code) of the 17-residue BIV Tat peptide used to determine the structure of the Tat-TAR complex.

TAR RNA system have challenged the notion of the formation of a U•AU base triple in the complex [8].

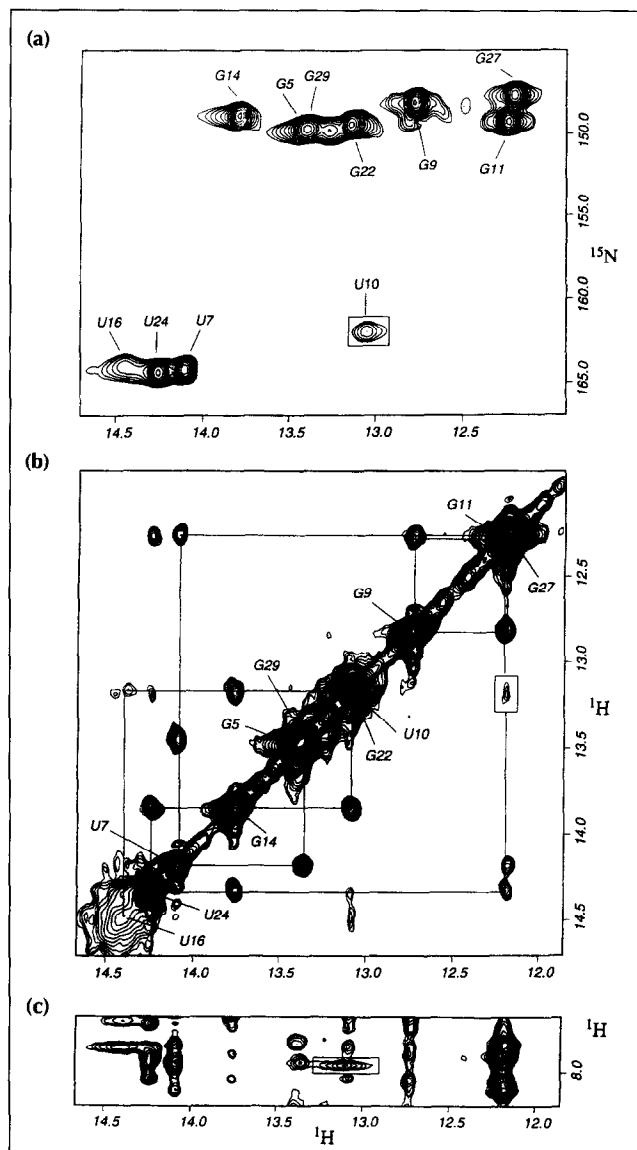
The corresponding BIV system provides an attractive alternative system for study since the minimal Tat peptide and TAR RNA necessary for sequence-specific complex formation has been determined, as well as the effects of a large number of Tat and TAR mutants on binding affinity [9,10]. The tight binding constant for the BIV Tat-TAR complex suggests that this system should be amenable to NMR structural characterization.

There have been significant recent advances in the application of NMR spectroscopy to elucidate the folded conformations of RNA in solution (reviewed in [11–14]). These efforts have been greatly aided by the availability of uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled RNA [15,16] and the application of heteronuclear multidimensional NMR techniques (reviewed in [17,18]) to extract the maximum structural information in crowded regions of the spectrum. We report here on the solution structure of the complex formed between the 28-nucleotide (28-mer) BIV TAR RNA (structure 1, Fig. 1) and a 17-amino-acid (17-mer) BIV Tat peptide, representing the TAR recognition domain, (structure 2, Fig. 1) in solution using a combined NMR–molecular dynamics approach. Our goal was to define the structure and identify the intermolecular contacts that account for the sequence-specificity and affinity of complex formation. We have characterized the critical role in complex formation of the large number of arginines and glycines and of the single isoleucine in the BIV Tat sequence 2 and the two uracil bulges separated by a G•C base pair in the BIV TAR RNA sequence 1. As the Tat and TAR sequences of HIV and BIV exhibit similarities, the principles deduced from our results on the the BIV Tat-TAR complex may be applicable to the HIV system.

## Results

### Complex formation

We monitored the formation of the BIV Tat-TAR complex on gradual addition of BIV 17-mer Tat peptide 2 to the BIV 28-mer TAR RNA 1  $\text{H}_2\text{O}$  solution, pH 6.8 by recording the imino proton NMR spectra



**Fig. 2.** Spectroscopic analysis of exchangeable nucleic acid protons in the BIV Tat-TAR RNA complex. (a)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra correlating imino proton and attached nitrogen regions of the complex containing uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled TAR RNA in  $\text{H}_2\text{O}$  buffer at pH 5.5 and 10 °C. The assignments are listed next to the crosspeaks. The boxed crosspeak is the U10 assignment, which correlates the U10 imino proton (13.01 ppm) with its attached nitrogen (162.08 ppm). (b) Expanded NOESY contour plots (150 ms mixing time) in the BIV Tat-TAR complex in  $\text{H}_2\text{O}$  buffer at pH 5.5 and -2 °C. Distance connectivities between TAR RNA imino protons in the symmetrical 12.0–14.7 ppm region. The imino proton assignments are listed along the diagonal and the lines trace the connectivities between imino protons on adjacent base pairs along the TAR RNA helix. The boxed crosspeak corresponds to a weak NOE between the broad imino proton of U10 (13.18 ppm) and the imino proton of G11 (12.18 ppm). (c) NOE connectivities between TAR RNA imino protons (12.0–14.7 ppm) and a portion of the base and amino proton region. The boxed crosspeak is the NOE between the broad imino proton of U10 (13.18 ppm) and the H8 proton of A13 (7.92 ppm).

(12.0–14.7 ppm) at 1 °C (see Supplementary material). We detect separate resonances for free and bound TAR RNA at substoichiometric ratios, indicating slow exchange due to the formation of a high affinity complex.

We detect the same number of imino-proton resonances in the complex except that several imino protons undergo significant downfield (G14, U16 and U24) and upfield (G11 and G22) shifts upon complex formation.

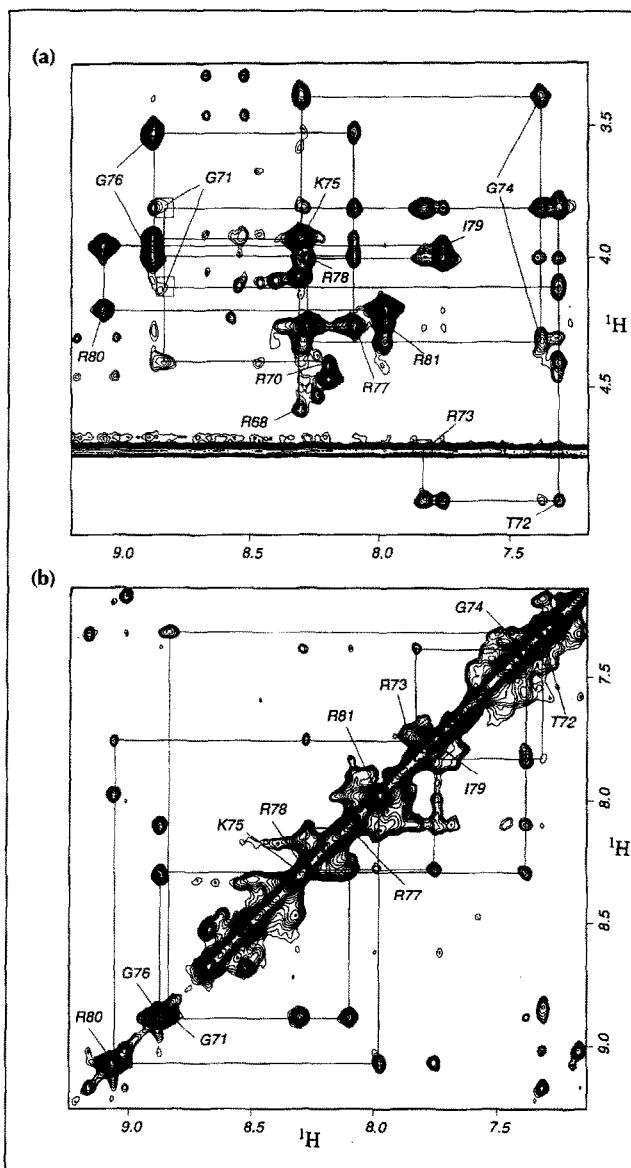
#### Exchangeable nucleic acid protons

The imino protons of the 28-mer TAR RNA in the BIV Tat-TAR complex have been assigned using standard nucleic acid assignment procedures [19]. The experimental data on the complex in  $H_2O$  buffer were collected at both pH 6.8 and pH 5.5 over the -2 to 25 °C temperature range. We anticipated that low pH and temperature would facilitate the detection and identification of the imino protons of the U10 and U12 bulge residues in the complex.

The imino protons can be correlated with their attached nitrogens in a  $^1H$ ,  $^{15}N$  heteronuclear single quantum coherence (HSQC) contour plot of the BIV Tat-TAR complex containing uniformly  $^{15}N$ ,  $^{13}C$ -labeled TAR RNA in  $H_2O$  buffer, pH 5.5 at 10 °C (Fig. 2a). The guanine (147–151 ppm) and uracil (161–165 ppm) nitrogens exhibit distinct chemical shifts and can be readily differentiated by class. An expanded region of the nuclear Overhauser effect spectroscopy (NOESY) contour plot correlating NOE connectivities between imino protons (12.0–14.5 ppm) and base and amino protons (5.2–9.2 ppm) is given in the Supplementary materials. We detect the characteristic NOEs between uracil imino protons and adenine H2 and amino protons across Watson-Crick A•U base pairs and between guanine imino protons and cytosine amino protons across Watson-Crick G•C base pairs in the complex.

An expanded NOESY contour plot of the BIV Tat-TAR complex in  $H_2O$  buffer, pH 5.5 at -2 °C correlates NOE connectivities in the symmetrical 12.0–14.5 ppm region (Fig. 2b). The NOE connectivities can be traced from the imino proton of G29 at one end of the helix towards the broad imino proton of U16 located at the other end of the TAR RNA helix. It is significant that NOEs are detected between the imino protons of G9 and G11, which span the intervening U10 bulge residue, and between the imino protons of G11 and U24, which span the intervening U12 bulge residue in the complex.

We have identified the U10 imino proton resonance in the BIV Tat-TAR complex. It can be clearly detected in low temperature NMR spectra recorded at pH 5.5, but is very weak at pH 6.5, reflecting the slowing of its exchange rate with water at acidic pH. The U10 imino proton is broad, resonates at 13.18 ppm (-2 °C) and shifts upfield with increasing temperature. It exhibits a broad  $^1H$ – $^{15}N$  HSQC correlation crosspeak (boxed peak, Fig. 2a) similar to that detected for the broad U16 residue (Fig. 2a) in the complex. Further, the imino proton of U10 exhibits a weak and broad NOE to the imino proton of G11 in the NOESY (150 ms mixing time) contour plot of the complex recorded in  $H_2O$  buffer,



**Fig. 3.** Expanded  $^{13}C$ -filtered NOESY (250 ms mixing time) contour plots in the BIV Tat-TAR complex containing uniformly labeled  $^{13}C$ ,  $^{15}N$ -TAR RNA in  $D_2O$  buffer, pH 6.8 at 25 °C. (a) NOE connectivities between the NH (7.3–9.1 ppm) and CaH (3.3–5.0 ppm) protons of the Tat peptide from R70 to R81 in the complex. The lines trace the sequential NOE connectivities along the peptide chain. (b) NOE connectivities between amide protons (symmetrical 7.2–9.2 ppm region) of the Tat peptide from G71 to R81 in the complex.

pH 5.5 at -2 °C (boxed peak, Fig. 2b). It is important to note that a broad crosspeak is clearly detectable between the imino proton of U10 and the H8 proton of A13 in NOESY (150 ms mixing time) spectra of the complex in  $H_2O$  buffer, pH 5.5, recorded at 10 °C and -2 °C (boxed peak, Fig. 2c). This NOE pattern is characteristic of the Watson-Crick edge of U10 interacting with the major-groove edge of A13 which is in a Watson-Crick A•U pair to form a U•AU triple in the complex. The formation of the U10•A13•U24 base triple can occur if the bulge U10 base does not stack between its flanking G•C base pairs. The observation of an NOE between the

**Table 1. Intermolecular NOEs in the BIV Tat-TAR complex**

Amino acid	Nucleotide	NOE assignments
G66	$\alpha\text{CH}_2$	C17 H5
P67	$\beta\text{CH}_2, \gamma\text{CH}_2$ $\delta\text{CH}_2$	A18 H1' C17 H5
R68	$\beta\text{CH}_2, \gamma\text{CH}_2$ $\delta\text{CH}_2$ $\text{NH}_2$	A18 H2 A21 H8, A18 H2, U20 H1', U19 H1' A21(H8, H1')
R70	$\beta\text{CH}_2, \gamma\text{CH}_2$ $\delta\text{CH}_2$ $\epsilon\text{NH}$ $\text{NH}_2$	G22 NH1, G14 NH1, A13 NH2 U24 NH3, G14 NH1, G22 NH1, C23 NH2, U10 H5, A13 NH2 G14 NH1, C15 H5 A13 H8
G71	$\alpha\text{CH}_2$	A21 (H8, H1'), G22 H8
T72	$\alpha\text{H}$ $\beta\text{H}$ $\gamma\text{CH}_3$ $\gamma\text{OH}$	C23 H5 C23 (H5, H6) C23 (H5, H6), A21 (H8, H1', H2', H3'), G22 (H8, H1', H2', H3', H4', H5', H5'') C23 (H5, H6)
R73	$\beta\text{CH}_2, \gamma\text{CH}_2$ $\delta\text{CH}_2$ $\epsilon\text{NH}$	G9 NH1, G11 NH1, C25 NH2 G9 NH1, G11 NH1 G11 NH1
G74	$\text{NH}$ $\alpha\text{CH}_2$	C23 (H6, H5) U24 (H5, H6), C23 (H6, H5, H2')
R77	$\beta\text{CH}_2, \gamma\text{CH}_2$ $\delta\text{CH}_2$ $\epsilon\text{NH}$	C8 H5 C8 (H5, H6), G9 H8 G9 H8, C8 H6
I79	$\alpha\text{H}$ $\beta\text{H}, \gamma\text{CH}_2$ $\gamma\text{CH}_3, \delta\text{CH}_3$	U10 H5 U10 (H5, H6) U10 (H5, H6), U24 NH3

imino protons of G9 and G11 in the complex (Fig. 2b) is consistent with this conclusion resulting in a mutual stacking of the G9•C26 and G11•C25 base pairs in the complex. Finally, the NOE patterns involving the U10 imino protons are also detected in the BIV Tat-TAR complex generated from a TAR mutant where U12 is replaced with C.

In contrast, we have been unable to detect the imino proton of U12 between 12.0 and 15.0 ppm in the complex at acidic pH and low temperature conditions. This result, together with the observation that we detect an NOE between the imino protons of G11 and U24, is indicative of U12 looping out of the helix resulting in mutual stacking of the flanking G11•C25 and A13•U24 base pairs in the complex.

#### Non-exchangeable nucleic acid protons

The non-exchangeable nucleic acid base and sugar protons have been assigned following analysis of through-bond and through-space connectivities in the BIV Tat-TAR RNA complex. Expanded NOESY (250 ms mixing time) contour plots correlating NOEs between base protons (6.9–8.4 ppm) with their own and 5'-flanking-sugar H1' protons (5.0–6.1 ppm) in the complex in

$\text{D}_2\text{O}$  buffer, pH 6.8 at 25 °C are given in the Supplementary materials. The chain tracing can be readily followed from C17 to C30 including the U24-C25-C26 segment facing the U10 and U12 bulge residues in the TAR RNA. It was possible to differentiate between U10 and U12 by studying the BIV Tat-TAR complex containing a C12 for U12 substitution in the TAR RNA. There is a break in the connectivity at the G9-U10 and G11-U12 steps, whereas it is difficult to trace the connectivities through the U10-G11 and U12-A13 steps because of the similar chemical shifts for the sugar H1' protons of U10 and G11 and also because of crosspeak overlaps between the base/sugar H1' protons of U12 and A13 in the complex.

The assigned sugar H1' protons can then be correlated with the remaining sugar protons through a combination of  $^1\text{H}-^{13}\text{C}$  HCCH-correlation spectroscopy (COSY) and HCCH-total correlation spectroscopy (TOCSY) experiments on the BIV Tat-TAR complex containing uniformly  $^{13}\text{C}, ^{15}\text{N}$ -labeled TAR RNA in  $\text{D}_2\text{O}$  solution at 25 °C. We have also used several heteronuclear three- and four-dimensional NOESY experiments to try to resolve ambiguities due to crosspeak overlaps in homonuclear two-dimensional NOESY spectra of the complex. This information can be found in the Supplementary materials.

A distinction can be made between C3'-*endo* (negligible J[H1'-H2']) coupling constant) and C2'-*endo* (large J[H1'-H2']) coupling constant) families of sugar puckers following analysis of the double-quantum-filtered (DQF)-COSY spectrum of the BIV Tat-TAR complex in  $\text{D}_2\text{O}$  buffer at 25 °C (Supplementary materials). These results indicate that the U17-A18-U19-U20 hairpin loop, adjacent residue A21 and the U10-G11-U12 bulge-containing segment (and to a lesser extent adjacent A13) adopt sugar pucker in the C2'-*endo* range in the BIV Tat-TAR complex. The remaining non-terminal sugars adopt the characteristic A-RNA C3'-*endo* conformation in the complex.

#### Exchangeable and non-exchangeable peptide protons

The amino acid backbone and side chain protons of the 17-mer Tat peptide in the BIV Tat-TAR complex have been assigned using standard procedures [20]. These assignments were aided by recording  $^{13}\text{C}$ -filtered NOESY (250 ms mixing time) spectra on the complex containing  $^{13}\text{C}, ^{15}\text{N}$ -labeled TAR RNA in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  buffer at pH 6.8 and 25 °C (Fig. 3). The peptide protons can be selectively monitored without detection of the nucleic acid protons in such a  $^{13}\text{C}$ -filtered NOESY experiment on the complex.

The expanded  $^{13}\text{C}$ -filtered NOESY contour plot exhibiting NOEs between the amide protons (7.2–9.2 ppm) and  $\text{H}\alpha$  protons (3.3–5.0 ppm) in the BIV Tat-TAR complex in  $\text{H}_2\text{O}$  buffer is plotted in Figure 3a. The connectivities involving amide protons with their own and adjacent  $\text{H}\alpha$  protons can be traced

**Table 2.** Statistics of NMR data and the four final structures of BIV Tat-TAR complex.

<b>NMR restraints</b>	
TAR RNA	
Distance restraints	691
Intra-residue	417
Sequential ( $ i-j =1$ )	226
Inter-strand	
(G4-U16 $\leftrightarrow$ A21-C31)	35
Loop (C17-U20)	120
U10 $\leftrightarrow$ Rest of RNA	9
U12 $\leftrightarrow$ Rest of RNA	2
Torsion restraints (6/ribose)	132
Tat peptide	
Distance restraints	259
Intra-residue	150
Sequential	46
Medium range ( $2 \geq  i-j  \geq 4$ )	25
Long range ( $ i-j  \geq 5$ )	38
Tat peptide-TAR RNA	
Distance restraints	102
U10 $\leftrightarrow$ peptide	10
<b>Structure statistics</b>	
NOE violations	
Number $> 0.2 \text{ \AA}$	$23.7 \pm 6.2^a$
Maximum violations ( $\text{\AA}$ )	$0.53 \pm 0.07$
r.m.s.d. of violations ( $\text{\AA}$ )	$0.047 \pm 0.007$
Deviations from ideal covalent geometry	
Bond length ( $\text{\AA}$ )	$0.013 \pm 0.004$
Bond angle ( $^\circ$ )	$3.017 \pm 0.018$
Improper ( $^\circ$ )	$1.913 \pm 0.173$
<b>Pairwise r.m.s.d. among the four final structures (<math>\text{\AA}</math>)</b>	
TAR RNA heavy atoms	
C6-G11, A13-U16, A21-G29	$1.11 \pm 0.06$
Tat peptide	
P69-I79	
Backbone (N, C $\alpha$ , C')	$0.49 \pm 0.08$
All heavy atoms	$1.03 \pm 0.07$
S65-R81	
Backbone (N, C $\alpha$ , C')	$1.48 \pm 0.25$
All heavy atoms	$1.90 \pm 0.32$
Tat peptide-TAR RNA <sup>b</sup>	
C6-G11, A13-U16, A21-G29, P69-I79	$1.16 \pm 0.08$

<sup>a</sup>Number of  $> 0.2 \text{ \AA}$  violations for the core region of the complex (C6-U16, A21-G29, P69-I79) =  $7.8 \pm 3.2$

<sup>b</sup>Superimposed on all RNA heavy atoms plus peptide backbone

from R68 to R81 for the 17-mer Tat peptide in the complex (Fig. 3a). There are breaks in the connectivities at the R68-P69-R70 step because of the intervening P69 residue, and at the R73-G74 step. These connectivities can be traced in other regions of the corresponding  $^{13}\text{C}$ -filtered NOESY spectrum of the complex in  $\text{D}_2\text{O}$  buffer (Supplementary materials). An NOE characteristic of a  $\beta$ -sheet is detected between the H $\alpha$  protons of T72 and R78 in the complex.

The expanded  $^{13}\text{C}$ -filtered NOESY contour plot exhibiting NOEs between amide protons in the symmetrical 7.2–9.2 ppm region in the complex in  $\text{H}_2\text{O}$  buffer is shown in Figure 3b. The connectivities between adjacent amide protons can be traced from G71 to R81. In essence there is a break in the connectivity at the

T72-R73 step and strong crosspeaks between adjacent amide protons at the K75-G76, G76-R77 and R80-R81 steps in the complex (Fig. 3b). The backbone amide and C $\alpha$  protons and the assignable side chain protons of the Tat peptide in the BIV Tat-TAR complex in  $\text{D}_2\text{O}$  buffer are given in the Supplementary materials.

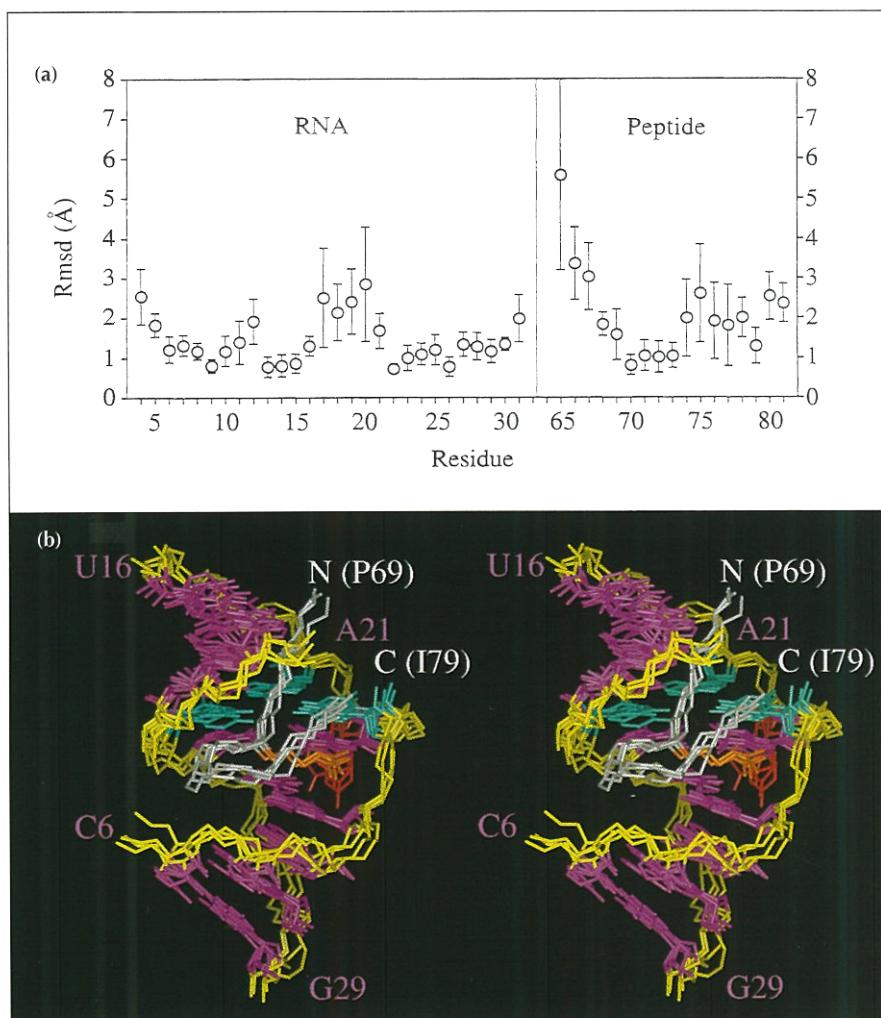
### Intermolecular restraints

We have identified 102 intermolecular restraints between the 17-mer Tat peptide and the 28-mer TAR RNA in the BIV Tat-TAR complex in solution. These intermolecular NOEs are listed in Table 1 and establish that R68, R70, R73 and R77 are involved in a significant number of NOEs with protons on the TAR RNA. In addition, T72 and I79 together with G71 and G74 exhibit a significant number of intermolecular NOEs. The behavior of bulge uracils U10 and U12 of the TAR RNA in the complex is distinctly different; U10 exhibits 10 intermolecular NOEs in contrast to none for U12 (Table 1). It should be noted that the major-groove edge protons (H5, H6) of U10 exhibit NOEs to the H $\alpha$ , H $\beta$ , H $\gamma$  and H $\delta$  protons of I79. These 102 intermolecular restraints are critical in guiding the molecular dynamics calculations towards a well converged structure of the BIV Tat-TAR complex in solution.

### Input restraints and structure calculations

Inter-proton distance restraints were quantified from NOE buildup curves in NOESY experiments on the BIV Tat-TAR complex in  $\text{H}_2\text{O}$  (four mixing times) and  $\text{D}_2\text{O}$  (five mixing times). Additional distance restraints were obtained from analysis of three-dimensional NOESY-heteronuclear multiple quantum coherence (HMQC) (120 ms mixing time) and four-dimensional HMQC-NOESY-HMQC (200 ms mixing time) spectra. Hydrogen-bond restraints were incorporated for the Watson-Crick base pairs in the helical stem of the RNA in the complex. Torsion-angle restraints were imposed on selected ribose sugars to maintain appropriate puckers. (For details, see Materials and methods.) The input restraint statistics, and their partitioning into several categories, are summarized in Table 2.

The structure calculations on the BIV Tat-TAR complex were carried out in three stages. Initially, the RNA and peptide were independently folded, starting from idealized models, using restrained molecular dynamics with a simulated annealing protocol. This provided four TAR RNA and seven Tat peptide structures. These TAR RNA and Tat peptide models were docked against each other using restrained molecular dynamics in two cycles of simulated annealing refinement (see Materials and methods). All 22 of the successfully completed docking trials exhibited common structural features. The peptide forms a  $\beta$ -turn that lies in the major groove of the RNA. The U•AU base triple was also observed in all 22 structures. Differences and NOE restraint violations in the family of structures were confined largely to the peptide side chains and to the amino (S65-P67) and carboxy (R80, R81) termini of the



**Fig. 4.** Evaluation of four distance-refined structures of the BIV Tat-TAR complex. (a) Pairwise all-heavy-atom r.m.s.d.'s ( $\text{\AA}$ ) in the atomic-coordinate positions calculated by residue for the four distance-refined structures. A least-squared all-heavy-atom superpositioning of all RNA heavy atoms and the peptide backbone for residues C6–G11, A13–U16, A21–G29, P69–I79 of the BIV Tat-TAR complex was performed before the calculation. The average values (open circles) are shown along with the standard deviations (vertical bars). (b) A stereopair of the four distance-refined structures of the core segment of the BIV Tat-TAR complex. The P69 to I79 domain of the BIV Tat peptide is shown in a white backbone with the side chain of R73 depicted in orange. The C6 to U16 and A21 to G29 segments of the BIV TAR RNA are shown with a yellow sugar-phosphate backbone and magenta bases. The U10•A13•U24 triple is shown in cyan and the external U12 base in red. The ends of the peptide and the RNA domains are labeled by residue number. This stereoview was generated by superpositioning all RNA heavy atoms and the peptide backbone for residues C6–G11, A13–U16, A21–G29, and P69–I79 of the complex. The backbone phosphate oxygens of the RNA have been deleted for clarity.

peptide and to the RNA loop (C17–U20). Four structures of the complex were chosen and analyzed. The structure closest to the average of these four was chosen as the representative structure for detailed description. This structure exhibited the lowest violations relative to the experimental data.

#### Structure analysis

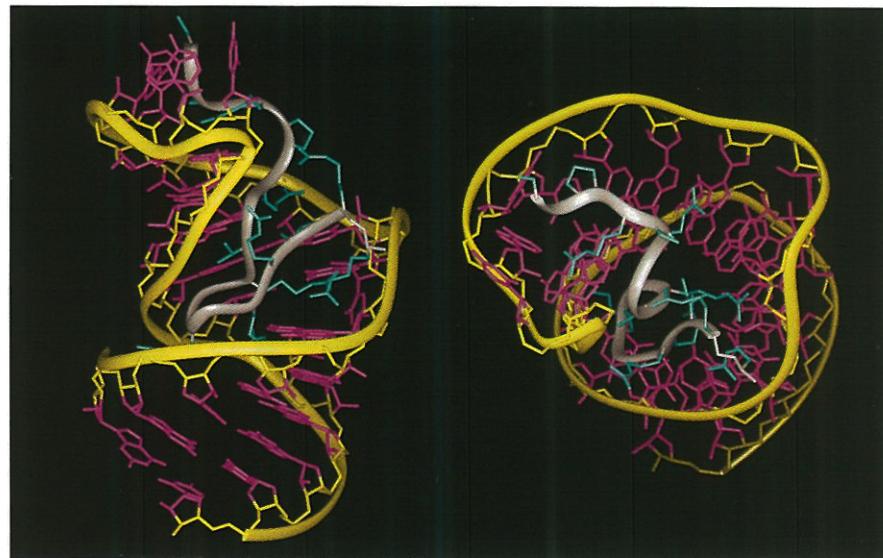
The root mean square deviation (r.m.s.d.) values at individual TAR RNA residues and at individual Tat peptide residues between the four distance-refined structures of the BIV Tat-TAR complex are plotted in Figure 4a. It is evident that the resolution is not uniform for all residues of the complex. Only those features that are accurately defined are described here.

A stereoview of the four superpositioned distance-refined structures of the BIV Tat-TAR complex looking normal to the helix axis is shown in the Supplementary materials. A stereoview of the four superpositioned distance-refined structures of the core domain of the BIV Tat-TAR complex looking at right angles to the helix axis is shown in Figure 4b. This view shows the backbone segment of the BIV Tat peptide from residues P69 to I79, including the side chain of R73, and the BIV TAR RNA from residues C6 to U16 and A21 to G29.

Two views of a representative distance-refined structure of the entire BIV Tat peptide–TAR RNA complex are shown in Figure 5. The peptide and nucleic acid backbones are shown in white and yellow ribbons, respectively, with these views emphasizing the deep penetration of the Tat peptide into the widened major groove of the TAR RNA. The key elements of the solution structure of the BIV Tat-TAR complex are the edge-on positioning of the Tat peptide in a  $\beta$ -hairpin fold in the widened major groove formed by the continuous helical stem of the TAR RNA. The bulged U10 and U12 residues are clearly important (Fig. 4b). The U10 base participates in buttressing the Watson–Crick A13•U24 base pair whereas U12 loops out into the minor groove.

Two close-up views of the key intermolecular contacts of residues R73 and I79 of the Tat peptide with the (G9–U10–G11–U12–A13)•(U24–C25–C26) TAR RNA segment in one representative distance-refined structure of the complex are shown in Figure 6. These views identify the packing alignment of the side chain of I79 (shown in dotted white spheres) relative to the major-groove edge of the U10•A13•U24 triple, and the intermolecular hydrogen bonding alignments of the side chain of R73 with the major-groove edge of G11 and the phosphate backbone at the G9–U10 step (Fig. 6b).

**Fig. 5.** Two views of a representative distance-refined structure of the BIV Tat-TAR complex. The BIV Tat peptide, depicted with a white backbone (ribbon representation) and cyan side chains, extends from residues S65 to R81. The BIV TAR RNA, depicted with a yellow sugar-phosphate backbone (ribbon representation) and magenta bases, extends from G4 to C31. Left, view looking into the major groove; right, view looking down the major groove.



A view of the intermolecular interactions between the P69-R70-G71-T72 segment of the Tat peptide and the (A13-G14-C15-U16)•(A21-G22-C23-U24) TAR RNA segment in one representative distance-refined structure of the complex is shown in Figure 7a. The intermolecular hydrogen bonds involving the Tat peptide backbone and the side chains of R70 and T72 with the major-groove edge and backbone phosphates of the TAR RNA are identified.

A view of the location of the (G71-T72-R73)•(R77-R78-I79)  $\beta$ -sheet segment of the Tat peptide and the (G9-U10-G11)•(C25-C26) TAR RNA segment in one representative distance-refined structure of the complex is shown in Figure 7b. This view shows intermolecular hydrogen bonds involving the Tat peptide backbone and

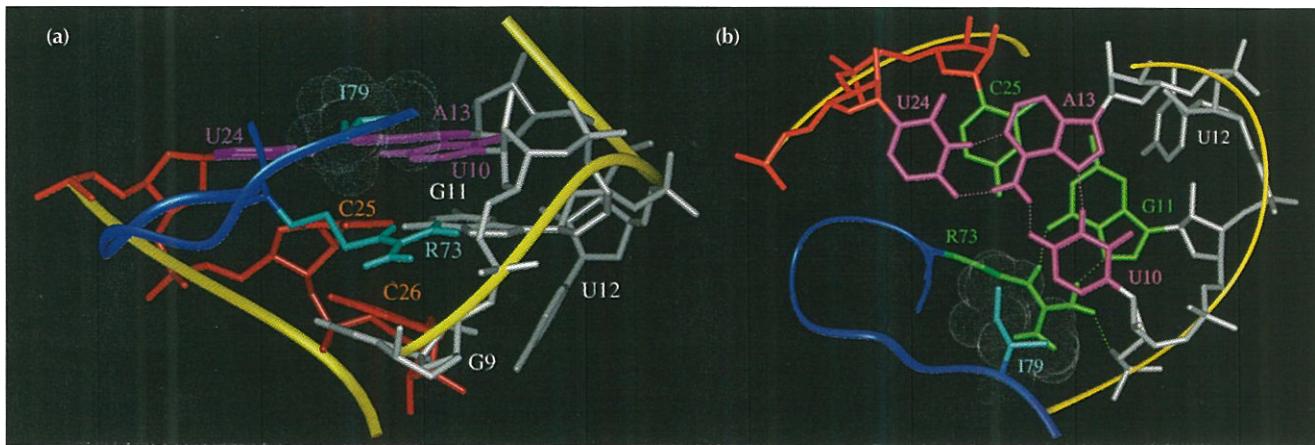
the side chains of R73 and I79 with the major-groove edge and the backbone phosphates of the TAR RNA in the BIV Tat-TAR complex.

## Discussion

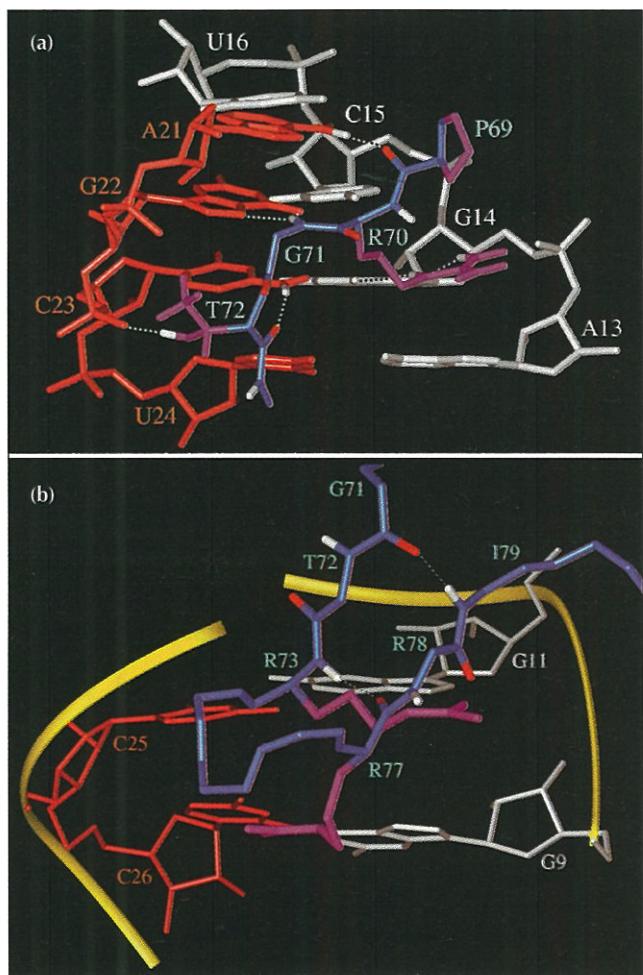
### Global features of bound TAR RNA

The stem segment of the 28-mer TAR RNA adopts a continuous stacked helix starting from the G4•C31 pair at one end of the stem to the U16•A21 pair adjacent to the hairpin loop at the other end of the stem. It is significant that the G9•C26 and G11•C25 base pairs that flank the U10 bulge stack on each other as do the G11•C25 and A13•U24 base pairs that flank the U12 bulge.

The two stems of the Tar RNA helix are bent by  $\sim 40^\circ$  in the complex. Measurements of roll angles at individual



**Fig. 6.** Intermolecular interactions between the R73 and I79 residues of the Tat peptide and the (G9-U10-G11-U12-A13)•(U24-C25-C26) TAR RNA segment in a representative distance-refined structure of the BIV Tat-TAR complex. **(a)** A view normal to the helix axis. The RNA backbones are shown as yellow ribbons with bases from the individual strands represented in white and red colors. The U10•A13•U24 triple is colored in magenta. The U12 base is looped out of the helix. The reverse  $\beta$ -sheet of the Tat peptide is shown by a blue ribbon with the side chain of R73 (cyan) positioned opposite the major-groove edge of G11 in the complex. The side chain of I79 is shown as a cyan-colored stick representation and as a white dotted surface. **(b)** A view down the helix axis showing the U10•A13•U24 base triple (pink) and the Watson-Crick G11•C25 base pair in green. The hydrogen-bonding alignments in the triple are indicated by dotted lines. The side chain of R73 (green) is anchored through hydrogen bond formation with the major groove edge of G11 and the phosphate backbone at the G9-U10 step. The side chain of I79 (cyan) is shown in a white space-filling representation and is positioned along one edge of the U10 base involved in the U10•A13•U24 base triple in the complex.



**Fig. 7.** Intermolecular interactions in segments of a representative distance-refined structure in the BIV Tat-TAR complex. (a) Interactions between the P69-R70-G71-T72 residues on the Tat peptide and the (A13-G14-C15-U16)-(A21-G22-C23-U24) TAR RNA segment. The peptide backbone is in blue with the side chains of R70 and T72 in magenta. Bases from the individual RNA strands are shown in white and red. Intermolecular hydrogen bonds are indicated by dashed lines and involve both backbone and side chains on the Tat peptide and base-pair edges and phosphate backbone on the TAR RNA. (b) Intermolecular interactions between the (G71-T72-R73---R77-R78-I79) residues on the Tat peptide and the (G9-G10)-(C25-C26) TAR RNA segment. The peptide backbone is in blue with the side chains of R73 and R77 in magenta. Bases from the individual RNA strands are shown in white and red. Hydrogen bonds across the  $\beta$ -sheet are indicated by dashed white lines.

steps along the helix (given in the Supplementary materials) show that the bend is localized primarily at the adjacently stacked C8•G27, G9•C26 and G11•C25 base pairs in the complex. The helix bends diagonally towards the A21-C31 strand and into the major groove as reflected by a large twist at the (G9•C26)-(G11•C25) step accompanied by a roll at the (C8•G27)-(G9•C26) and (G9•C26)-(G11•C25) steps.

The twist angles between adjacent base pairs for the segment centered about the bulged uracils in the BIV Tat-TAR complex are listed in the Supplementary materials. There is significant overwinding of the TAR

helix between the G9•C26 and G11•C25 base pairs (twist =  $52 \pm 3^\circ$ ) in the complex.

The width of the major groove increases dramatically proceeding along the stem towards the hairpin loop segment of the TAR RNA in the BIV Tat-TAR complex. The shortest phosphorus-phosphorus distance across the groove increases from 9.0 Å (separation between C6 and C23) to 16.6 Å (separation between G11 and A21). This increase in major-groove width greatly facilitates the insertion of the Tat peptide into this groove to form the complex.

#### Global features of bound Tat

A segment of the 17-mer Tat peptide adopts a  $\beta$ -hairpin structure defined by an antiparallel  $\beta$ -sheet between the G71-T72-R73 and R77-R78-I79 segments. We detect the characteristic NOE patterns across the  $\beta$ -sheet which is linked by a three-residue turn. The  $\beta$ -sheet is approximately parallel to the major groove with its G71-T72-R73 strand directed towards the interior of the major groove while the R77-R78-I79 strand is directed towards the exterior of the groove in the complex (Fig. 7b). There is also an abrupt turn in the peptide backbone at the  $\text{C}\alpha$  positions of R73 and R77 across the  $\beta$ -sheet (Fig. 7b) which accommodates the peptide in the major groove of the bent helical stem of the TAR RNA. The R73 and R77 side chains that extend from these branch points anchor the  $\beta$ -sheet to the floor of the major groove in the complex (Fig. 7b). The remaining segments of the Tat peptide adopt an extended structure with few intermolecular restraints to define the amino- and carboxy-terminal ends of the chain in the complex.

#### Global views of Tat-TAR interactions

The  $\beta$ -hairpin fold of the 17-mer Tat peptide is bound in the major groove of the 28-mer TAR RNA in the complex (see Fig. 8). The amino- and carboxy-terminal ends of the peptide project towards the hairpin loop end of the TAR RNA. The Tat peptide penetrates deep into the major groove with the  $\beta$ -hairpin segment completely engulfed within the diameter of the TAR RNA helix. Recognition is associated with specific peptide-backbone and side-chain hydrogen bonds as well as hydrophobic contacts with the base-pair edges and the sugar-phosphate backbone of the TAR RNA. It is interesting to note that the amino-terminus of the Tat peptide interacts extensively with the hairpin loop of the TAR RNA.

#### U10•A13•U24 base triple

A base triple is formed between the bulged U10 base and the Watson-Crick A13•U24 base pair in the distance-refined structures of the complex. This interaction is stabilized by two hydrogen bonds involving the Watson-Crick edge of U10 and the Hoogsteen edge of A13 (Fig. 6b) similar to those observed for the T•A•T triples in the pyrimidine-purine-pyrimidine class of DNA triplets [21]. The key experimental features in support of this base triple are the direct observation of the U10 imino proton (13.2 ppm at  $-2^\circ\text{C}$ ) (Fig. 2a) and

the observed NOE between it and the H8 proton of A13 (boxed peak in Fig. 2c; see also Supplementary materials).

Base triple formation buttresses the A13•U24 base pair, which is centrally positioned in the complex and proximal to the helical bend site. In addition, the base triple positions U10 for a key intermolecular contact with the side chain of I79. TAR RNA mutants where U10 is replaced with C or A have a 50% reduction in binding affinity compared to wild type, whereas replacement of the Watson-Crick A13•U24 base pair with its U•A counterpart results in a 75% reduction in binding affinity [9]. It is evident from the structure that formation of the U10•A13•U24 triple facilitates the adoption of a specific orientation by the G9-U10 backbone phosphate, which participates in a key intermolecular interaction with the side chain of R73. Disruption of the U•A•U base triple would modulate this interaction and affect the binding affinity of the BIV Tat-TAR complex.

#### Bulged U12 residue

The bulged U12 base is looped out into the minor groove in our distance-refined structures of the BIV Tat-TAR complex. We do not detect its imino proton in the 12.0–15.0 ppm region in the complex at low pH and low temperature. The base protons of U12 do not exhibit intermolecular NOEs to Tat peptide protons in the complex. These results are consistent with the observation that replacement of U12 with cytosine results in a moderate reduction of binding affinity to 75% of wild type [9]. It appears that the major groove widens significantly in the BIV Tat-TAR complex as a result of the looped out U12 residue, consistent with earlier predictions that linked major groove width to perturbations in the RNA helix [22].

#### R73-G11-phosphate intermolecular interaction

The side chain of R73 is positioned in the plane of the G11•C25 base pair and directed towards the sugar-phosphate backbone at the G9-U10 step in the complex (Fig. 6a). The guanidinium side chain of R73 forms a pair of intermolecular hydrogen bonds with the major-groove edge of G11 and with the phosphate oxygen at the G9-U10 step in the complex (Fig. 6b). Thus, recognition is dependent both on the nature of the residue at position 11 (sequence specificity) and on the conformation of the sugar-phosphate backbone (conformational specificity). Indeed, replacement of the G11•C25 base pair with either its C•G counterpart or a G•U mismatch pair results in a reduction in the binding affinity to < 5% of wild type [9]. Furthermore, the ethylation of the G9-U10 phosphate hinders the binding of Tat to TAR RNA [9]. Such arginine-guanine interactions have been observed previously in a zinc finger-DNA complex [23] and a similar arginine-guanine-phosphate interaction has been proposed in an arginineamide-HIV TAR complex [6]. It is probable that the phosphate oxygens face into the RNA, which would maximize their interaction with the guanidinium group of R73 in the complex.

#### I79-U10 intermolecular interaction

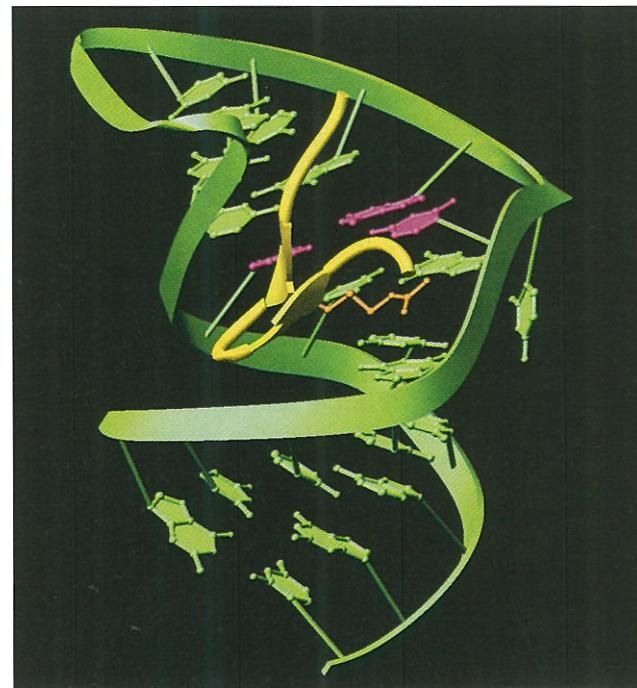
Replacement of I79 with other hydrophobic amino acids results in a significant loss in TAR-binding affinity, indicating that it is important in modulating binding in solution. The side chain of I79 is positioned in the plane of the U10•A13•U24 base triple (Fig. 6a) and directed towards the hydrophobic (H5, H6) edge of the U10 residue in our distance-refined structures of the complex (Fig. 6b). We have observed strong intermolecular NOEs between the H5, H6 protons of U10 and the side chain protons of I79 (Table 1) that are consistent with this. The complementary arrangement of intermolecular packing seen in the structure can be disrupted by substituting isoleucine with other hydrophobic amino acids, accounting for the loss of binding affinity seen experimentally [10].

#### Intermolecular contacts involving T72

Threonine 72 is another key amino acid in BIV Tat; replacement of this residue with alanine results in a reduction in binding affinity to < 3% of wild type for formation of the BIV Tat-TAR complex [10]. The T72 residue forms a pair of intermolecular hydrogen bonds to the TAR RNA, with the backbone carbonyl oxygen hydrogen bonding to the amino group of C23 and the side chain hydroxyl hydrogen bonding to the phosphate oxygen at the G22-C23 step in the complex (Fig. 7a).

#### Intermolecular contacts involving R70 and R77

Replacement of R70 and R77 with lysines in BIV Tat results in a reduction in binding affinity to < 3% of wild



**Fig. 8.** Ribbon representation of the complex of the BIV Tat-TAR recognition domains. The Tat peptide folded in a  $\beta$ -hairpin is shown in yellow with arrows indicating the  $\beta$ -sheet segment of the polypeptide chain. The side chain of R73 is shown in orange. The RNA is shown in green with the bases of residues U10, A13 and U24, which form the base triple, highlighted in magenta.

type [10]. We observe sequence-specific contacts between the guanidinium side chains of these arginines and the major-groove edge of guanines in our distance-refined structures of the complex. Specifically, R77 pairs with G9 through two hydrogen bonds (Fig. 7b) and R70 pairs with G14 through two bifurcated hydrogen bonds (Fig. 7a). Such arginine–guanine bonding interactions have been proposed in the literature [24]. It is worth noting that replacement of the G14•C23 base pair with its C•G counterpart or with a G•U mismatch results in a decrease in binding affinity to < 3 % of wild type for complex formation [9]. This can be readily explained by the observed intermolecular hydrogen bonding interactions for the R70–G14 alignment and the T72–C23 alignment (outlined in the previous section) upon complex formation. The conformation of the R77 side chain is not very well converged, indicating that it may hydrogen bond at the G9•C26 step in any of several orientations.

#### Intermolecular contacts involving G71, G74 and G76

Replacement of residues G71, G74 and G76 in the Tat peptide with alanines results in a pronounced reduction in the binding affinity to < 12 % of wild type for formation of the BIV Tat–TAR complex [10]. Our solution structures provide an explanation for the uniqueness of glycine residues at these positions in the Tat sequence and their role in complex formation. Residues G71 and G74 are located along the inwardly-directed strand of the anti-parallel  $\beta$ -sheet and along the turn segment, respectively, of the Tat peptide in the complex. The intermolecular hydrogen bond between the amide proton of G71 and the major-groove edge of guanine 22 (Fig. 7a) leaves no room for introduction of larger side chains which would point directly into the base and sugar rings of A21. Glycine 74 is packed against the sugar-phosphate backbone of the C23-U24-C25 segment of TAR RNA, leaving no room for accommodating a larger side chain in the complex. Glycine 76 is located along the outwardly-directed strand of the antiparallel  $\beta$ -sheet and its associated turn in the complex. There is sufficient room to accommodate the methyl side chain of an alanine residue but nothing larger in the solution structure of the complex. We also note that positive  $\phi$  values are observed for G74 ( $110 \pm 8^\circ$ ) and G76 ( $71 \pm 6^\circ$ ) in the solution structure of the complex. This domain is sterically allowed only for glycine residues in the Ramachandran plot and hence the two glycines facilitate the formation of the turn in the peptide backbone.

#### Peptide interactions with the C17–U20 hairpin loop of TAR

The C17-A18-U19-U20 hairpin loop of TAR RNA in the BIV Tat–TAR complex is poorly defined (high r.m.s.d. values in Fig. 4a) despite a total of 120 NOEs between protons within the loop segment. An analysis of the entire family of 22 distance-refined structures of the docked Tat–TAR complex reveals that there are two dominant conformers for the RNA tetraloop and two families of conformers for the amino-terminal (S65–G66–P67–R68) segment of the Tat peptide in the complex. The

four structures presented in this paper have a mostly extended conformation of this amino-terminal tetrapeptide segment, whereas in five other structures (data not shown) the amino-terminus forms a turn at P67 stabilized by a hydrogen bond between the amide proton of G66 and the carbonyl oxygen of R68. This turn conformation folds the amino-terminus of the peptide and completely engulfs it in the major groove of the RNA. The side chain of R68 stretches over the plane of the U16•A21 base pair reaching to the sugar-phosphate backbone of residues A18–U19–U20 in both conformations of the amino-terminal peptide in the complex. It will be necessary to characterize Tat–TAR complexes containing extensions of the amino-terminus to definitively differentiate between extended and folded conformations for the S65 to R68 domain of the Tat peptide.

#### Molecular recognition in HIV and BIV Tat–TAR

The BIV TAR RNA bears significant similarity to the HIV TAR RNA; both consist of pyrimidine bulges that divide an otherwise regular RNA duplex stem into two halves. In BIV TAR, two bulged uracils are interspersed with a G•C base pair, whereas HIV TAR has a stretch of three pyrimidines forming a bulge with a G•C base pair immediately 3' to the bulge. The next base-pair step above the G•C pair is an A•U pair in both sequences.

The principal similarity between the Tat proteins from HIV and BIV is that both contain an arginine-rich basic domain that interacts with the TAR RNA. Based on several biochemical experiments on the HIV Tat–TAR system and NMR data from the arginineamide–TAR complex, it has been proposed that the guanidinium group of an arginine in the Tat protein forms hydrogen bonds with the major-groove edge of a G•C pair [5,6]. The model further proposed that the 5' U of the U–C–U bulge in HIV TAR reaches up to form a base triple with the A•U pair at the expense of significant distortion to the RNA backbone [7]. An alternative view based on a detailed structural characterization from NMR and molecular dynamics studies concluded that all bases in the tri-residue bulge are unstacked with the 5' U positioned in the major groove in arginineamide–TAR and Tat peptide–TAR complexes in solution [8]. Therefore, there is disagreement about the specific location and role of the bulged uracils in the HIV Tat–TAR complex [6,8], and this issue remains to be resolved through additional experimentation.

In the structure of the BIV Tat peptide–TAR RNA complex, we observe that the sugar-phosphate backbone at the G9–U10 step is stretched over the next base step (G11•C25 base pair), enabling U10 to reach the A13•U24 base plane to form a U10•A13•U24 base triple (Fig. 6a). This brings the phosphate between G9 and U10 almost into the plane of the G11•C25 base pair and in close proximity to the guanidinium group of R73, facilitating the formation of an additional R73–RNA hydrogen bond, with the O5' oxygen of this phosphate group (Fig. 6b). This arginine–guanine–phosphate motif readily explains the previous demonstration that ethylation of

the phosphate between G9 and U10 interferes with the peptide binding in the BIV Tat-TAR complex [9].

The formation of a U•AU base triple by extending a bulged uracil over a G•C base pair, while an arginine side chain forms hydrogen bonds with the guanine and the stretched phosphate, constitutes a novel feature of peptide-RNA recognition. Our studies on the BIV Tat-TAR system conclusively establish the presence of this structural motif, which was originally postulated based on structural studies on the HIV arginineamide-TAR system [5,6]. Indeed, this may be the only common structural motif for peptide-RNA recognition in BIV and HIV Tat-TAR systems.

## Significance

**The BIV Tat-TAR complex is an attractive model system for understanding peptide-RNA recognition in HIV Tat-TAR. Our structural studies on the BIV Tat-TAR complex have identified a new structural motif in protein-RNA recognition. This motif positions an anti-parallel  $\beta$ -sheet edgewise in a widened RNA major groove, and shows that an arginine-guanine-phosphate interaction constitutes the major determinant of binding specificity. The TAR RNA, which contains two single-uracil bulges separated by a G•C base pair, forms a continuously stacked helix with a ~40° bend centered about the bulged sites in the complex. Specific intermolecular hydrogen bonding and van der Waals interactions identified in the BIV Tat-TAR complex readily account for the consequences of peptide and RNA mutations reported previously [9,10]. Our structural studies on the BIV Tat-TAR complex and earlier studies on the related HIV counterpart [5,6] establish that the binding specificities share a common principle of arginine-guanine-phosphate interaction achieved through a change in backbone conformation of bulged residues associated with base-triple formation.**

## Materials and methods

### Sample preparation

The 28-mer TAR RNA was prepared by *in vitro* transcription using T7 polymerase and synthetic DNA templates [25,26]. The T7 polymerase was purified from an overexpressing *E. coli* strain [27]. The DNA template oligonucleotides were chemically synthesized on an Applied Biosystems DNA synthesizer and purified using 15% denaturing polyacrylamide-gel electrophoresis (containing 7 M urea). The TAR RNA product from large scale transcription (50 ml) was ethanol precipitated and purified using 20% denaturing polyacrylamide-gel electrophoresis (containing 7 M urea). The TAR RNA product band was monitored by UV shadowing, cut and electroeluted using an elutrap (Schleicher and Schuell, Inc.). The eluted RNA was ethanol precipitated and equilibrated with NMR buffer (10 mM phosphate, 0.1 mM EDTA, pH 6.8) by spinning through an Amicon centricon (3 kDa cut off) with the

procedure repeated five to seven times. A 50-ml transcription yielded about 8 mg of 28-mer TAR RNA. The same transcription conditions were used for the uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled TAR RNA except that  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled nucleoside triphosphates (NTPs) were used instead of unlabeled NTPs. The  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled NTPs were isolated from an *E. coli* strain overexpressing rRNA grown in minimal medium containing  $^{13}\text{C}$  glucose and  $^{15}\text{N}$  ammonium sulfate as carbon and nitrogen sources [15].

The 17-mer Tat peptide was chemically synthesized and purified in the Institutional Microchemistry Core Facility. The peptide was purified by high-pressure liquid chromatography to about 95% purity based on a mass spectroscopic analysis.

The 1:1 Tat-TAR complex was formed by gradually titrating the Tat peptide at pH 6.8 to a TAR RNA solution in phosphate buffer (10 mM phosphate at pH 6.8). The additions were monitored by recording the imino proton NMR spectral region of the TAR RNA.

### NMR spectroscopy

NMR spectra were collected on Varian Unity Plus 600 MHz and 500 MHz NMR spectrometers. All NMR experiments were acquired in the phase-sensitive mode. The quadrature detection in the indirect dimensions in homonuclear and heteronuclear multi-dimensional experiments was accomplished using the States-TPPI method [28]. The NMR data were processed on SUN work stations using VNMR software and analyzed using the FELIX program (Biosym). The three- and four-dimensional NMR data were processed and analyzed using NMRPipe and PIPP programs [29,30]. Proton chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) while carbon and nitrogen chemical shifts were indirectly referenced to DSS and ammonia, respectively [31]. Details of two-, three- and four-dimensional NMR experiments are outlined in the Supplementary material.

### Input restraints

Input distance restraints involving non-exchangeable protons in the complex were derived from the NOE buildup experiments at 60, 90, 130, 180 and 250 ms mixing times on the Tat-TAR complex in  $\text{D}_2\text{O}$  solution at 25 °C. Interproton distance restraints involving exchangeable protons in the complex were derived from the NOE buildup experiments at 50, 80, 120 and 180 ms mixing times on the complex in  $\text{H}_2\text{O}$  solution at 25 °C. Volume integrals of each non-overlapped NOE crosspeak were measured and plotted against the mixing time as buildup curves and fitted to second-order polynomials. Interproton distances were calculated from the initial buildup rates using the slopes of the cytosine and uracil H5-H6 intrar residue crosspeaks as a reference, calibrated to a 2.45 Å distance. The distance restraints were assigned a uniform lower and upper bound of 30% of the computed distance. For peaks with intensities that were too weak, yielding distances greater than 5.0 Å by this calibration mechanism, a cutoff distance of 5.0 Å was imposed with bounds at 30%. For crosspeaks involving methyl and other unresolved protons, the volume integral was divided by the number of protons before curve fitting. Restraints corresponding to crosspeaks arising from non-stereospecific assignments were treated with  $<\mathbf{r}^6>^{-1/6}$  averaging. The cutoff distance for peaks involving methyl protons was 5.5 Å. Crosspeaks in the  $\text{D}_2\text{O}$  data set that were not amenable to such analysis due to reasons of overlap were manually classified into three categories with corresponding

distances and bounds: strong ( $2.7 \pm 0.8 \text{ \AA}$ ), medium ( $4.0 \pm 1.0 \text{ \AA}$ ), and weak ( $4.5 - 1.0 / + 2.0 \text{ \AA}$ ).

Additional restraints were obtained from quantitative analysis of three dimensional NOESY-HMQC spectra. Peak heights on both proton dimensions were calibrated against those corresponding to pyrimidine H5–H6 crosspeaks, employing an  $r^{-1/6}$  approximation. Crosspeaks that could only be resolved in the four dimensional HMQC-NOESY-HMQC spectra were qualitatively classified into three categories and assigned corresponding restraints: strong ( $2.7 \pm 0.8 \text{ \AA}$ ), medium ( $4.0 - 2.0 / + 1.0 \text{ \AA}$ ) and weak ( $4.5 - 2.5 / + 2.0 \text{ \AA}$ ).

Distance restraints were imposed on the Watson–Crick hydrogen bond participants in the RNA base pairs. These distances were taken from the standard base-pair geometries of nucleic acids and were assigned an upper and lower bound of  $\pm 0.1 \text{ \AA}$ .

Torsion restraints on the endocyclic torsion angles within ribose sugars of RNA were obtained from analysis of COSY experiments. The presence of strong H1'-H2' crosspeaks was interpreted as C2'-*endo* ( $P = 144\text{--}180^\circ$ ) sugar pockers, whereas the absence of the crosspeak was interpreted as C3'-*endo* ( $P = 0\text{--}36^\circ$ ) sugar pockers and translated to corresponding values and ranges for the sugar endocyclic torsion angles [32]. All other residues were left unrestrained during the computations.

#### *Starting structures of TAR RNA and Tat peptide*

The initial model for the 28-mer TAR RNA was generated in standard A-form. A continuous A-form duplex was created for residues G4–U16 and A21–C31. The adenine residues opposite U10 and U12 were deleted while both the bulge uracils were retained in standard helical geometry. The C17–U20 loop was extended from U16 as a single strand A-form helix. Energy minimization by the conjugate-gradient method was used to bring about favorable covalent geometry. The starting model of the Tat peptide was constructed in completely extended geometry.

#### *Distance-restrained molecular dynamics*

Restrained molecular dynamics calculations were performed on the Tat–TAR complex with a simulated annealing protocol in vacuum with a distance-dependent dielectric constant in three phases using the program X-PLOR [33] on Silicon Graphics Power Challenge L computers. In the first phase, the TAR RNA and the Tat peptide were independently refined against the NMR restraints corresponding to the respective components in the complex. The van der Waals interactions of the bulge residues (U10 and U12 to the rest of the RNA) and within the loop (C17–U20) were turned off initially, and scaled back in while the RNA was being equilibrated at 300 K. This yielded four RNA and seven peptide structures. In the second phase, each peptide structure was placed  $\sim 30 \text{ \AA}$  away from each RNA structure and then docked by a gradual scaling up of the inter-proton distance restraints. The intra-loop (C17–U20) and all intermolecular van der Waals interactions were turned off initially and scaled back in while the complex was being equilibrated at 300 K. Electrostatic interactions were turned off completely. In the third phase, the protocol of the second phase was repeated on the docked complex. The electrostatic interactions were scaled up simultaneously with the van der Waals interactions.

In each phase, the dynamics calculations were initiated at 5 K and the temperature was gradually raised to 1000 K in 5 ps

and then equilibrated for 1 ps. Then the force constants for the distance restraints were scaled up from 0 to a final value of  $30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  over 6 ps. During the heating and high temperature stages, while there were no NMR restraints on the system, the base atoms of the RNA Watson–Crick base pairs were held coplanar using planarity restraints with a force constant of  $40 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . These planarity restraints were scaled off gradually as the NOE distance restraints were scaled up. The system was then allowed to evolve for 20 ps at 1000 K before slow cooling to 300 K in 14 ps and was then equilibrated for another 10 ps. Selected van der Waals and electrostatic interactions that had been turned off were now brought on gradually over 12 ps and the system was equilibrated for another 10 ps. The coordinates saved every 0.5 ps for the last 4 ps were averaged and the resulting structure was subjected to conjugate gradient energy minimization until a gradient of  $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  was reached. All dynamics were carried out with a time step of 1 fs and the force constants on the Watson–Crick hydrogen-bond-related distance restraints were maintained at  $60 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  throughout. The force constant on all torsion restraints was  $50 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ .

Of the 28 attempts to dock the peptide on the RNA, 22 runs were completed successfully. Of these, four were selected based on the criteria of acceptable covalent geometry and low restraint violations and analyzed.

#### *Structure analysis*

The helicoidal parameters of the RNA in the final structures were analyzed with the program CURVES [34,35]. The peptide structure was analyzed using the program package RIBBONS [36]. Color figures were prepared using Insight II (Biosym Technologies, Inc., San Diego, CA) and RIBBONS.

#### *Coordinates deposition*

The coordinates of the BIV Tat–TAR complex (accession number not yet available) have been deposited with the Protein Data Bank, Brookhaven National Laboratory, Upton, New York, 11923, USA, from whom copies can be obtained.

#### *Supplementary material available*

Details of the NMR experiments and the relevant references. Four tables listing proton (S1) and carbon (S2) chemical shifts of TAR RNA and proton chemical shifts of Tat peptide (S3) and twist and roll parameters at the central steps of TAR RNA (S4) in the BIV Tat–TAR complex. Fifteen figures showing imino proton spectra of free TAR RNA and TAR RNA bound to Tat peptide (S1), expanded NOESY contour plots in  $\text{H}_2\text{O}$  solution (S2) and  $\text{D}_2\text{O}$  solution (S3), expanded NOESY contour plot of the natural and U12 to C12 mutant complex in  $\text{D}_2\text{O}$  solution (S4), expanded  $^1\text{H}$ – $^{13}\text{C}$  HCCH-TOCSY contour plot in  $\text{D}_2\text{O}$  solution (S5), expanded  $^1\text{H}$ – $^{13}\text{C}$ – $^{31}\text{P}$  HCP contour plot in  $\text{D}_2\text{O}$  solution (S6), expanded  $^1\text{H}$ – $^{13}\text{C}$  NOESY-HMQC contour plot in  $\text{D}_2\text{O}$  solution (S7), expanded  $^1\text{H}$ – $^{13}\text{C}$ – $^1\text{H}$  HMQC-NOESY-HMQC contour plot in  $\text{D}_2\text{O}$  solution (S8), expanded DQF-COSY contour plot in  $\text{D}_2\text{O}$  solution (S9), expanded  $^{13}\text{C}$ -filtered NOESY contour plot in  $\text{D}_2\text{O}$  solution (S10), expanded DQF-COSY contour plot of peptide fingerprint region in  $\text{D}_2\text{O}$  solution (S11), color plate of four superimposed structures of the Tat–TAR complex including amino-acid side chains (S12), schematic of the  $\beta$ -sheet and attached loop together with NOE patterns (S13), shortest phosphate–phosphate separation across the major groove (S14) and color plate showing the R70–G14 and R77–G9 hydrogen bonding alignments in the BIV Tat–TAR complex (S15).

**Acknowledgements:** This research was supported by Start-Up funds to D.J.P. from the Memorial Sloan-Kettering Cancer Center. R.A.K. is a postdoctoral fellow of the Miriam and Benedict Wolf Cancer Research Fund. We thank Weijun Xu for preparation of T7 polymerase and uniformly labeled  $^{13}\text{C}$ ,  $^{15}\text{N}$ -nucleoside triphosphates. We acknowledge the help of Drs Radovan Fiala and David Live in our laboratory in the design, data collection and processing of three- and four-dimensional NMR data sets. We are grateful to Drs Frank Delaglio and Daniel Garrett of the Laboratory of Chemical Physics, NIDDK, NIH for making NMR software available for the processing of three- and four-dimensional NMR data sets and for valuable suggestions concerning its use.

## References

- Nagai, K. (1992). RNA-protein interactions. *Curr. Opin. Struct. Biol.* **2**, 131–137.
- Nagai, K. & Mattaj, I.W., eds. (1994). *RNA-Protein Interactions*. IRL Press, New York.
- Karn, J., Gait, M.J., Churcher, M.J., Mann, D.A., Mikaelian, I. & Pritchard, C. (1994). Control of human immunodeficiency virus gene expression by the RNA-binding proteins Tat and Rev. In *RNA-Protein Interactions*. (Nagai, K. & Mattaj, I.W., eds), pp 193–220, IRL Press, New York.
- Frankel, A.D. (1994). Using peptides to study RNA-protein recognition. In *RNA-Protein Interactions*. (Nagai, K. & Mattaj, I.W., eds), pp 221–247, IRL Press, New York.
- Calnan, B.J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A.D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science* **252**, 1167–1171.
- Puglisi, J.D., Tan, R., Calnan, B.J., Frankel, A.D. & Williamson, J.R. (1992). Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257**, 76–80.
- Puglisi, J.D., Chen, L., Frankel, A.D. & Williamson, J.R. (1993). Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Acad. Sci. USA* **90**, 3680–3684.
- Aboul-ela, F., Karn, J. & Varani, G. (1995). The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313–332.
- Chen, L. & Frankel, A.D. (1994). An RNA-binding peptide from bovine immunodeficiency virus Tat protein recognizes an unusual RNA structure. *Biochemistry* **33**, 2708–2715.
- Chen, L. & Frankel, A.D. (1995). A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl. Acad. Sci. USA* **92**, 5077–5081.
- Wimberly, B., Varani, G. & Tinoco, Jr., I. (1991). Structural determinants of RNA function. *Curr. Opin. Struct. Biol.* **1**, 405–409.
- Varani, G. & Tinoco, Jr., I. (1991). RNA structure and NMR spectroscopy. *Quart. Rev. Biophys.* **24**, 479–532.
- Moore, P.B. (1993). Recent RNA structures. *Curr. Opin. Struct. Biol.* **3**, 340–344.
- Moore, P.B. (1995). Determination of RNA conformation by nuclear magnetic resonance. *Accounts Chem. Res.* **28**, 251–256.
- Nikonowicz, E.P., Sirr, A., Legault, P., Jucker, F.M., Baer, L.M. & Pardi, A. (1992). Preparation of  $^{13}\text{C}$  and  $^{15}\text{N}$  labelled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res.* **20**, 4508–4513.
- Batey, R.T., Inada, M., Kujawinski, E., Puglisi, J. & Williamson, J.R. (1992). Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res.* **20**, 4515–4523.
- Nikonowicz, E.P. & Pardi, A. (1993). An efficient procedure for assignment of the proton, carbon and nitrogen resonances in  $^{13}\text{C}$ / $^{15}\text{N}$  labeled nucleic acids. *J. Mol. Biol.* **232**, 1141–1156.
- Dieckmann, T. & Feigon, J. (1994). Heteronuclear techniques in NMR studies of RNA and DNA. *Curr. Opin. Struct. Biol.* **4**, 745–749.
- van de Ven, F.J. & Hilbers, C.W. (1988). Nucleic acids and nuclear magnetic resonance. *Eur. J. Biochem.* **7**, 1–38.
- Wuthrich, K. (1986). *NMR of Proteins and Nucleic Acids*. John Wiley & Sons, New York.
- Moser, H.E. & Dervan, P.B. (1987). Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650.
- Weeks, K.M. & Crothers, D.M. (1993). Major groove accessibility of RNA. *Science* **261**, 1574–1577.
- Pavletich, N.P. & Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å resolution. *Science* **252**, 809–817.
- Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **73**, 804–809.
- Milligan, J.F., Grobe, D.R., Witherall, G.W. & Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798.
- Wyatt, J.R., Chastain, M. & Puglisi, J.D. (1991). Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques* **11**, 764–769.
- Davanloo, P., Rosenberg, A.H., Dunn, J.J. & Studier, F.W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **81**, 2035–2039.
- Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393–399.
- Delaglio, F., Grzesiek, S., Vuister, G., Zu, G., Pfeiffer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293.
- Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. (1991). A common sense approach to peak picking in two-, three- and four-dimensional spectra using automatic computer analysis of contour programs. *J. Magn. Reson.* **95**, 214–220.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. & Sykes, B.D. (1995).  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
- Brügel, A.T. (1992). *X-PLOR User Manual*. (Version 3.1), Yale University, New Haven.
- Lavery, R. & Sklenar, H. (1988). The definition of generalized helical parameters and axis of curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91.
- Lavery, R. & Sklenar, H. (1989). Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dynam.* **6**, 655–667.
- Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* **24**, 958–961.

Received: 18 Nov 1995; revisions requested: 22 Nov 1995  
revisions received: 27 Nov 1995. Accepted: 27 Nov 1995.

## Note added in proof

A related solution structure of the BIV Tat-TAR complex has recently been published (Puglisi, J.D., Chen, L., Blanchard, S. & Frankel, A.D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science* **270**, 1200–1203). The structure determined by Puglisi *et al.* was based on 26 intermolecular restraints, in contrast to our structure, which was based on 102 intermolecular restraints. Both structures globally position the β-hairpin fold of the Tat peptide in the major groove of TAR RNA. The description of the structure by Puglisi *et al.* differs significantly, however, in the position of the key residues R73, I79, U10, and G11 and of the phosphate contacted by the side chain of R73, all of which are associated with the specificity of BIV Tat-TAR complex formation.

Puglisi *et al.* did not identify the imino proton of bulged U10 in their spectra of the complex. We identified this bulged U10 imino proton and established that a U10•A13•U24 triple is formed, based on the observation of an NOE between the imino proton of U10 and the H8 proton of A13 in the complex. Thus, U10 is positioned in the plane of the A13•U24 base pair in our structure, whereas U10 is shown to be positioned between the G11•C25 and G9•C26 base pairs in the schematic drawing outlining intermolecular contacts in their structure. Since I79 is positioned adjacent to U10

in this schematic drawing, this critical amino acid, which cannot be mutated to any other hydrophobic residue, is also in different positions in the two structures. An examination of their published stereo figure, however, leads to an alternative interpretation of the positioning of U10 and I79 which would be closer to our own structure. The side chain of R73 (the specificity-determining arginine) is hydrogen bonded to the phosphate backbone at the C9-G10 step in our structure. By contrast, it is said to be hydrogen bonded at the U10-G11

step in the text of the Puglisi *et al.* paper, whereas their published stereo figure shows the R73 side chain approaching the phosphate at the C8-G9 step. Previous studies using ethylation interference experiments [9] identify the phosphate that is contacted by R73 as that in the C9-G10 step. Significant discrepancies between the text, the schematic drawing outlining intermolecular contacts, and the stereo figure in the paper by Puglisi *et al.* limit the scope of further detailed comparison between the two structures.