Structural Features of an RNA Containing the CUGGGA Loop of the Human Immunodeficiency Virus Type 1 Trans-Activation Response Element†

Richard A. Colvin,‡,§ Stephen W. White,*,‡ Mariano A. Garcia-Blanco,*,‡,§,∥ and David W. Hoffman‡

Department of Microbiology, Section of Cell Growth, Regulation and Oncogenesis, and Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710

Received November 5, 1992

ABSTRACT: A 19-nucleotide RNA containing the CUGGGA loop sequence corresponding to nucleotides 30–35 of the HIV-1 trans–activation response element (TAR) was synthesized in vitro and analyzed by biochemical methods and one- and two-dimensional NMR spectroscopy. Diagnostic RNase cleavage patterns were similar for the loops in the full-length HIV-1 TAR and the 19-nucleotide RNA, indicating that they are similar in structure. NMR data showed that the loop is stabilized by base-stacking interactions. The first loop nucleotide is stacked upon the A-helical stem, and the loop uridine is stacked upon this cytosine. On the opposite side of the loop, the third loop guanosine is stacked upon the adenosine, which is stacked upon the stem. No specific Watson–Crick or non-Watson–Crick base pairing across the loop was identified. Unusually short interribose distances indicate a significant distortion of the sugar–phosphate backbone centered at the adenosine. Relatively short NMR relaxation times for protons of the adenosine and its adjacent guanosine, as well as rapidly exchanging imino protons, provide evidence for dynamic processes occurring in the loop.

The trans-activation response element (TAR)[1] of the human immunodeficiency virus type 1 (HIV-1) consists of the first 60 nucleotides of all HIV-1 RNAs. TAR forms a highly structured element containing a stem, a 4-base bulge, and a 6-nucleotide CUGGGA loop (Muesing et al., 1987; Harper & Logsdon, 1991; Berkhout, 1992; Colvin & Garcia-Blanco, 1992). The integrity of this element is required for the transcriptional trans-activation mediated by the viral trans-activator protein Tat. Recent studies have shown that the role of TAR is to bring Tat into proximity with the transcription machinery (Berkhout et al., 1989; Selby & Peterlin, 1990; Southgate et al., 1990). Tat binds to the bulge of TAR in vitro, although with low specificity (Dingwall et al., 1989; Weeks et al., 1990; Roy et al., 1990; Harper & Logsdon, 1991; Colvin & Garcia-Blanco, 1992). Although the loop of TAR is not required for Tat binding in vitro, mutational analyses show that it is essential for transcriptional trans-activation in vivo and, therefore, for viral replication (Feng & Holland, 1988; Berkhout & Jeang, 1989). Consistent with this, the loop is likely to be a site for binding to cellular factors involved in stabilizing the interaction between Tat and TAR (Gatignol et al., 1989; Gaynor et al., 1989; Marciniak et al., 1990; Sheline et al., 1991; Wu et al., 1991). The specificity of the interaction between the loop and binding factors may depend on the structure and/or sequence of the loop.

The complete TAR RNA element is 60 nucleotides in length and too large for high-resolution NMR analysis by currently available methods. Recently, a detailed NMR study of the HIV-1 TAR element in the region of the bulge was reported (Puglisi et al., 1992). In our paper, we focus on the structure of the CUGGGA loop. A 19-nucleotide RNA molecule was made in vitro and was designed to optimize several features that could not easily be satisfied by a single ideal nucleotide sequence. First, and most importantly, we recognized that the CUGGGA loop sequence had to be preserved. Second, the isolated stem sequence should maintain the stable A-form helix predicted to be found just preceding the loop (Weeks & Crothers, 1991). Third, the RNA molecule was made using T7 RNA polymerase (see below), and transcription is most efficient when two or preferably three guanine residues are at the 5′ end. Fourth, with such a short molecule, it is absolutely essential to avoid sequences that can potentially form alternative stem–loop structures. Finally, to facilitate the NMR assignment process (complicated with even relatively small RNA molecules), it was preferable to avoid four or more consecutive purines or pyrimidines.

The final sequence 5′-GGGCGCUGGGACGCCCGUC-3′ represents a compromise molecule, referred to as TAR-19, that we considered most likely to produce relevant biological information. Although TAR-19 does contain some differences from the native sequence, we believe that these result in minimal changes in the stem–loop structure. Our evidence comes from experiments presented in this paper and from work performed in other laboratories which is discussed later.

MATERIALS AND METHODS

Nomenclature. Nucleotides which are numbered C30–A35 refer to the loop nucleotides of full-length HIV-1 TAR. Nucleotides numbered C6[30]–A11[35] refer to the loop nucleotides of TAR-19, the RNA used in the NMR experiments, with the numbers in brackets referring to the equivalent nucleotides in full-length HIV-1 TAR. Nucleotides numbered G1–G5 and C12–C19 refer to the remaining nucleotides of TAR-19.

† This work was supported in part by a National Science Foundation Grant (MCB9118369 to S.W.W.) and by start-up funds from the Duke University Medical Center (to M.A.G.-B).

* Authors to whom correspondence should be addressed.

‡ Department of Microbiology.

§ Section of Cell Growth, Regulation and Oncogenesis.

∥ Department of Medicine.

[1] Abbreviations: CMCT, 1-cyclohexyl-3′-[2-(*N*-methyl-4-morpholino)ethyl]carbodiimide metho *p*-toluene sulfonate; DMS, dimethyl sulfate; HIV-1, human immunodeficiency virus type 1; HMQC, heteronuclear multiple quantum coherence spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TAR, trans–activation response element; Tat, trans-activator protein; TOCSY, total correlated spectroscopy; TSP, 3-(trimethylsilyl)-1-propanesulfonate; 2QF-COSY, two quantum-filtered correlated spectroscopy.

0006-2960/93/0432-1105$04.00/0

RNA Synthesis and Purification. TAR-19 RNA was transcribed using T7 RNA polymerase, essentially as described by Milligan et al. (1987). Template DNA consisting of the oligonucleotides 5′-AATTCTAATACGACTCACTATAG-GGCGCTGGGACGCCCGTC-3′ and 5′-GACGGGCGTC-CCAGCGCCCTATAGTGAGTCGTATTAG-3′ was synthesized on an Applied Biosystems Model 392 DNA synthesizer and purified by gel electrophoresis under denaturing conditions. T7 RNA polymerase was purified from an overproducing strain of *Escherichia coli* (pAR1219) provided by John Dunn. Small-scale transcriptions were performed to optimize the yield of TAR-19 RNA by varying the concentrations of RNA polymerase, template DNA, nucleotide triphosphates (NTPs), and magnesium. A yield of approximately 1 mg of purified RNA/10 mL of transcription volume was obtained using 2.2 mM NTPs, 16 mM $MgCl_2$, and 200 nM template DNA.

Preparative-scale transcription reactions were incubated for 4–6 h at 37 °C and then ethanol precipitated. The resulting crude RNA pellet was dissolved in 8 M urea and subsequently passed through a Sephadex G25 gel filtration column in 8 M urea. The 19mer RNA product eluted at the front and was resolved from unreacted nucleotide triphosphates. The TAR-19 RNA was then separated from transcripts of incorrect size by electrophoresis on 20% polyacrylamide gels under denaturing conditions (8 M urea). Removal of unreacted nucleotide triphosphates prior to electrophoresis resulted in faster running gels and more uniform migration of the RNA bands. RNA was visualized by UV shadowing and then was removed from the gel using a Bio-Rad Model 422 electroeluter. The RNA was further purified by ethanol precipitating twice, and it was finally passed through a Sephadex G25 gel filtration column in 3 mM phosphate buffer and lyophilized.

As an alternative method of producing the TAR-19 RNA, the oligonucleotides were cloned into the *Eco*R1 and *Hin*cII sites of pUC 19 to create the plasmid pTAR.19. The plasmid (3 mg) was linearized by digestion with *Hin*cII (New England Biolabs) prior to use in a 30-mL transcription reaction. RNA transcribed from this plasmid DNA was purified the same as from synthetic oligonucleotide template. Full-length (60 nucleotide) HIV-1 TAR RNA was transcribed from the plasmid pT7TAR (Marciniak et al., 1990) using T7 RNA polymerase (Stratagene) under standard conditions (Melton et al., 1984).

Oligoribonucleotides with sequences 5′-GGGGGCUGG-GAGGGGCGUC-3′ (RNA A) and 5′-GCCCCCUGG-GACCCCCGUC-3′ (RNA B) were prepared from 5-mL transcription reactions as described above. DNA templates consisted of synthetic oligonucleotide 5′-TAATACGACT-CACTATAG-3′ annealed to 5′-GTCGCCCCTCCCAGC-CCCCTATAGTGAGTCGTATTA-3′ for RNA A and 5′-GTCGGGGGTCCCAGGGGGCTATAGTGAGTCGTA-TTA-3′ for RNA B.

Nuclease Probing. The synthetic RNAs were labeled at their 5′ ends to a specific activity of 500 μCi/mmol with T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]-ATP (New England Nuclear). Each RNA (10^{-12} mol) was incubated with the indicated amount of RNase T1, a nuclease with specificity for single-stranded guanosines, in a total volume of 20 μL [buffer conditions: 100 mM KCl, 10 mM $MgCl_2$, 20 mM HEPES·KOH (pH 7.9), 1 mM dithiothreitol (DTT), and 0.1% (v/v) Nonidet P-40] for 30 min on ice. Following incubation, the reactions were extracted once with phenol, once with phenol:chloroform (1:1), and once with chloroform and were precipitated with ethanol. The labeled RNAs were then separated in a 20% polyacrylamide (acrylamide–bisacrylamide, 29:1) sequencing gel.

Dimer versus Monomer Assay. An assay to distinguish between RNA hairpins and monomers was carried out essentially as described by Hoffman et al. (1993). Solutions of TAR-19 RNA, designed dimer RNA consisting of a mixture of RNA A and RNA B, and separate RNA A and RNA B were prepared in a buffer of 15 mM phosphate, pH 6, at concentrations of 6 mg/mL in volumes of 0.01 mL. The RNA samples were heated to 90 °C for 3 min and then slowly cooled to room temperature. Each RNA solution (5 μL) was then loaded onto a cold (4 °C) 20% polyacrylamide gel in 1× TBE buffer. Electrophoresis was carried out at 7 V/cm for 24 h at 4 °C. The RNA bands were initially visualized by UV shadowing. For a more permanent record, the gel was stained with Stains-all (Sigma Chemical Co.), destained with water for several hours, and photocopied.

NMR Experiments. NMR data were collected at 500 or 600 MHz on Varian Unity 500 or 600 spectrometers, with the exception of a ^{1}H–^{31}P correlated spectrum, which was acquired on a 500 MHz Bruker AMX spectrometer. Spectra in D_2O solvent were obtained using samples with 1–12 mg TAR-19 in 550 μL of 15 mM Na/K phosphate buffer at pH 7.0. Proton spectra in 90% H_2O/10% D_2O solvent were obtained for TAR-19 in the same buffer at pH 7 and pH 5, using either the 1–1 (jump and return) method (Hore, 1983), a selective excitation method using a 2–20-ms soft pulse for water suppression (Sklenar & Bax, 1987), or presaturation. 2D nuclear Overhauser effect spectra were acquired using the method of Macura and Ernst (1979), with several mixing times between 200 and 280 ms, except for one spectrum acquired with a mixing time of 60 ms. 2QF-COSY spectra were acquired by the method of Müller et al. (1986). TOCSY spectra with mixing times of 70 and 120 ms were acquired by the method of Bax and Davis (1985). All 2D NMR were acquired in the phase-sensitive mode by the method of TPPI (Marion & Wüthrich, 1983). 2D spectra were typically acquired with 450–600 blocks of 2048 complex points, using sweep widths of 4000–5500 Hz for spectra in D_2O and 11000 Hz for spectra in H_2O. Two-dimensional data were acquired at 2, 26, 30, and 35 °C, resulting in relatively small temperature-dependent shifts in proton resonance frequencies; these shifts were useful in identifying overlapping resonances. H1′–H2′ coupling constants were estimated from the separation of anti-phase components in a 2QF-COSY spectrum acquired with a resolution of 2 Hz per point in the *t2* dimension, zero-filled once to a resolution of 1 Hz per point, and processed without apodization. NMR data were processed on a Silicon Graphics 4D30TG workstation using FELIX software by Hare Research, Inc.

A ^{13}C–^{1}H correlated heteronuclear multiple quantum coherence spectrum of TAR-19 was acquired at 600 MHz proton frequency using the pulse sequence of Bax and Davis (1985). A total of 116 blocks of 1024 complex points were acquired with 512 scans per block, for a total acquisition time of 18 h. A sweep width of 5500 Hz was used for protons, and a sweep width of 9000 Hz centered at 140 ppm was used in the ^{13}C dimension, so all carbons except the C8, C6, and C2 were folded over. No carbon decoupling was used during the acquisition time. A ^{31}P–^{1}H correlated spectrum was acquired at 500 MHz proton frequency using the method of Williamson and Bax (1988). A total of 160 blocks of 256 scans each were acquired, with sweep widths of 1202 and 1000 Hz in the proton and phosphorus dimensions, respectively.

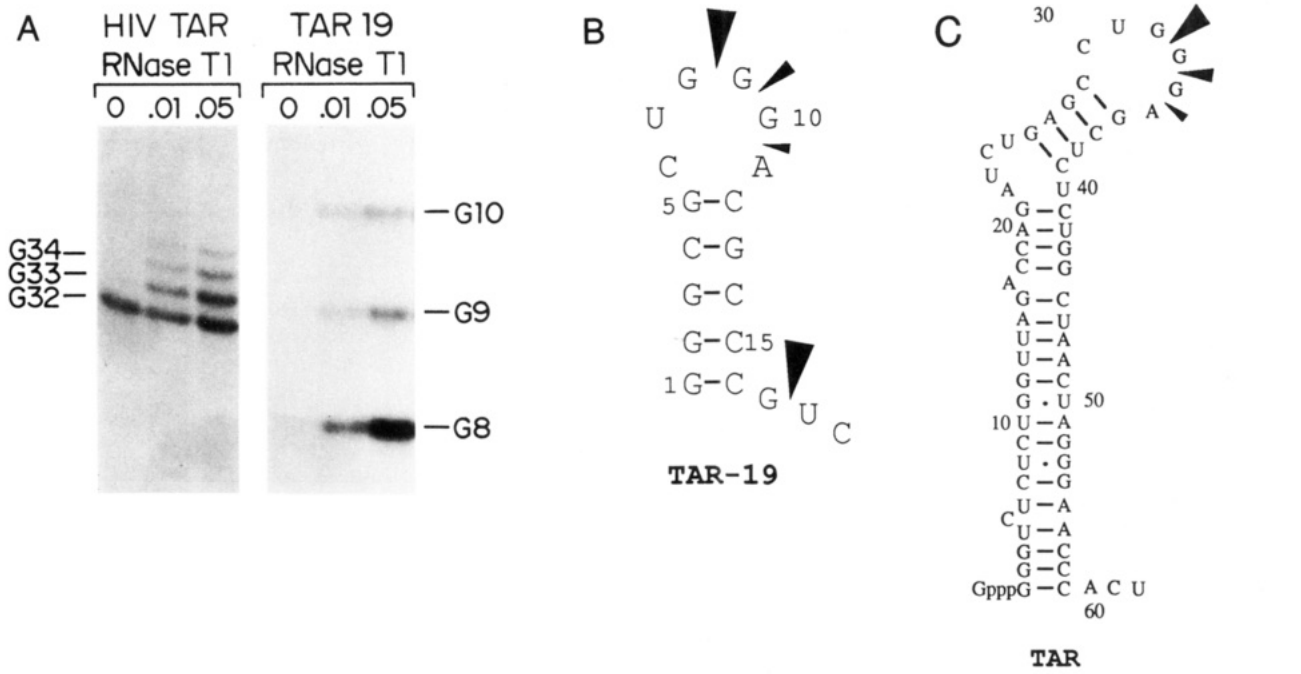

FIGURE 1: (A) Partial RNase T1 digestions of HIV-1 TAR and TAR-19. Residues where cleavage occurred are indicated. The upper three bands in HIV-1 TAR correspond to the three bands in TAR-19. The lower band is a nonspecific cleavage appearing in all lanes and should not be confused with G32. (B) Summary of cleavages by RNase T1 on TAR-19. Arrows indicate sites of cleavage by the enzyme. The relative frequency of cleavage is indicated by the relative size of the arrow. (C) Summary of cleavages by RNase T1 on full-length HIV-1 TAR. The relative frequency of cleavage is indicated by the relative size of the arrow. Position G8[32] of TAR-19 is sensitive to RNase T1, while G9[33] and G10[34] are progressively less sensitive to the nuclease. This pattern is similar to the pattern observed in full-length TAR at the equivalent loop positions G32, G33, and G34. Furthermore, the T1 cleavage pattern indicates that other G residues are not accessible to cleavage by RNase T1 except for G17, which is part of the 3-nucleotide tail (data not shown).

T_1 relaxation measurements were performed using the inversion–recovery method at 30 °C on a sample of 1 mM TAR-19 in D_2O solvent and 15 mM phosphate buffer, pH 7.0. A 15-s recycle time was used between scans, which exceeded five times the T_1 of the slowest relaxing proton in TAR-19. Sixteen scans were acquired for each of the 14 inversion–recovery delays between 0 and 4.0 s. The residual HDO resonance was not saturated for the T_1 experiments.

Base-Catalyzed Exchange of H8 Protons. Guanosine and adenosine H8 protons of TAR-19 were deuterated by heating approximately 1 mg of RNA to 70 °C for 18 h in 550 μL of D_2O, pH 8, and 50 mM phosphate buffer. Deuteration of H8 protons was greater than 75% as judged by proton resonance intensities, with relatively little degradation of the TAR-19 RNA.

RESULTS

Verifying Similarity of TAR-19 and HIV-1 TAR RNA Loops. (*i*) *RNAse Cleavage Patterns.* Enzymatic cleavage assays are very sensitive to structure, and a comparison of distinctive nuclease cleavage patterns provided an indication that the loop structures of full-length TAR and TAR-19 are at least highly similar. RNase T1 preferentially cleaves RNA after single-stranded guanosines. The CUGGGA loop of full-length HIV-1 TAR exhibits a characteristic RNase T1 cleavage pattern (Figure 1). Under enzyme limited digestion, G32 is most frequently cleaved, while G33 and G34 are cleaved with progressively less frequency, presumably due to their partial protection by the tertiary structure of the RNA loop. TAR-19 exhibits this same characteristic RNase T1 cleavage pattern and is, therefore, likely to have a loop structure which is similar or identical to that of the full-length HIV-1 TAR. Single-point mutations in the loop of the full-length TAR alter the RNase T1 cleavage pattern significantly (Colvin & Garcia-Blanco, 1992), providing evidence that RNase T1 is indeed a sensitive probe of the loop structure. Other investigators have found very different RNase T1 cleavage patterns in similar stretches of guanosine residues, indicating a structure dependence on this cleavage (Egebjerg et al., 1987). The same pattern is observed when the RNA structure is probed with the chemicals kethoxal and CMCT (Colvin & Garcia-Blanco, 1992), providing further evidence that this cleavage pattern is a result of structural features rather than sequence.

(*ii*) *Concentration Dependence of TAR-19 Structure.* Although RNase T1 cleavage suggests that the loops TAR-19 and HIV-1 TAR are similar in structure, we recognize that the NMR experiments and RNase cleavage experiments were performed at RNA concentrations that differ by several orders of magnitude. At high concentrations, RNAs may be more likely to form multimers rather than hairpins. Therefore, we decided to investigate the concentration dependence of the TAR-19 structure. Samples of TAR-19 in 15 mM phosphate buffer at pH 6 were prepared at concentrations ranging from 0.01 mg/mL to 10 mg/mL. The samples were heated briefly, cooled, subsequently run on a 10% native acrylamide gel, and visualized by UV shadowing and ethidium bromide staining (data not shown). Concentrations from 0.01 mg/mL to 6 mg/mL formed a single sharp band on the gel. At 10 mg/mL, most of the TAR-19 migrated as a single band at the same position as the lower concentration samples; however, several additional bands corresponding to larger RNA species were also apparent. The larger species are presumably aggregates of TAR-19, suggesting that it is possible to force aggregation as the solubility limit of the RNA is approached. These results argue that under the conditions of the NMR experiments, the majority of TAR-19 is a hairpin.

(*iii*) *Dimer versus Monomer Experiment.* All RNA hairpins have the potential to form palindromic dimers, where

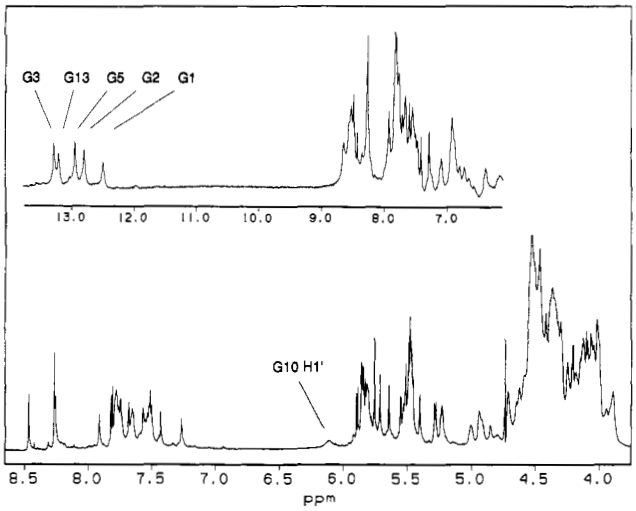

FIGURE 2: One-dimensional 600 MHz proton NMR spectrum of TAR-19 in D_2O solvent at pH 7.0, 26 °C. The inset shows a section of the spectrum in H_2O solvent at 2 °C obtained using the method of Sklenar and Bax (1987). Similar spectra were obtained with the 1-1 method of water suppression, while spectra obtained with presaturation showed a decrease in the intensities of the exchangable proton resonances due to saturation transfer from the H_2O. The imino resonances of the five stem guanosines are labeled. The spectra of all RNA preparations contained some minor resonances in addition to those of the major TAR-19 RNA species, regardless of whether the RNA was synthesized from plasmid or synthetic oligonucleotide template. These minor RNA species are likely to be alternative conformations of TAR-19, possibly dimers.

the nucleotides corresponding to the loop in a hairpin structure form a bulge in the dimer structure. For example, an RNA which forms a tetraloop hairpin in solution was recently found to form a dimer in the crystal structure (Holbrook et al., 1991). A bulged dimer structure can be difficult to distinguish from a hairpin loop on the basis of NMR data alone, since cross-loop NOEs in a hairpin are difficult to distinguish from cross-bulge NOEs in a dimer (Hoffman et al., 1993). An additional experiment has been developed to provide confirming evidence for the hairpin or dimer nature of RNA stem–loop sequences and has been successfully used to show that the TAR RNA of equine infectious anemia virus forms a monomer hairpin (Hoffman et al., 1993). This additional experiment was performed in our work to assess the potential for TAR-19 to dimerize.

Two 19-nucleotide RNAs (designated RNA A and RNA B) were designed with sequences chosen so that they have little potential to form hairpins but will potentially form a stable heterodimer when mixed: 5′-GGGGGCUGGGAG-GGGCGUC-3′ (RNA A) and 3′-CUGCCCCCAGGGU-CCCCCG-5′ (RNA B). If TAR-19 does indeed form a dimer, it is likely that a heterodimer would form by mixing RNA A and RNA B. The structure and behavior of this heterodimer would likely be similar to a TAR-19 homodimer. Migration rates on native gels were compared for TAR-19 and a mixture of RNA A plus RNA B. At RNA concentrations comparable to those used in the NMR experiments (6 mg/mL), the TAR-19 migrated as a single sharp band. In comparison, the mixture of RNA A and RNA B migrated as multiple bands, all of which migrated slower than TAR-19. These results provide further evidence that TAR-19 has a hairpin structure under the conditions of the NMR experiments.

NMR Analysis. (i) Spectral Features and Resonance Assignments. The NMR spectra of the TAR-19 stem–loop (Figure 2) provide evidence for a well-defined RNA structure. The proton resonances are generally sharp and disperse. With

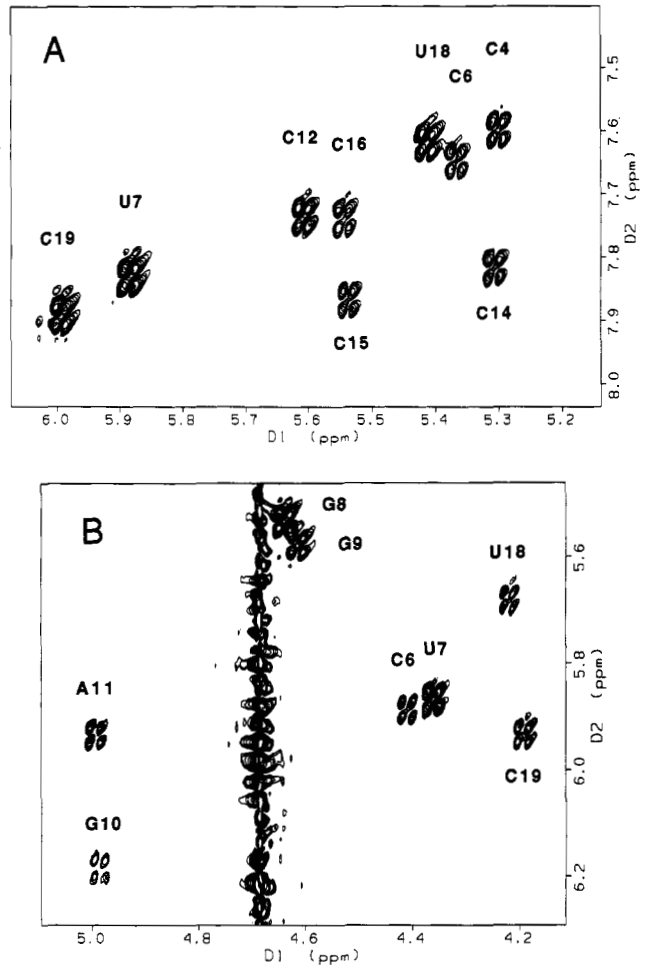

FIGURE 3: Sections of the two quantum-filtered COSY spectrum of TAR-19 obtained at 500 MHz, 35 °C, in 15 mM phosphate buffer, pH 7.0. (A) H5–H6 correlation peaks each of the nine pyrimidines in TAR-19 are clearly resolved, provided starting points for making sequence-specific assignments. (B) Correlation peaks are observed for the eight strongly coupled pairs of H1′–H2′ protons in TAR-19.

few exceptions, the resonances exhibit little change in chemical shift and line width over a range of temperature (2–55 °C), pH (5–7), $MgCl_2$ concentration (0–10 mM), and RNA concentration (0.2–3 mM). Resonance assignment was accomplished by first identifying protons belonging to the same pyrimidine ring (Figure 3) or ribose using through-bond couplings in 2QF-COSY and TOCSY spectra. Next, specific assignments were made by using NOE spectra to sequentially "walk" from ribose to base to ribose, as illustrated in Figure 4. Sequential NOE connectivities were made for all bases except G8[32], which has NOEs to only its own ribose, and G9[33], which has NOEs to its own ribose and ribose G8[32]. The assignments of G8[32] and G9[33] are therefore less certain than the other assignments for TAR-19, to the extent that there remains the unlikely possibility that the G9[32] and G9[33] assignments are interchanged. The H2 resonance of A11[35] was assigned by its NOE to the H1′ of C12. Resonance assignments are summarized in Table I.

The five sharp imino proton resonances observed near 13 ppm were assigned to the five base-paired G imino protons in the stem (Figure 2). The amino protons of the stem cytosines were assigned by NOEs their own H5 protons and to cross-strand G imino protons. The loop nucleotides contain a total of four imino protons (one for each U and G) which are potentially observable by NMR. None of these resonances, however, were detected at temperatures as low as 0 °C, pH 5 and pH 7, and upon addition of up to 10 mM $MgCl_2$. The

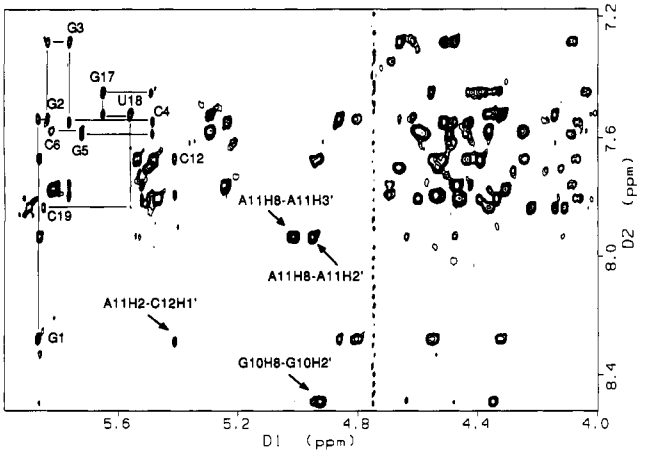

FIGURE 4: Section of the 2D NOE spectrum of TAR-19 in D_2O solvent. Examples of intraresidue H6/H8–H1′ NOEs are identified by their residue type and number. Other selected NOEs involving the CUGGGA loop are labeled with the identities (residue type, number, and proton type) of each of the two interacting nuclei. A sequential walk between H6/H8 and H1′ protons on one side of the stem and in the 3-nucleotide tail is demonstrated by connecting lines.

loop imino protons were therefore presumed to have intrinsically fast exchange rates, exchanging with the solvent protons on a time scale of milliseconds or less. Although the rapid exchange of loop imino protons indicates their relatively high accessibility to the solvent, the possibility of their transient involvement in hydrogen bonds cannot be ruled out.

Several additional experiments were particularly useful in confirming resonance assignments. The H5 resonances of cytidine and uridine were distinguished by the chemical shifts of their associated carbons in an HMQC spectrum as previously suggested by Varani and Tinoco (1991). The resonances of the C5 carbons of U7[31] and U18 are more than 5 ppm higher than any of the cytidine C5 resonances (Figure 5). In a similar manner, the assignment of H2 of A11[35] was confirmed by the chemical shift of its associated carbon, 9 ppm downfield from any of the C6 and C8 carbons. Adenosine and guanosine H8 assignments were verified by their exchange with deuterium upon heating a TAR-19 in D_2O at pH 8 for 18 h. Although in some cases H3′–P–H5′ linkages were verified by the ^{31}P–^{1}H correlated spectrum, this spectrum was of limited usefulness due to the considerable amount of overlap, especially in the H5′ and ^{31}P resonances.

TAR-19 RNA prepared in transcription reactions using synthetic oligonucleotide template appeared to be identical to that prepared from plasmid template as judged by one-dimensional NMR spectra. Plasmid DNA template was found to yield approximately 100 mol of TAR-19/mol of template, as compared to approximately 10 mol of TAR-19/mol of synthetic oligonucleotide template. Despite this difference in efficiency of the transcription reaction, the synthetic oligonucleotide template was judged to be preferable due to its ease of preparation. Plasmid template which was recovered after transcription, gel purified, and subsequently used in additional transcription reactions did not yield significant amounts of TAR-19 product.

(*ii*) *A-Helical Stem* (*Nucleotides G1–G5 and C12–C16*). Stem nucleotides yielded typical resonance data expected for an A-form double-stranded RNA helix, as indicated by the pattern of sequential and intranucleotide NOE intensities (Figure 4). The sequential H6/H8(*i*)–H2′/H3′(*i*–1) NOEs are stronger than the corresponding intranucleotide NOEs. H8/H6(*i*)–H5(*i*+1) NOEs are also observed, as expected, where this distance is 3.8 Å in an A-helix. In some cases, weak NOEs are detected between sequential base-stacked H8 and H6 protons. Cross-strand NOEs from the stem imino to H1′ resonances confirmed the A-helical structure of the stem (Heus & Pardi, 1991a).

(*iii*) *Loop Nucleotides*. The base-stacked structure of the stem extends into loop positions C6[30] and U7[31] (Figure 6). The intensities of the ribose to base NOEs are typical of A-helix for nucleotide C6[30] but more closely resemble B-helix for U7[31]. The H6 proton of C6[30] is closer to the H2′ and H3′ protons of ribose 5 than those of its own ribose, while the H6 proton of U7[31] is closer to its own H2′ and H3′ than those of C6[30]. No NOEs were observed from U7[31] to G8[32]. G8[32] and G9[33] are at the "top" of the hairpin, but the orientation of these purines relative to the other loop bases is uncertain, since their positions can only be defined by NOEs from one H8 proton per ring. Several H8–ribose–H8 NOE pathways, as well as a direct H8 to H8 NOE, clearly show that G10[34] is base-stacked upon A11[35]. NOEs from the H8 proton of A11[35] to the H5 and H6 protons of C12 show that this purine is stacked upon the base of C12 (Figure 6). The H2 proton of A11[35] has only a single NOE (to the H1′ of C12), confirming that this base is stacked upon C12. H8/H6–H1′ NOEs of medium intensity indicate that C6[30], U7[31], G10[34], and A11[35] are in the C2′-endo anti conformation.

Unusual interribose NOEs between G10[34], A11[35], and C12 indicate a substantial local distortion of the sugar–phosphate backbone (Figure 6). NOEs from the H3′ and H4′ of G10[34] to the H2′ and H3′ of A11[35] indicate short distances between these ribose protons (Table II). Additionally, a strong NOE between the H1′ of A11[35] and the H1′ of C12 shows that these protons are also unusually close. Consistent with this backbone distortion, the phosphorus which links A11[35] with C12 is shifted downfield to –3.52 ppm; all other phosphorus nuclei resonate in a relatively narrow range between –3.81 and –4.62 ppm. The compression of riboses 10, 11, and 12 suggest that A11[35] is bulged from the normal helical base-stacked orientation. Further evidence for a bulged adenosine is provided by the absence of any cross-strand NOEs from the adenosine H2 to the H1′ of U7[31], as would be observed if these bases were all normally base-stacked. This is consistent with previous studies of single nucleotide bulges in double-stranded DNA which show that a bulged adenosine continues to be stabilized by base-stacking, although with significant distortion of the sugar–phosphate backbone (Kalnick et al., 1989).

Adenosine H2 protons that are stacked inside RNA helices often have upfield chemical shifts (Varani et al., 1989) and relatively long T_1 relaxation times (Hoffman et al., 1993). For example, although the TAR RNA stem–loop of equine infectious anemia virus contains four adenosines, only the two adenosines in the central region of the A-helical stem exhibit the characteristic long T_1 relaxation times and upfield chemical shifts (Hoffman et al., 1993). In TAR-19, the H2 of A11[35] exhibits neither of these features. Mutational analyses indicate that A11[35] can be mutated to cytosine with no effect on trans-activation in vivo and can even be deleted in one of the stem–loops of HIV-2 (Berkhout, 1992). This is different from the other loop nucleotides that cannot be altered without greatly reducing function.

Sugars in oligoribonucleotides are usually exchanging rapidly between structures in the C2′ endo (S-type) and C3′ endo (N-type) families of conformers. The percentage of time that a sugar is in each conformer can be estimated from the H1′–H2′ coupling constant (Puglisi et al., 1990). H1′–

Table I: Chemical Shift Assignments for TAR-19[a]

	H8/H6	H5/H2	C5/C2	H1′	H2′	H3′	H4′	J(1′–2′)	imino/amino
G1	8.28	na[b]	na	5.88	4.87	4.81	4.56	<3	12.50
G2	7.54	na	na	5.84	4.67	4.62	4.54	<3	12.81
G3	7.28	na	na	5.77	4.52	4.44	4.49	<3	13.30
C4	7.55	5.25	97.4	5.49	4.50	4.60	4.44	<3	8.54, 6.74
G5	7.59	na	na	5.73	4.26	4.61	4.39	<3	12.96
C6[30]	7.57	5.30	97.7	5.84	4.37	4.60	4.38	8	
U7[31]	7.77	5.83	105.3	5.82	4.31	4.53	4.26	9	
G8[32]	7.81	na	na	5.48	4.60	4.66	4.12	9	
G9[33]	7.70	na	na	5.51	4.54	4.68	3.92	9	
G10[34]	8.48	na	na	6.14	4.94	4.36	4.65	8	
A11[35]	7.93	8.29	154.5	5.87	4.96	5.02	4.49	8	
C12	7.67	5.55	98.2	5.42	4.55	4.58	4.52	<3	8.63, 6.65
G13	7.79	na	na	5.77	4.53	4.70	4.48	<3	13.21
C14	7.76	5.26	97.1	5.53	4.47	4.35	4.43	<3	8.66, 6.93
C15	7.81	5.48	97.6	5.51	4.40	4.45		<3	8.58, 6.94
C16	7.68	5.51	98.0	5.50	4.56	4.38	4.43	<3	8.27, 6.81
G17	7.45	na	na	5.66	4.33	4.37	4.35	3.5	
U18	7.52	5.30	104.0	5.57	4.15	4.36	4.31	7	
C19	7.83	5.92	98.7	5.86	4.12	4.22	4.16	7	

[a] Chemical shifts (ppm) and intraribose coupling constants (Hz) are for TAR-19 in 15 mM phosphate buffer, pH 7.0. Data for nonexchangeable protons are reported at 26 °C, with chemical shifts relative to the residual HDO resonance at 4.74 ppm. Carbon chemical shifts are relative to external TSP. Intraribose coupling constants J(H1′–H2′) that were undetectable are reported as <3 Hz; other couplings have an uncertainty of 2 Hz. Amino and imino protons are reported at 2 °C, where their exchange rate with the solvent is reduced. [b] na = not applicable.

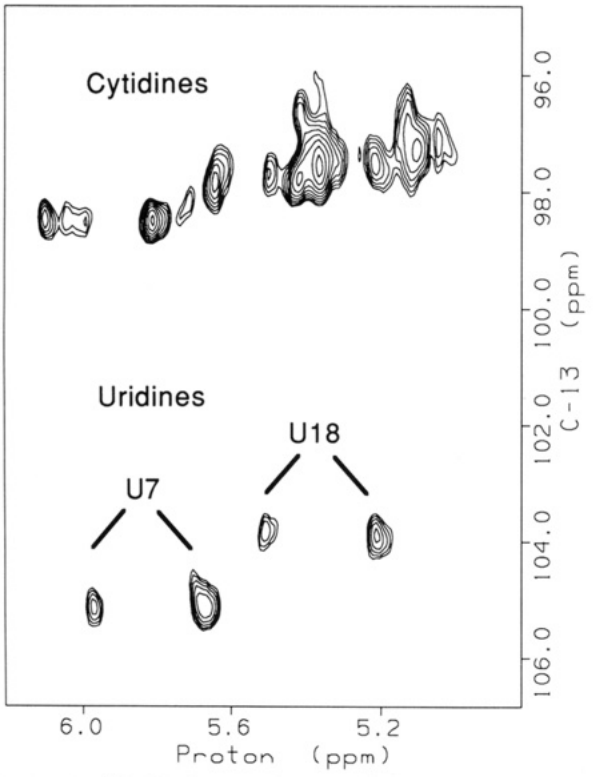

FIGURE 5: ^{13}C–^{1}H heteronuclear multiple quantum coherence spectrum of TAR-19 in D_2O solvent provided confirmation of resonance assignments. A section of the spectrum containing the H5–C5 correlations of the two uridines and seven cytidines is shown. The resonances of the two uridines are distinguished by their higher carbon chemical shifts. In a similar manner, the H2–C2 resonance of the single adenosine was distinguished from H8 and H6 resonances by its higher carbon chemical shift. Although the TAR-19 sample contained only natural abundance ^{13}C, a usable two-dimensional spectrum was easily obtainable in an 18-h acquisition time.

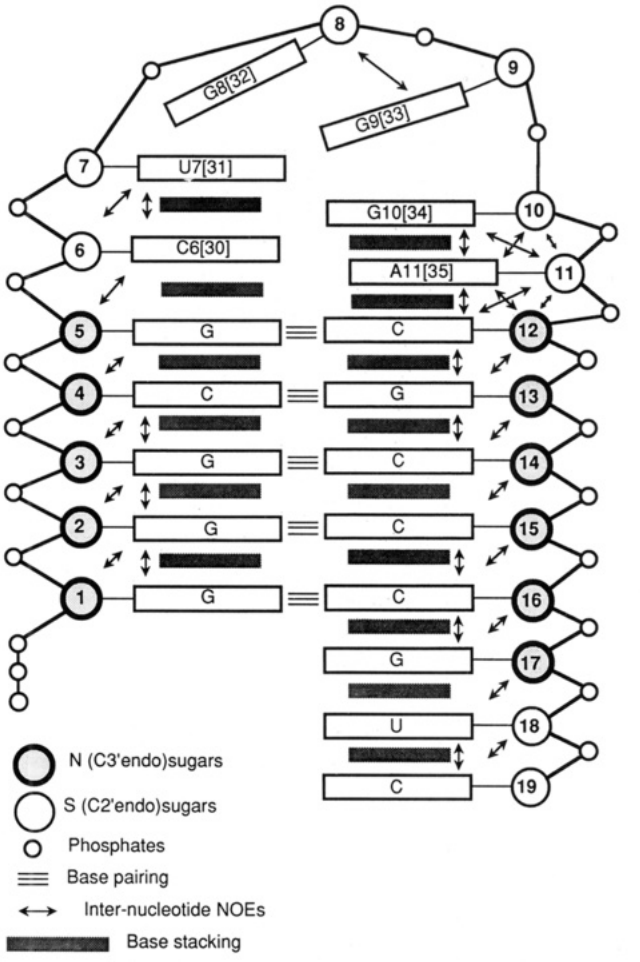

FIGURE 6: Schematic representation of the TAR-19 RNA is shown to illustrate important structural features and to summarize the NMR results. Nucleotides are numbered according to their position in the TAR-19 sequence, loop bases are numbered consistent with the TAR-19 sequence, and the loop bases position in the full HIV-1 TAR sequence are in brackets. Observed internucleotide NOEs are indicated by arrows. Additional NOEs may be present but are unobservable due to chemical shift overlap. No cross-loop NOEs between C6-A11 or U7-G10 were observed.

H2′ couplings were either very weak or undetected for 11 of the riboses, located in A-helical parts of TAR-19. Relatively strong 7–9-Hz couplings were observed for each of the six loop riboses and the last two tail nucleotides (Figure 3). Stem riboses were assigned to the N (C3′ endo) family of conformers, based on their small H1′–H2′ coupling constants (Table I). All six loop sugars are primarily in the C2′ endo (S-type) conformer family. S-type sugar puckers permit the backbone to extend, facilitating the reversal of direction in the sugar–

Table II: Summary of NOEs Involving Protons in CUGGGA Loop[a]

proton pair	proton pair
C6 H6-G5 H1′ (w)	G10 H8-G10 H4′ (w)
C6 H6-G5 H2′ (m)	G10 H8-A11 H2′ (m)
C6 H6-G5 H3′ (o)	G10 H8-A11 H3′ (w)
C6 H5-G5 H3′ (w)	G10 H8-A11 H4′ (w)
C6 H6-C6 H1′ (m)	G10 H8-A11 H8 (w)
C6 H6-C6 H2′ (o)	G10 H3′-A11 H2′ (m)
C6 H6-C6 H3′ (o)	G10 H3′-A11 H3′ (w)
C6 H6-U7 H6 (w)	G10 H4′-A11 H2′ (m)
U7 H6-C6 H2′ (m)	A11 H2-C12 H1′ (m)
U7 H6-U7 H4′ (w)	A11 H8-G10 H2′ (o)
U7 H6-U7 H2′ (s)	A11 H8-G10 H3′ (w)
G8 H8-G8 H2′ (s)	A11 H8-G10 H4′ (w)
G8 H8-G8 H3′ (m)	A11 H8-A11 H1′ (s)
G9 H8-G8 H3′ (m)	A11 H8-A11 H2′ (s)
G9 H8-G8 H2′ (o)	A11 H8-A11 H3′ (s)
G9 H8-G9 H1′ (m)	A11 H8-A11 H4′ (w)
G9 H8-G9 H2′ (s)	A11 H8-C12 H5 (w)
G9 H8-G9 H3′ (s)	A11 H1′-C12 H1′ (m)
G9 H8-G9 H4′ (m)	C12 H5-A11 H2′ (w)
G10 H8-G10 H1′ (w)	C12 H6-A11 H1′ (m)
G10 H8-G10 H2′ (m)	C12 H6-A11 H2′ (m)
G10 H8-G10 H3′ (m)	

[a] Peaks which are relatively strong in a two-dimensional NOE spectrum with a 60-ms mixing time are designated as (s). Peaks which have integrated volumes less than one-tenth of the intranucleotide H6–H5 NOEs in a 280-ms spectrum are designated as (w). Other NOE peaks observed in spectra with mixing times of less than 280 ms are designated as (m). Peaks designated with (o) are present but their intensity is uncertain due to resonance overlap. Additional NOEs are present in the data but are obscured or ambiguous due to overlap. Intraribose and intrabase NOEs are not included.

phosphate backbone as has been observed in other RNA hairpins (Cheong, et al., 1990; Saenger, 1984; Heus & Pardi, 1991b).

Several observations provided evidence of a dynamic process involving loop residues G10[34] and A11[35]. The H1′ resonance of G10[34] is unusually broad, as can be clearly seen in Figure 2. The width of this resonance is temperature sensitive, increasing from 14 Hz at 42 °C to 36 Hz at 26 °C (Figure 2), while other resolved resonances in TAR-19 showed little or no change in width. This behavior is most consistent with an exchange between conformations which occurs at the intermediate rate on the NMR time scale, which is on the order of milliseconds. In addition, the H8 protons of G10-[34] and A11[35] exhibit T_1 relaxation times which are slightly shorter than those of the H8 and H6 protons in the stem, with the exception of the H8 resonance of G1 which relaxes quickly. More significantly, the H2 of A11[35] also relaxes faster than most H8 and H6 resonances in TAR-19, which is unusual since adenosine H2 resonances are usually identified by their long T_1 relaxation times (Hoffman et al., 1993). The short T_1 relaxation times of the G1, G10[34], and A11[35] protons are best attributed to internal motion involving these nucleotides. While fast internal motion with a correlation time less than the rotational time of the molecule would be expected to increase T_1, relatively slow motion could provide additional relaxation pathways.

(*iv*) *Tail Nucleotides (G16–C19).* The three nucleotide "tail" was included in the TAR-19 sequence to permit subcloning of the template DNA into a plasmid. It is interesting that near normal base-stacking continues from the stem into the three tail nucleotides, despite their lack of base-pairing partners. Intensities of the intra- and internucleotide ribose to base NOEs are similar to those seen in the stem, with the exception of the U18-C19 sequential NOEs which are of somewhat weaker intensity (Figure 4). An RNA hairpin loop recently analyzed by Heus and Pardi (1991b) contained a three-nucleotide tail at the 3′ end of the molecule similar to that of TAR-19. The ribose of G17 is almost always in the C3′ endo family of conformers, while riboses of U18 and C19 are in rapid exchange between the C2′ endo and C3′ endo conformer families, indicated by the intermediate value of their H1′–H2′ couplings. Chemical shifts of U18 and C19 protons exhibit a greater temperature dependence than other nonexchangeable protons of TAR-19. Variations of up to 0.1 ppm over a 30 °C range are consistent with slight changes in the orientation of the tail nucleotides with temperature.

DISCUSSION

From a practical standpoint, the choice of RNA sequence was justified by our success in making large quantities of TAR-19 from both oligonucleotide and plasmid templates, and our ability to unambiguously assign the NMR spectrum. With regard to the biological relevance of the molecule, we have provided strong evidence that it adopts a stable hairpin conformation in solution and that the loop has a structure that is similar by nuclease probing to that of full-length TAR. Certainly, the difference from the native molecule that is most likely to cause structural and functional perturbations is the G-C base pair at the top of the TAR-19 stem that corresponds to a C-G base pair in the wild-type HIV-1 TAR. We fully appreciate that this switch may introduce changes into the loop conformation that are too subtle to detect by nuclease probing. Work by Feng and Holland (1988), however, clearly shows that full-length HIV-1 TAR with this change exhibits wild-type levels of trans-activation in vivo. We would also point out that the TAR RNA structure with the C-G closing base pair is currently being investigated by NMR in several other laboratories. This will offer a unique functional insight into the TAR element since any structural differences that can be attributed to the closing base pair would indicate features that are not important for TAR function.

This study allows a comparison of the structure determined by NMR spectroscopy with the structure determined by nuclease and chemical probing methods. Of particular importance, C30 of TAR is not methylated by DMS while A35 is methylated (Colvin & Garcia-Blanco, 1992). This unequal susceptibility to methylation is consistent with the NMR data showing that A11[35] is bulged from the normal helical orientation and somewhat mobile, while C6[30] is normally base-stacked upon G5. Additional data show that while G32 is frequently cleaved by RNAse T1 and modified by kethoxal and CMCT, G33 and G34 are progressively less sensitive to these reagents (Figure 1; Harper and Logsdon, 1991; Colvin & Garcia-Blanco, 1992). These results are consistent with the structure of the loop of TAR-19 derived from the NMR studies where G10[34] is located on the interior of the loop and stabilized by base-stacking (Figure 6).

How does our structure for the CUGGGA loop correlate with biochemical data on how this region of the HIV-1 TAR RNA element interacts with protein? In general, apart from the special case of the glutaminyl-tRNA synthetase/tRNA-Gln complex (Rould et al., 1989), very little is known about the structural aspects of protein–RNA interactions, although information is starting to emerge (Wu & Uhlenbeck, 1987; Hoffman et al., 1990; Jessen et al., 1991; Ramakrishnan & White, 1992). In the case of the HIV-1 TAR loop, it appears that cellular factors may recognize not only the structure of the loop but also nucleotides G8[32] and G9[33]. Mutational studies show that these guanosines are not required for loop stability but are absolutely required for wild-type levels of trans-activation (Feng & Holland, 1988; Berkout & Jeang,

1989). Conversely, other mutations in the loop that destroy the ability to trans-activate transcription do apparently disrupt the wild-type structure (Colvin & Garcia-Blanco, 1992). Our model which shows G8[32] and G9[33] accessible at the top of the loop and not involved in helical interactions appears to agree with these data.

The TAR element from the HIV-1 related lentivirus EIAV (equine infectious anemia virus) shares some structural features with the TAR element of HIV-1 that may be important for function. EIAV TAR functions in a similar way to that of HIV-1 (Carvalho & Derse, 1991), and we have recently analyzed its structure by NMR and biochemical methods (Hoffman et al., 1993). Although smaller than HIV-1 TAR, it does appear to have functionally homologous regions. There is an exposed guanosine at the top of the hairpin loop that has been shown to be essential for EIAV TAR function (Carvalho & Derse, 1991). This guanosine may be the functional counterpart of G8[32] and/or G9[33] in HIV-1 TAR. Also, there is a CAGAUU sequence that matches the uridine bulge region of HIV-1 TAR which is known to interact with the HIV-1 Tat protein (Carvalho & Derse, 1991).

It was recently shown that overexpression of TAR renders cells in culture resistant to HIV-1 infection (Sullenger et al., 1990). This confirmed that transactivation is a viable target for chemotherapeutic treatment of AIDS. Although it is clear that both an intact bulge and loop are required for inhibition (Sullenger et al., 1991), this study provides invaluable structural information toward the design of chemical analogs of TAR that prevent trans-activation and thus viral reproduction.

ACKNOWLEDGMENT

We thank Xiaolian Gao for assistance in collecting the ^{31}P–^{1}H correlated spectrum and Stephanie Porter for purifying the T7 RNA polymerase. We are indebted to Drs. J. Jaeger and I. Tinoco, whose communication of results prior to publication and helpful suggestions allowed us to identify and resolve assignment ambiguities in the loop.

REFERENCES

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Berkhout, B. (1992) *Nucleic Acids Res. 20*, 27–31.
Berkhout, B., & Jeang, K.-T. (1989) *J. Virol. 63*, 5501–5504.
Berkhout, B., Gatignol, A., Rabson, A. B., & Jeang, K.-T. (1989) *Cell 62*, 757–767.
Carvalho, M., & Derse, D. (1991) *J. Virol. 65*, 3468–3474.
Cheong, C., Varani, G., & Tinoco, I. (1990) *Nature 346*, 680–682.
Colvin, R. A., & Garcia-Blanco, M. A. (1992) *J. Virol. 66*, 930–935.
Dingwall, C., Ernberg, I., Gait, M. J., Green, S., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, A., & Valerio, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6925–6929.
Egebjerg, J., Leffers, H., Christensen, A., Andersen, H., & Garret, R. A. (1987) *J. Mol. Biol. 196*, 125–136.
Feng, S., & Holland, E. (1988) *Nature 334*, 165–167.
Gatignol, A., Kumar, A., Rabson, A., & Jeang, K.-T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7828–7832.
Gaynor, R. B., Soultanakis, E., Kuwabara, M., Garcia, J. A., & Sigman, D. S. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4858–4862.
Harper, J. W., & Logsdon, N. J. (1991) *Biochemistry 30*, 8060–8066.
Heus, H. A., & Pardi, A. (1991a) *J. Am. Chem. Soc. 113*, 4360–4361.
Heus, H. A., & Pardi, A. (1991b) *Science 253*, 191–193.
Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495–2499.
Hoffman, D. W., Colvin, R. A., Garcia-Blanco, M. A., & White, S. W. (1993) *Biochemistry* (preceding paper in this issue).
Holbrook, S. R., Cheong, C., Tinoco, I., & Kim, S.-H. (1991) *Nature 353*, 579–581.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Kalnik, M. W., Norman, D. G., Swann, P. F., & Patel, D. J. (1989) *J. Biol. Chem. 264*, 3702–3712.
Macura, S., & Ernst, R. R. (1979) *Mol. Phys. 41*, 95–117.
Marciniak, R. A., Garcia-Blanco, M. A., & Sharp, P. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3624–3628.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035–7056.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Muesing, M. A., Smith, D. H., & Capon, D. J. (1987) *Cell 48*, 691–701.
Müller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482–6492.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I. (1990) *Biochemistry 29*, 4215–4226.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76–80.
Rould, M. A., Perona, J. J., Söll, D., & Steitz, T. A. (1989) *Science 246*, 1135–1142.
Roy, S., Parkin, N. T., Rosen, C., & Sonenberg, N. (1990) *Genes Dev. 4*, 1365–1373.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Selby, M. J., & Peterlin, B. M. (1990) *Cell 62*, 769–776.
Sheline, C. P., Milocco, L. H., & Jones, K. A. (1991) *Genes Dev. 5*, 2508–2520.
Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 75*, 378–383.
Southgate, C. M., Zapp, M. L., & Green, M. R. (1990) *Nature 345*, 640–642.
Sullenger, B. A., Gallardo, H. F., Ungers, G. E., & Gilboa, E. (1990) *Cell 63*, 601–608.
Sullenger, B. A., Gallardo, H. F., Ungers, G. E., & Gilboa, E. (1991) *J. Virol. 65*, 6811–6816.
Varani, G., & Tinoco, I. (1991) *J. Am. Chem. Soc. 113*, 9349–9352.
Varani, G., Wimberly, B., & Tinoco, I. (1989) *Biochemistry 28*, 7760–7772.
Weeks, K., & Crothers, D. M. (1991) *Cell 66*, 1–20.
Weeks, K., Ampe, C., Schultz, S. C., Steitz, T. A., & Crothers, D. M. (1990) *Science 249*, 1281–1285.
Williamson, D., & Bax, A. (1988) *J. Magn. Reson. 76*, 174–177.
Wu, F., Garcia, J. A., Sigman, D. S., & Gaynor, R. B. (1991) *Genes Dev. 5*, 2128–2140.
Wu, H. N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221–8227.