Structural Features of the Trans-Activation Response RNA Element of Equine Infectious Anemia Virus†

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

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

Received August 3, 1992; Revised Manuscript Received October 23, 1992

ABSTRACT: A 25-nucleotide RNA with the sequence of the trans-activation response (TAR) element of equine infectious anemia virus (EIAV) was analyzed by biochemical methods and by one- and two-dimensional NMR spectroscopy. NMR, nuclease probing, and polyacrylamide gel migration rates show that the RNA consists of an A-helical stem capped by two non-Watson–Crick U-G base pairs and a compact four-nucleotide loop. The loop is stabilized by base stacking, with loop nucleotides C12 and C15 stacked upon U11 and G16, respectively. Near the 5′ end of the molecule, the stem contains a bulge at nucleotide C2, most likely a result of base pairing between G1 and C25. A method for distinguishing RNA stem–loops from palindromic dimers is described and was used to confirm that the EIAV TAR RNA has a stem–loop structure at conditions used for NMR spectroscopy. RNase A and RNase T1 cleavage patterns are consistent with the structural features derived from the NMR data.

Equine infectious anemia is a disease of the horse immune system that is caused by the equine infectious anemia virus (EIAV)[1] (Cheevers & McGuire, 1985). This virus has recently become the focus of considerable scrutiny since it has several features in common with human immunodeficiency virus type 1 (HIV-1) (Carvalho & Derse, 1991). These lentiviruses are similar both in the organization of their genomes and in their infection of cells of the monocyte macrophage lineage. A more detailed knowledge of EIAV biology would directly contribute toward a better understanding of HIV-1.

A particularly interesting common feature of EIAV and HIV-1 is the trans-activation of transcription mediated by the Tat–TAR system (Dorn et al., 1990; Cullen & Green, 1989). Tat or trans-activator protein is encoded by the viral *tat* gene and binds specifically to the TAR (trans-activation response) element located at the 5′ end of the viral RNA. It appears that this places the Tat protein in such a way as to promote further stimulatory interactions with the transcriptional apparatus. Since this system has been shown to be essential for viral replication in HIV-1, it is a promising target for antiviral drugs and has been intensely studied [for a review, see Cullen (1990)].

HIV-1 Tat contains several distinct regions including an N-terminal core domain that is thought to perform the stimulatory function and a C-terminal arginine-rich domain that mediates the binding to the TAR element (Carroll et al., 1991). HIV-1 TAR contains a stem–loop structure closed by a 6-nucleotide loop and a pyrimidine bulge within the stem (Dingwall et al., 1989; Muesing et al., 1987; Harper & Logsdon, 1991; Berkhout, 1992; Colvin & Garcia-Blanco, 1992). Extensive mutational studies have been performed on TAR, and many mutations disrupt the Tat-mediated increase in gene expression (Muesing et al., 1987; Feng & Holland, 1988; Berkhout & Jeang, 1989; Dingwall et al., 1989). There now exists a rather detailed picture of the functional sites within the element; for example, Tat clearly binds to the pyrimidine bulge within the RNA stem (Weeks et al., 1990; Weeks & Crothers, 1991; Calnan et al., 1991). More recently, the HIV-1 TAR element and its complex with arginine have been analyzed by NMR methods, and the stem–loop/bulge structure has been confirmed (Puglisi et al., 1992).

The EIAV Tat and TAR have been less well characterized but appear to operate in the same general manner. The Tat protein has a similar domain structure to its HIV-1 counterpart in which the N-terminal half mediates activation and the C-terminal half recognizes and binds TAR (Carroll et al., 1991). The TAR RNA is also predicted to be a stem–loop structure, and although a direct binding to EIAV Tat has yet to be demonstrated, mutations in the loop and stem nucleotides have a large effect on the Tat-mediated response which suggests a Tat–TAR interaction (Carvalho & Derse, 1991). However, a more detailed comparison reveals important differences between the EIAV and HIV-1 Tat–TAR systems. EIAV Tat lacks a rather conserved cysteine-rich region (Carroll et al., 1991), the nucleotide sequence of EIAV TAR does not predict a pyrimidine bulge, and there are no obvious similarities in the TAR loop sequences (Carvalho & Derse, 1991).

In this paper we report an analysis of the structure of the EIAV TAR element by NMR spectroscopy. Our aim is to identify any common features between the EIAV and HIV-1 TAR elements that might form the basis of the Tat–TAR interaction and to attempt to understand the structural basis of the differences between the two systems. This work also represents a continuation of our efforts to understand the structural aspects of RNA–protein interactions (Hoffman et al., 1991; Wilson et al., 1986; Ramakrishnan & White, 1992).

† This work was supported in part by a National Science Foundation grant (MCB 9118369 to S.W.W.) and by start-up funds by the Duke University Medical Center (M.A.G.-B.).

* Author to whom correspondence should be addressed.
‡ Department of Microbiology.
§ Section of Cell Growth, Regulation and Oncogenesis.
‖ Department of Medicine.

[1] Abbreviations: EIAV, equine infectious anemia virus; 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; NTPs, nucleotide triphosphates; 2QF-COSY, two quantum-filtered correlated spectroscopy; TAR, trans-activation response; Tat, trans-activator protein; TOCSY, total correlated spectroscopy; TSP, 3-(trimethylsilyl)-1-propanesulfonate.

0006-2960/93/0432-1096$04.00/0

MATERIALS AND METHODS

RNA Synthesis and Purification. The 25-nucleotide EIAV TAR RNA with the sequence 5′-GCACUCAGAUUCUGCGGUCUGAGUC-3′ was transcribed using T7 RNA polymerase, essentially as described by Milligan et al. (1987). Template DNA consisting of the oligonucleotides 5′-TAATACGACTCACTATAG-3′ and 5′-GACTCAGACCGCAGAATCTGAGTGCTATAGTGAGTCGTATAA-3′ was synthesized on an Applied Biosystems Model 391 DNA synthesizer and purified by polyacrylamide 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 EIAV RNA by varying the concentrations of RNA polymerase, template DNA, nucleotide triphosphates, and magnesium. Preparative-scale transcription reactions containing 3 mM NTPs, 13 mM $MgCl_2$, and 300 nM template DNA were incubated for 4–6 h at 37 °C, then ethanol precipitated, and dissolved in 8 M urea. The RNA was then separated from transcripts of incorrect size by electrophoresis on 15–20% polyacrylamide gels under denaturing conditions (8 M urea). RNA was visualized by UV shadowing and then removed from the gel using a Bio-Rad Model 422 electroeluter. The RNA was further purified by ethanol precipitating three times, and it was finally passed through a Sephadex G25 gel filtration column in 2 mM phosphate buffer and lyophilized. A typical yield was 1 mg of purified RNA/10 mL of transcription reaction volume.

Oligoribonucleotides with sequences 5′-GCAGAGAGAUUCUGCGGUGAGAGUC-3′ (RNA A) and 5′-GCACUCUCAUUCUGCGGUCUCUCUC-3′ (RNA B) were prepared from 5-mL transcription reactions as described above. DNA templates consisted of 5′-TAATACGACTCACTATAG-3′ oligonucleotide annealed to 5′-GACTCTCACCGCAGAATCTCTCTGCTATAGTGAGTCGTATAA-3′ for RNA A and 5′-GAGAGAGACCGCAGAATGAGAGTGCTATAGTGAGTCGTATAA-3′ for RNA B.

Dimer versus Monomer Assay. Solutions of EIAV TAR RNA, designed dimer RNA consisting of a mixture of RNA A and RNA B, and separate RNA A and RNA B were prepared in NMR buffer (15 mM phosphate, pH 6) at concentrations of 6 mg/mL in volumes of 0.02 mL. The RNA samples were heated to 90 °C for 3 min and then cooled on ice. Five milliliters of each RNA solution was then loaded onto a cold (4 °C) 18% polyacrylamide gel in a buffer of 50 mM Tris, 30 mM borate, and 2 mM EDTA, pH 7.8. Electrophoresis was carried out at 7 V/cm for 6 h at 4 °C. The RNA bands were initially visualized by UV shadowing. For a more permanent record, the gel was stained for 10 min with Stains-all (Sigma Chemical Co.), destained with water for several hours, and photocopied. Gels with several polyacrylamide concentrations between 9% and 18% were used; the 18% gel resulted in the best resolution of the 25-nucleotide RNAs.

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 units of RNase T1 for 30 min on ice or with RNase A for 10 min on ice, 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]. 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.

NMR Experiments. NMR data were collected at 500 or 600 MHz on Varian Unity 500 or 600 spectrometers, and in some cases at 500 MHz on a Bruker AMX spectrometer. Spectra in D_2O solvent were typically obtained using samples with 5–12 mg of RNA dissolved in 550 μL of 15 mM Na/K phosphate buffer and 0.1 mM EDTA at pH 6.0. Samples were heated to 90 °C for 3 min and then cooled on ice prior to NMR analysis. Proton spectra in 90% H_2O/10% D_2O solvent were obtained using the 1–1 (jump and return) method (Hore, 1983). 2D nuclear Overhauser effect spectra were acquired using the method of Macura and Ernst (1979), with mixing times between 200 and 280 ms. 2QF-COSY spectra were acquired by the method of Müller et al. (1986). TOCSY spectra with mixing times of between 40 and 80 ms were acquired by the method of Bax and Davis (1985), modified by a 50-ms "clean TOCSY" delay as described by Ernst (1988). All 2D NMR were acquired in the phase-sensitive mode by the time-proportional phase incrementation method (Marion & Wüthrich, 1983). 2D spectra were typically acquired at 500 (or 600) MHz with 450–600 blocks of 2048 complex points, using sweep widths of 4000 (or 4800 Hz) for spectra in D_2O and 12200 Hz for spectra in H_2O. Two-dimensional NOE data were acquired at 2, 15, 20, 27.5, and 30 °C. H1′–H2′ coupling constants were estimated from the separation of antiphase components in a 2QF-COSY spectrum acquired with a resolution of 1.2 Hz per point in the t_2 dimension, zero-filled once to a resolution of 0.6 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 EIAV TAR was acquired at 600 MHz proton frequency using the pulse sequence of Bax et al. (1983), with a $^1/_2J$ delay of 2.2 ms. A total of 200 blocks of 1024 complex points was acquired with 128 scans per block, for a total acquisition time of 18 hours. A sweep width of 5500 Hz was used for protons, and a sweep width of 18 000 Hz centered at 108 ppm was used in the ^{13}C dimension.

T_1 relaxation measurements were performed using the inversion–recovery method at 27.5 °C on a sample of 1 mM EIAV TAR in D_2O solvent, 15 mM phosphate buffer, pH 7.0. A 16-s recycle time was used between scans, which exceeded four times the T_1 of the slowest relaxing proton in EIAV TAR. Sixteen scans were acquired for each of 12 inversion–recovery delays between 0 and 8.0 s.

RESULTS

Verification of Stem–Loop Structure. All RNA stem–loops have the potential to form palindromic dimers, where the nucleotides corresponding to the loop in a stem–loop structure could form an internal loop in the dimer structure. For example, the EIAV TAR can in principle form either of the structures shown in Figure 1, parts A and B. A stem–loop is extremely difficult to distinguish from a symmetric dimer on the basis of NMR data alone, since the set of interproton distances is almost the same in either structure. Several methods have previously been used to distinguish RNA stem–loops from dimers, including gel filtration chromatography (Puglisi et al., 1990; Varani et al., 1991), nuclease cleavage assays, and melting curves monitored by UV absorbance. None of these methods provides an unambiguous test of monomer

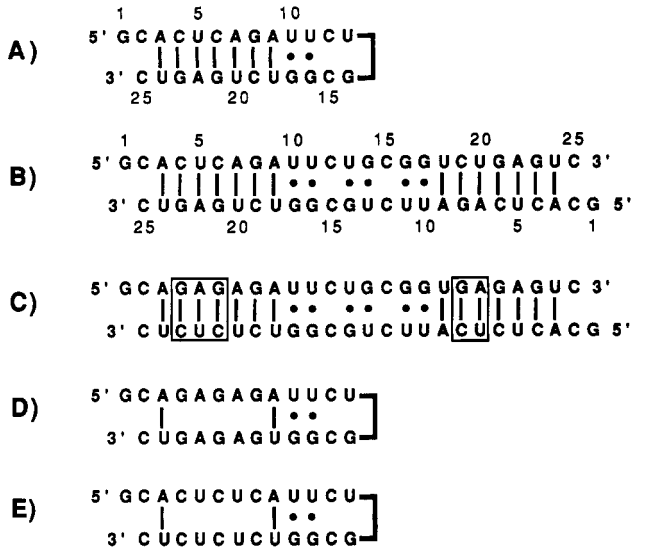

FIGURE 1: (A) Schematic diagram of the stem–loop structure of the EIAV TAR RNA. (B) Diagram of a palindromic dimer that could potentially be formed by the EIAV TAR. (C) Dimeric RNA consisting of strands RNA A and RNA B, designed to have a structure similar to that of the EIAV TAR dimer model. Differences between this designed dimer RNA and the EIAV dimer model are boxed. (D and E) Stem–loops that can be formed by RNA A and RNA B, respectively. These structures have little stability; therefore, when RNA A and RNA B are mixed, a dimer structure is much more likely than separate stem–loops.

versus dimer structure. Indeed, the results of these experiments can often be rationalized as supporting either structural model. For example, gel filtration chromatography and gel migration rates are sensitive to molecular shape in an unpredictable and complex manner. It would not be surprising for a compact, helical dimer to migrate faster than a hairpin with a bulky loop. Melting behavior cannot be unambiguously interpreted, since mismatched base pairs in a dimer or a loop in a monomer can either stabilize or destabilize an RNA structure, and melting may not be a simple one-step transition. Nuclease cleavage assays are highly sensitive to RNA structure, in the sense that small changes in structure are likely to result in changes in the cleavage pattern. However, either a hairpin loop or mismatched base pairs in a dimer could be rationalized as likely sites of nuclease cleavage. Recently, NMR spectroscopy has proved to be a powerful method for studying RNA structure, and several NMR studies of small RNA stem–loops have been reported (Puglisi et al., 1990; Cheong et al., 1990; Heus & Pardi, 1991a; Varani et al., 1991). In one instance, an RNA which formed a hairpin monomer in solution was found to form a dimer in the crystal structure (Holbrook et al., 1991). Clearly, an independent method for determining whether an RNA forms a stem–loop or dimer structure under the conditions of an NMR experiment would be of significant value.

The EIAV TAR almost certainly forms a stem–loop structure in vivo. It would not be surprising if the high concentrations required for the NMR experiments favor dimer formation. The dimer structure would probably be quite stable, since the "bulge" region would contain U-G base pairs and only two C-C mismatched pairs (Figure 1B). In an effort to unambiguously determine which secondary structure is adopted by the EIAV TAR under the conditions of the NMR experiment, two 25-nucleotide RNAs were prepared with sequences as shown in Figure 1, parts D and E. These RNAs, referred to as RNA A and RNA B, were designed so that they have little potential to form stem–loops, but when annealed together they will form a structure that is extremely similar

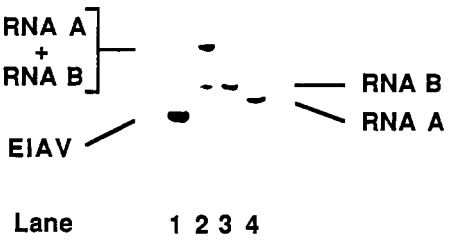

FIGURE 2: 18% polyacrylamide gel showing the relative migration rates of the EIAV TAR stem–loop (lane 1), the designed dimer RNA consisting of strands RNA A plus RNA B (lane 2), strand RNA A (lane 3), and strand RNA B (lane 4). Lane 2 contains a slight molar excess of RNA A.

to an EIAV TAR dimer. The only differences between the EIAV dimer model and the designed dimer are GC and AU base pairs which are switched to CG and UA (boxed nucleotides in Figure 1C). Otherwise the structures are identical, including the central bulge and mismatched ends. The base pair switches were made in regions of the EIAV TAR which are helical, independent of whether the dimer or stem–loop structure is assumed, and are flanked on either side of the helical structure.

Nondenaturing gel electrophoresis was chosen as a method for comparing the EIAV TAR with the designed dimer RNA. Ideally, comparison of the RNAs should be made under the conditions of the NMR experiment. Although it is not possible to simulate the NMR conditions within the nondenaturing gel, it is easily possible to anneal the RNA strands at the conditions of the NMR experiment. After being annealed, the RNAs were kept cold to minimize the chance of interchange between the possible dimer and stem–loop structures. Polyacrylamide gels were run in the cold and at a slow migration rate of 1–2 cm/h for the RNA species (7 V/cm), to ensure that the RNA structures were not disrupted. Figure 2 shows an 18% polyacrylamide gel comparing the migration rates of the EIAV TAR RNA, the designed dimer of RNAs A and B, and separate RNA A and RNA B. The EIAV TAR RNA migrates much faster than the designed dimer, indicating that these RNAs are significantly different in molecular shape. Thus, it is very likely that the EIAV TAR adopts a monomeric stem–loop conformation rather than a dimer conformation. The fast migration of the stem–loop RNA compared to the RNA dimer and the single-stranded RNAs is indicative of its compact structure. The relative migration rates of the RNAs are somewhat sensitive to the percentage of acrylamide in the gel. For example, on a 9% gel the RNA dimer and the single-stranded RNAs have similar migration rates, which are all slower than the migration of the EIAV TAR stem–loop RNA.

This strategy for determining whether an RNA forms a stem–loop or dimer can be generally applied to other RNA sequences. For any potential RNA stem–loop, a dimer RNA can be intentionally constructed by switching GC or AU base pairs in the stem to CG and UA pairs, respectively, in a manner analogous to that described for the EIAV TAR RNA (Figure 1). A physical property such as migration rate on a polyacrylamide gel can then be compared for the potential RNA stem–loop and designed RNA dimer.

NMR Analysis. (*i*) *Spectral Features.* NMR spectroscopy was used to determine the structural features of the EIAV TAR stem–loop. The one-dimensional spectrum in D_2O solvent (Figure 3) shows that the proton resonances are generally sharp, with line widths of typically 5–10 Hz. The nonexchangeable protons resonate in a relatively narrow chemical shift range between 3.9 and 8.3 ppm, as is typical

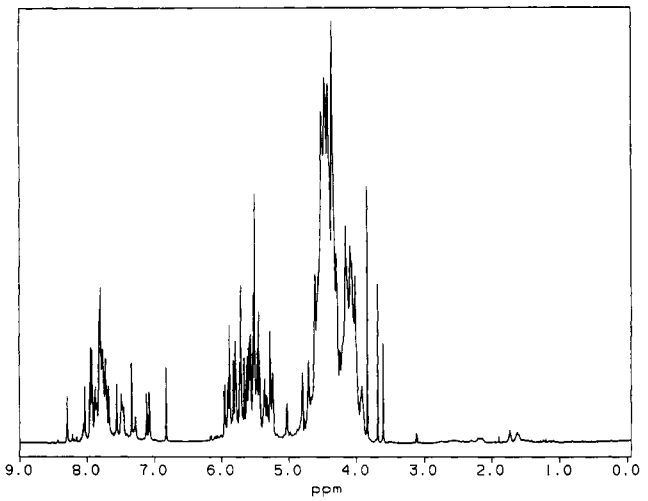

FIGURE 3: One-dimensional proton NMR spectrum of EIAV TAR in D_2O solvent at 600 MHz. The sample contains 15 mM phosphate buffer and 0.2 mM EDTA, at pH 6.0, 20 °C.

for RNA molecules. The spectrum is indicative of the high quality of the RNA sample. A few small resonances between 1.0 and 3.2 ppm are attributed to residue from the acrylamide gels used in purifying the RNA, and sharp peaks between 3.6 and 3.9 ppm are due to added EDTA. The spectrum also contains some minor peaks, indicating the presence of very small amounts of other RNA species. In H_2O solvent, sharp resonances are observed for 10 imino protons that have relatively slow exchange rates with the solvent (Figure 4). Detailed structural information can be extracted from the NMR data only after the resonances have been assigned to the specific protons from which they arise. Spectrum assignment was accomplished using a combination of 2D homonuclear and heteronuclear methods. Chemical shift data are summarized in Table I.

(ii) A-Helical Stem (Nucleotides 3–9 and 18–24). Assignments were first made for the resonances of the A-helical stem region of the RNA molecule, which includes nucleotides A3–A9 and U18–U24. The H6 protons of the pyrimidines were distinguished from the H8 protons of the purines by the presence of H5–H6 correlations in the 2QF-COSY spectrum, providing convenient starting points for making spectrum assignments. H5–H6 correlation peaks for all pyrimidines in the EIAV TAR were resolved (Figure 5). Using the NOE spectrum, it was then possible to sequentially "walk" from the H6 or H8 proton of a base to the H1′ and H2′ protons of a ribose, then to the H6 or H8 proton of the next base, as illustrated in Figure 6. NOEs were also observed from the H2 protons of the four adenosines to the H1′ proton of the ribose on the $i + 1$ nucleotide and to the H1′ of a cross-strand ribose (Figure 6), confirming the assignments of the stem nucleotides. NOEs were observed from the H5 protons of the stem pyrimidines to the H8 or H6 proton of the $i-1$ nucleotide, as expected for an A-helix where this distance is normally 3.8 Å (Figures 6 and 7). In several cases weak NOEs are detected between the sequential base-stacked H8 and H6 protons.

A two-dimensional NOE spectrum in 90% H_2O/10% D_2O solvent was used to assign exchangeable proton resonances. Uridine imino protons in the A-helical stem were identified by their strong NOEs to the H2 proton of their base-paired adenosine (Figure 4). The amino protons of the stem cytosines were assigned by NOEs to their own H5 protons and to cross-strand G imino protons. No amino resonances were assigned for A3, and no imino resonance was assigned for U24. Presumably these resonances were broadened due to fast exchange with the solvent, a common observation for exchangeable protons near the end of a helical stem.

H1′–H2′ coupling constants are sensitive to ribose conformation. These couplings are small for the stem nucleotides, and the corresponding peaks in the 2QF-COSY spectrum are weak or absent. Small H1′–H2′ coupling constants indicate that the riboses within the A-helical stem are primarily in the C3′ endo conformation, as is normally observed in A-helical RNA (Puglisi et al., 1990). The A-helical stem of EIAV TAR consists of 14 nucleotides; resonance assignment of the remaining 11 nucleotides was somewhat more complicated since no particular structure could be assumed in interpreting the NMR data.

(iii) UG Base Pairs. Nucleotides U10, U11, G16, and G17 form two "wobble" type base pairs (Crick, 1966) that are stacked on the top of the A-helical stem, with apparently little distortion from the regular A-helical geometry. Each UG base pair contains two hydrogen bonds, with the uridine imino proton hydrogen bonding to the guanosine O6 carbonyl and the guanosine imino proton hydrogen bonding with the uridine O2 carbonyl. Resonance assignments for the two UG base pairs were relatively straightforward, using sequential ribose–base–ribose NOE pathways extending from the well-defined A-helical stem (Figures 6 and 7). The intensities of the sequential H6/H8 to H1′/H2′ NOEs are similar to those observed for the A-helix. The normal base-stacking NOEs connecting A9 H8 to U10 H5 and U10 H6 to U11 H5 are observed, as well as a strong NOE from the A9 H2 resonance to the U10 H1′ (Figure 6). A strong NOE was clearly observed between the two imino protons of the U11-G16 base pair (Figure 4), as well as weaker NOEs to the iminos in the U10–G17 base pair. The U10-G17 imino protons have nearly identical chemical shifts, differing by only 0.05 ppm, which prevented the reliable observation of the NOE between them; however, these protons could be assigned by their NOEs to the imino of U18 (Figure 4). NOEs from the U10 and G17 imino protons to the H2 and amino protons of A9 were also observed and were slightly stronger for the imino of U10 than that of G17. The 2D NOE spectrum shows that all four of the imino protons in the UG base pairs have strong exchange peaks with the H_2O resonance, due to their relatively rapid exchange with the solvent. The riboses of U10, G16, and G17 have weak H1′–H2′ couplings and are, therefore, primarily in the C3′ endo conformer family, as is usually found in an A-helix. The ribose of U11 exhibits a strong H1′–H2′ coupling, indicating that it is primarily C2′ endo. The H2′ resonances of U10 and U11 are shifted slightly upfield of the other H2′ resonances in the A-helical stem.

The U-G base pairs at the top of the stem are consistent with the model derived from mutational data (Carvalho & Derse, 1991). In general, the structural features and observed NOEs for the two UG base pairs in the EIAV TAR RNA are similar to those described for a UG base pair in another RNA stem–loop (Puglisi et al., 1990) and GT base pairs in B-form DNA (Hare et al., 1986; Quignard et al., 1987; Kalnik et al., 1988). In all cases the mismatched base pair has been found to form the same imino–carbonyl hydrogen bonds and is incorporated into the helix with relatively little distortion.

(iv) Hairpin Loop (Nucleotides 12–15). Nucleotides C12, U13, G14, and C15 form a compact 4-nucleotide loop at the top of the RNA hairpin, linking the U11-G16 base pair. The bases of C12 and C15 are stacked on U11 and G16, respectively, while U13 and G14 form a tight 2-nucleotide turn at the top of the hairpin loop. The H6 resonance of C12 was assigned from NOEs to H1′ and H2′ of U11 (Figure 7).

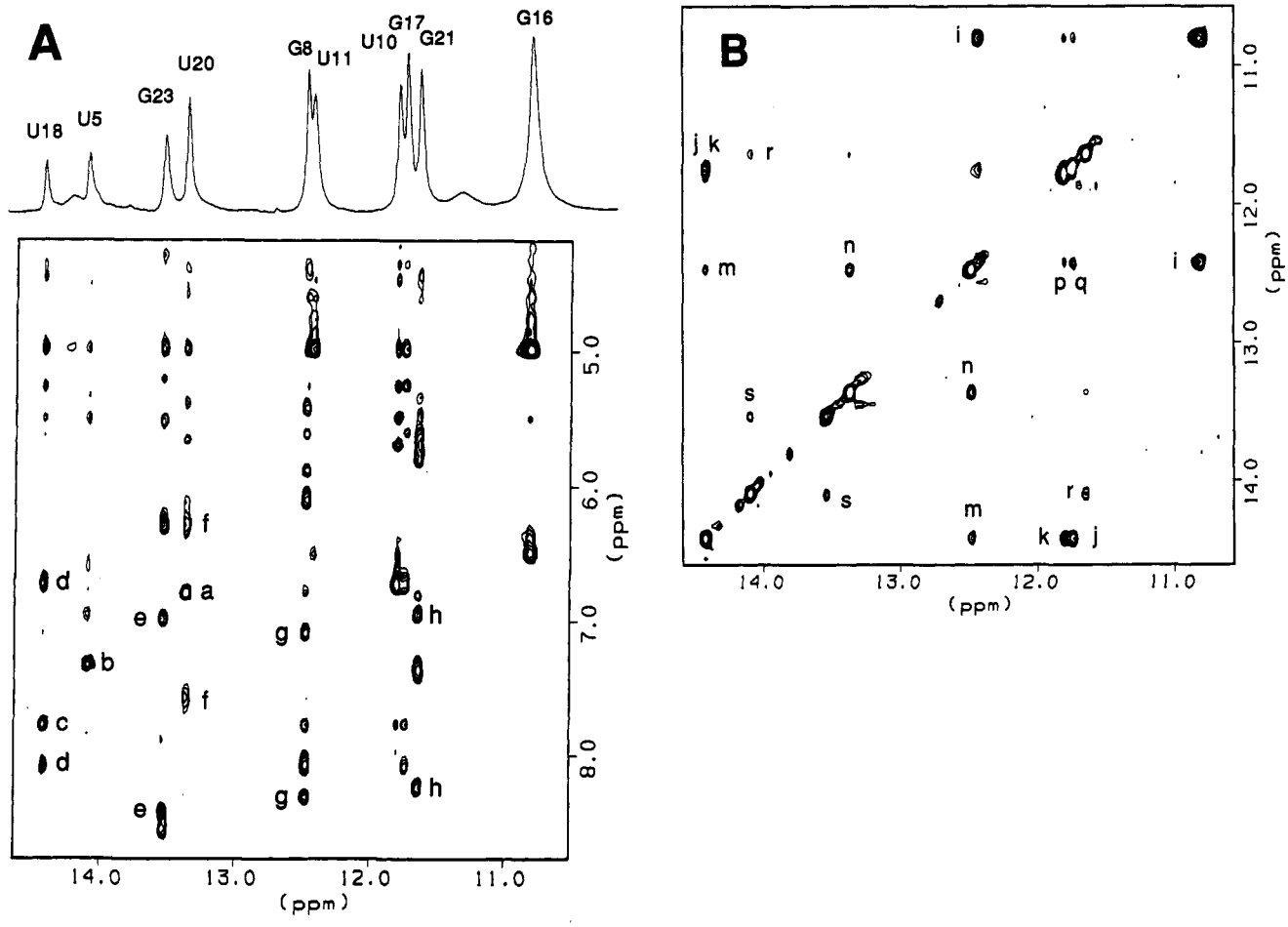

FIGURE 4: (A) Top: Section of the 1D NMR spectrum of the EIAV TAR RNA in 90% H_2O/10% D_2O solvent and 15 mM phosphate buffer, pH 6.0, at 2 °C showing assignments of imino proton resonances. Although the RNA contains a total of 14 imino protons, one for each G and U, only 10 sharp imino resonances are observed. Two very broad resonances, near 14 and 11 ppm, are likely due to the unassigned imino protons of G1, U13, G14, and U24. Bottom: A section of a 2D NOE spectrum in H_2O solvent (280 ms mixing time), with several assigned cross-peaks labeled. Peaks a, b, and c are NOEs from the stem adenosine H2 resonances to the imino proton of its base-paired uridine and provided starting points for sequence-specific resonance assignment. Other NOE peaks are labeled as follows: (d) U18 imino-A9 amino; (e) G23 imino-C4 amino; (f) U20 imino-A7 amino; (g) G8 imino-C19 amino; (h) G21 imino-C6 amino. NOEs are also observed from the iminos to H1′ and H5 resonances between 5.2 and 6.0 ppm (Heus & Pardi, 1991b). All imino protons have exchange peaks with the H_2O resonance at 5.0 ppm. (B) Another section of the same NOE spectrum, showing NOEs between imino protons. Peaks are labeled as follows: (i) U11 H3-G16 H1; (j) G17 H1-U18 H3; (k) U18 H3-U10 H3; (m) G8 H1-U18 H3; (n) G8 H1-U20 H3; (p) U10 H3-U11 H3; (q) G17 H1-U11 H3; (r) U5 H3-G21 H1; (s) U5 H3-G23 H1.

Similarly, the H6 of C15 was assigned from its NOE to H8 of G16; the H1′ of C15 has NOEs to its own H1′ and the H1′ of G16. NOEs from U11 H6 to C12 H5 and from G16 H8 to C15 H5 confirm the U11-C12 and C15-G16 base stacking; orientation of the bases U13 and G14 is less certain. Nonexchangeable base protons of U13 and G14 were assigned by default, since the 2QF-COSY and NOE spectra indicated resonances of only one purine and one pyrimidine that were not clearly assigned to other nucleotides in he EIAV TAR RNA sequence. The H8 resonance of G14 has a very strong NOE to an H1′, assigned to that of its own ribose. H1′–H2′ couplings indicate that riboses of C12–C15 are primarily in the C2′ endo family of conformers, as is commonly observed for RNA loops.

(*v*) *5′ and 3′ End Nucleotides.* G1, C2, and C25, located at the 5′ and 3′ ends of the stem, show substantial deviation from regular A-helical structure. The sequence of the EIAV TAR is such that these nucleotides would not be expected to normally base pair (Figure 1A). Resonances of these three nucleotides were assigned using NOE pathways from A3 and U24 in the A-helical part of the stem. Nucleotide C25, at the 3′ end of the stem, was assigned from the NOEs of its H5 and H6 protons to the ribose of U24. NOEs between the H6 protons of U24 and C25 were observed, as well as an NOE from the H6 of U24 to the H5 of C25. These NOE data indicate base stacking of U24 upon C25. The ribose of C25 is in the C2′ endo conformation as indicated by a large H1′–H2′ coupling constant. The chemical shift of the H2′ resonance of C25 is shifted upfield (to 4.03 ppm), as is commonly observed for the H2′ at the 3′ end of A-helices (Varani et al., 1989). Resonances of G1 and C2 were assigned using sequential H8-H6-H8 NOEs connecting G1 with C2 and C2 with A3. The H6 of C2 was distinguished from H8 of G1 by the presence of the H5-H6 peak in the 2QF-COSY spectrum (Figure 5). NOEs were identified which connect the H8 proton of G1 with H1′ and H2′ protons of ribose 1; similarly, NOEs were identified between the H6 and H5 protons of C2 with the H2′ and H3′ protons of ribose 2. The H8 proton on the purine ring of A3 has NOEs to protons of *both* riboses 1 and 2 as well as its own ribose (Figures 6 and 7). These data are most consistent with nucleotide G1 being significantly closer to A3 than would be expected if G1, C2, and A3 were all normally base stacked. This suggests the formation of a base pair between G1 and C25, resulting in a single nucleotide bulge at C2. Although no slowly exchanging imino and amino protons were assigned for G1 and C25, this is not inconsistent

Table I: Resonance Assignments for EIAV TAR RNA[a]

	H8/H6	C8/C6	H5/H2	C5/C2	H1′	H2′	J(1′–2′)	imino/amino
G1	7.95	138.2	na[b]	na	5.63	4.71	9	
C2	7.87	143.0	5.96	97.2	5.87	4.41	8	
A3	8.30	139.7	7.94	153.3	5.90	4.75	4	
C4	7.51	140.0	5.27	95.9	5.48	4.32	<3	8.44, 7.01
U5	7.86	141.0	5.35	102.0	5.56		<3	14.10
C6	7.79	140.2	5.65	96.4	5.49	4.46	<3	8.27, 6.97
A7	7.96	138.2	6.86	151.5	5.84	4.64	<3	
G8	7.11	134.4	na	na	5.53	4.53	<3	12.50
A9	7.57	137.9	7.83	153.1	5.91	4.62	<3	8.10, 6.75
U10	7.47	139.0	5.26	102.8	5.33	4.11	3	11.80
U11	7.88	141.5	5.60	103.4	5.77	4.28	7	12.43
C12	7.76	142.7	5.92	96.8	5.62	4.49	7	
U13	7.73	142.0	5.60	103.4	5.53	4.34	9	
G14	7.80	137.3	na	na	5.60	4.49	7	
C15	7.77	142.0	5.59	96.1	5.83	4.46	9	
G16	8.00	137.8	na	na	5.49	4.74	<3	10.82
G17	7.42	136.2	na	na	5.66	4.33	<3	11.75
U18	7.75	140.2	5.29	101.8	5.54	4.33	<3	14.42
C19	7.82	140.3	5.66	96.4	5.54	4.37	<3	8.33, 7.11
U20	7.76	142.7	5.38	102.4	5.47	4.44	<3	13.37
G21	7.68	135.1	na	na	5.69	4.57	<3	11.65
A22	7.72	138.0	7.36	152.4	5.85	4.59	<3	8.56, 6.30
G23	7.08	134.3	na	na	5.56	4.36	<3	13.55
U24	7.52	140.0	5.02	101.9	5.66	4.33	3	
C25	7.81	142.1	5.81	96.9	5.77	4.03	9	

[a] Assignments of chemical shifts (ppm) and intraribose H1′–H2′ coupling constants (Hz) are for EIAV TAR RNA in 15 mM phosphate buffer, pH 6. Data for nonexchangeable protons are reported at 30 °C, with chemical shifts relative to the residual HDO resonance at 4.7 ppm. Amino and imino resonances are reported at 2 °C where the exchange rate with the solvent is reduced. Carbon chemical shifts are relative to external TSP. H1′–H2′ couplings that were undetected are reported as <3 Hz, other couplings have an uncertainty of 1 Hz. [b] na = not applicable.

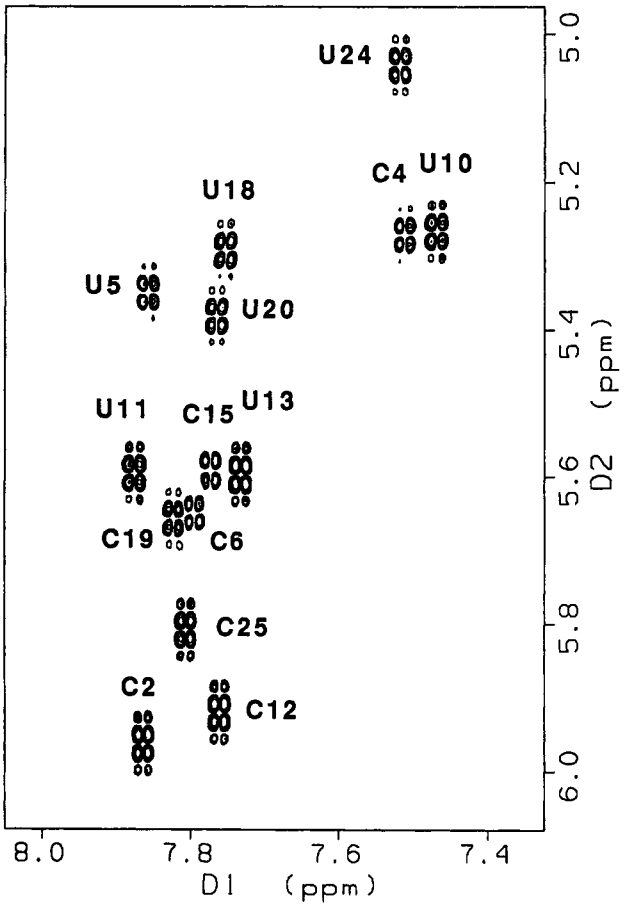

FIGURE 5: Section of the 2QF-COSY spectrum of EIAV TAR showing assignments of the H5–H6 resonances in the 14 pyrimidine rings.

with the base pair formation since it is not surprising for hydrogen-bonded protons in base pairs at the end of a stem to rapidly exchange with the solvent.

(*vi*) *Additional NMR Results.* The 25-nucleotide EIAV TAR is among the largest RNA molecules for which substantial NMR spectrum assignments have been made. Additional experimental approaches were essential in establishing the reliability of the spectrum assignments. These experiments included (1) analyzing the temperature dependence of proton resonance frequencies and (2) ^{1}H–^{13}C heteronuclear spectroscopy.

Two-dimensional homonuclear NMR data sets were acquired in D_2O solvent at a series of five temperatures between 2 and 30 °C. The benefits of these data were 2-fold. First, temperature-dependent changes in proton resonance frequencies could be exploited in identifying overlapping peaks in the 2D spectra; and second, the possibility of temperature-dependent structural changes could be assessed. Most protons in the EIAV TAR were found to exhibit small shifts of only a few one-hundredths of a ppm over the investigated temperature range. Though relatively small, these chemical shift changes were often valuable in assigning nearly overlapping resonances. In a few cases, larger temperature-dependent changes of up to 0.22 ppm were observed, primarily involving resonances of U10-G17 base pair. Two examples are the H8 resonance of G17 which moves from 7.30 to 7.42 ppm and H1′ of U10 which moves from 5.27 to 5.33 ppm between 15 and 30 °C. The intensities of NOEs involving these protons, however, show little change with temperature. The observed changes in resonance frequencies, therefore, are more likely reflective of the hypersensitivity of chemical shift to very small changes in the orientation of the bases rather than any substantial change in structure.

^{13}C NMR data has previously been used as an aid in assigning NMR spectra of DNA oligonucleotides (Leupin et al., 1987; LaPlante et al., 1988; Ashcroft et al., 1991). More recently, it has been shown that ^{13}C chemical shift information can also be a valuable tool in assigning spectra of RNA (Varani & Tinoco, 1991). A ^{1}H–^{13}C correlated HMQC spectrum was used in confirming the spectrum assignments for the EIAV

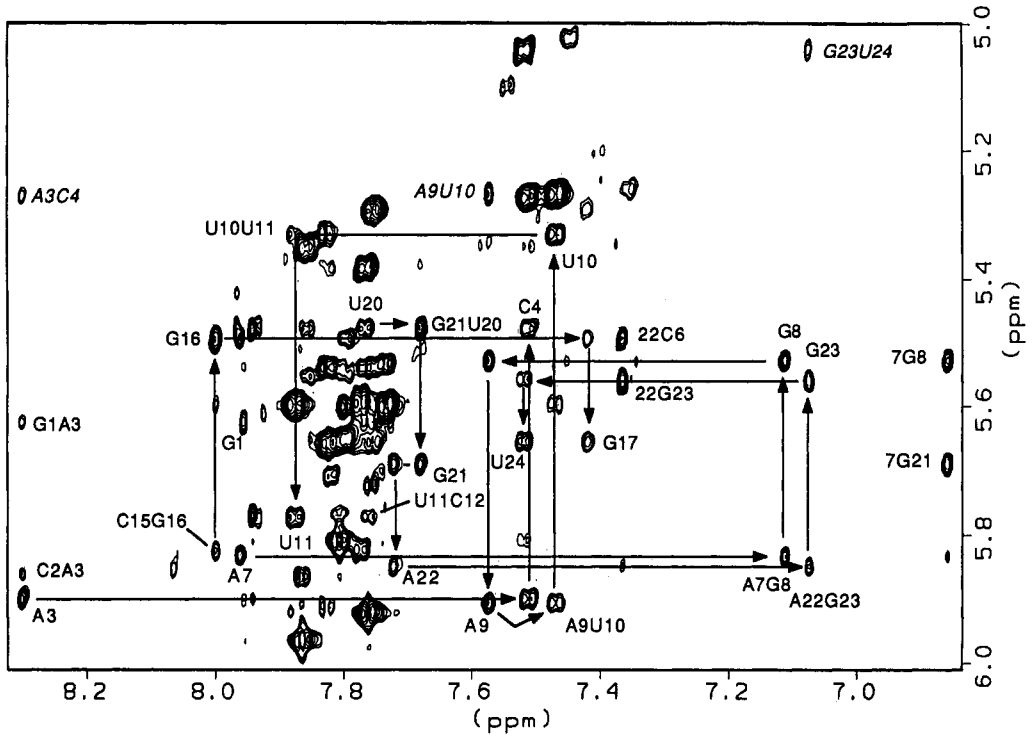

FIGURE 6: Section of the 2D NOE spectrum of the EIAV TAR in D_2O solvent, with a few of the assigned NOE peaks identified. Intranucleotide H6/H8-H1′ are labeled by type and number. Sequential H1′(i)–H6/H8(i+1) NOEs are labeled by the type and number of the two nucleotides. Sequential H6/H8(i)–H5(i+1) NOEs are labeled in italics by the type and number of the two nucleotides. NOEs from adenosine H2 to ribose H1′ resonances are labeled by the number of the adenosine and the type and number of the ribose H1′. For example, "A9" identifies the intranucleotide H8–H1′ NOE for A9, "G8A9" identifies the NOE from the G8H1′ to A9H8, "*A9U10*" identifies the NOE from A9H8 to U10H5, and "9C19" identifies the NOE from A9H2 to C19H1′. Examples of sequential NOE "walks" between nucleotides are shown by arrows.

TAR. The C2 resonances of the four adenosines exhibit characteristic downfield chemical shifts, clearly separated from the C6 and C8 resonances by about 8 ppm (Figure 8). Uridine and cytosine bases were clearly differentiated by the chemical shifts of their C5 carbons (Table I), thus adding an additional degree of confidence to the assignments of the UUCUGCGG loop resonances. Based on only their homonuclear H5-H6 couplings, uridines and cytidines would be indistinguishable. The ^1H–^{13}C HMQC spectrum of the EIAV TAR also suggests a correlation between carbon chemical shift and whether a resonance is from a pyrimidine C6 and purine C8, since the resonances of the these carbon types appear in two distinct groups (Figure 8).

Recently it was reported that H8, H6, and H1′ protons in an RNA helix can exhibit surprisingly long T_1 relaxation times, potentially resulting in substantial errors in distance measurements derived from NOE data (Wang et al., 1992). T_1 relaxation times were measured for the protons of EIAV TAR at 27.5 °C using the inversion–recovery method. In general, the T_1 relaxation times are significantly shorter than those reported by Wang et al. (1992). With the exception of three of the adenosine H2 protons, T_1's were found to be between 0.7 and 2.4 s, with most of the protons having T_1's of about 1.4 s. There was no obvious correlation of T_1 with the position of a proton within the stem–loop sequence. The ribose H1′ protons generally had longer T_1's than the other ribose protons, typically 2.2 s versus 1.2 s. The H2 protons of three adenosines in the EIAV TAR have significantly longer T_1's of 3.6 to 4.0 s, as is commonly observed for these protons in RNA and DNA helices. One adenosine, A3, has a relatively short T_1 of 1.5 s for its H2 resonance, which is about the same as most of the H6 and H8 resonances. It is interesting that no imino proton was observed for the A3-U24 base pair, suggesting

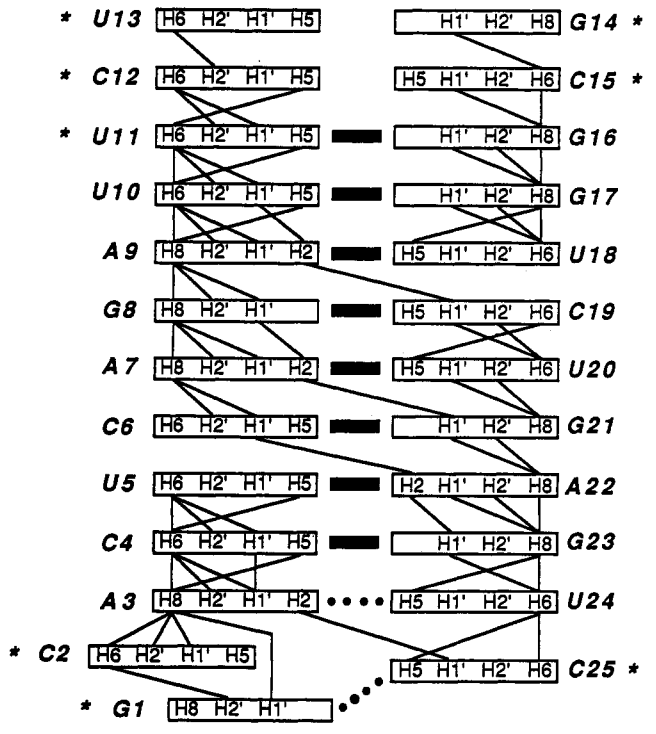

FIGURE 7: Schematic diagram summarizing internucleotide NOEs and hydrogen bonding in EIAV TAR. Base pairings confirmed by observed imino resonances are indicated by thick bars, and probable base pairings without observed imino resonances are indicated by dotted lines. Nucleotides with C2′ endo sugars are marked with asterisks. Assigned internucleotide proton–proton NOEs are indicated by lines; intranucleotide NOEs are not indicated. Unambiguous NOE peaks linking U5 with C6, U13 with G14, and U18 with C19 were not observed; several of these NOEs are probably present but unresolved due to overlap in the 2D spectra.

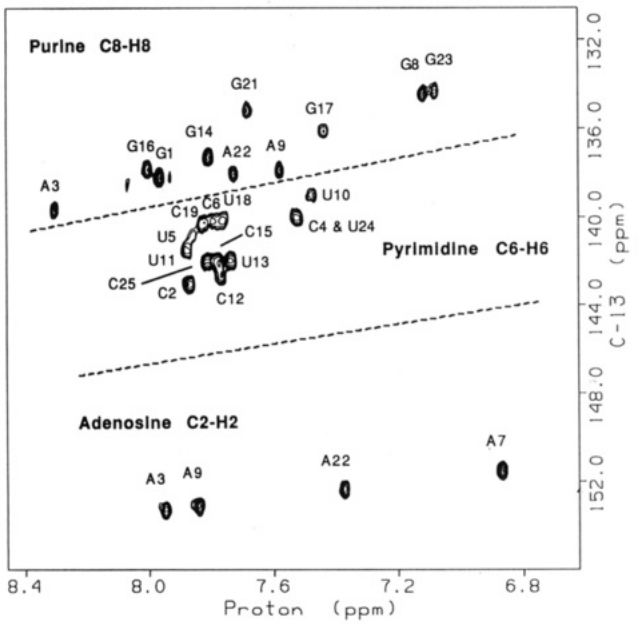

FIGURE 8: Section of the ^{13}C–^{1}H HMQC spectrum of the EIAV TAR RNA indicating assignments of C2, C6, and C8 carbons.

that the relatively short T_1 may be related to mobility of this base.

Nuclease Probing. Enzymatic cleavage assays are sensitive probes of RNA structure. RNase T1 preferentially cleaves RNA after single-stranded guanosines. Under enzyme-limited digestion, G14 was the only nucleotide of EIAV TAR cleaved by RNase T1 (Figure 9). Nucleotides G16 and G17 were protected from cleavage, as were the guanosines in the A-helical stem. The RNase T1 cleavage pattern is therefore exactly as would be predicted from the NMR data, which shows that G14 is the only unpaired guanosine.

RNase A is a structure-sensitive nuclease that usually cuts after single-stranded pyrimidines. RNase A cleaves the EIAV TAR at four sites: after U11, C12, U13, and C15 (Figure 9). The cleavage rates at the four sites are not equal, with U13 being the most frequently cleaved. This differential nuclease susceptibility is indicative of a complex loop structure. U13 is the most accessible loop pyrimidine to the enzyme and the least likely to be involved in interactions with other nucleotides, consistent with the NMR results which show that U13 is at the top of the hairpin loop. Although the NMR data show that U11 is base paired, RNase A often cleaves pyrimidines that are located at the top of stems (Moazed et al., 1986). The partial protection of C12 and C15 from RNase A is consistent with the NMR data, which shows these nucleotides to be base stacked upon the helical stem, though not base paired themselves.

Taken together, the RNase A and RNase T1 cleavage data are most consistent with an EIAV TAR element that contains a base-paired, helical stem capped by a 4-nucleotide hairpin loop consisting of C12, U13, G14, and C15. The nuclease cleavage patterns are, therefore, entirely consistent with the structural model derived from the NMR data.

DISCUSSION

A goal of this study is to use the structural information derived from the NMR data to evaluate the relationship between EIAV TAR structure and function. The functional importance of various stem and loop nucleotides was recently investigated by mutational analysis, where relative trans-activation efficiencies of wild-type and mutant EIAV TAR

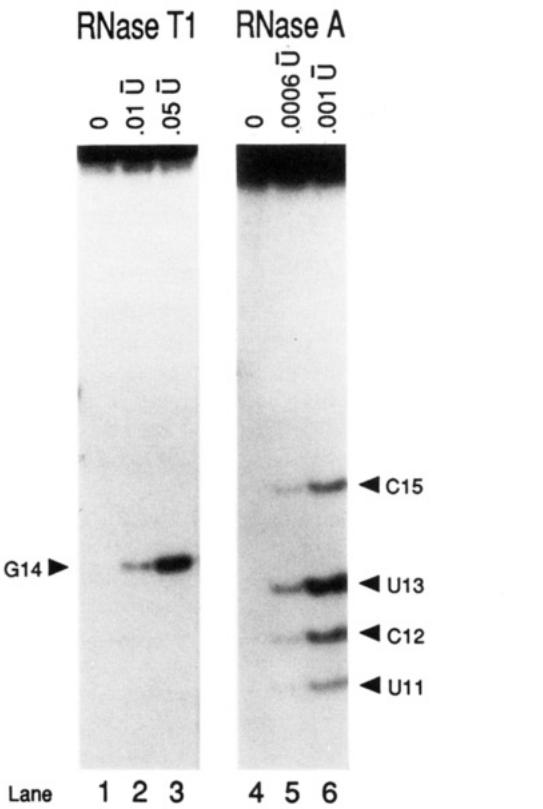

FIGURE 9: (A, left) Partial RNase T1 and RNase A digestions of EIAV TAR RNA. Residues where cleavage occurred are indicated. G13 was the only residue cleaved by RNase T1 at enzyme limited conditions. U11, C12, and U14 were cleaved by RNase A at enzyme limited conditions with the amount of enzyme indicated above each lane. (B, right) Summary of cleavages by RNase T1 and RNase A on EIAV TAR. Arrows indicate sites of cleavage by the enzyme. The relative frequency of cleavage is indicated by the relative size of the arrow.

RNAs were assessed through chloramphenicol acetyl transferase (CAT) assays (Carvalho & Derse, 1991). The NMR data provides an additional framework for interpreting the mutational results. Mutation of G14 was found to have a large negative effect on EIAV TAR function, while mutations of other nucleotides within the CUGC loop had less than 2-fold effects on activity. The NMR data and RNase T1 cleavage pattern each indicate that G14 is one of two nucleotides at the top of the tight turn in the hairpin loop and is, therefore, potentially accessible for interaction with the EIAV Tat protein or other factors. It is interesting that the HIV-1 TAR RNA also contains accessible guanosines at the top of its hairpin loop which are also essential for TAR function (Colvin et al., 1993). In HIV-1 these guanosines are probably involved in binding a cellular factor or factors that stabilize the interaction between Tat and TAR (Cullen, 1990).

The two UG base pairs were also found to be critical for EIAV TAR function. Specifically, mutations of U11-G16 to U-A and U10-G17 to U-U each resulted in a significant reduction in activity, leading to the suggestion that U10 and U11 may be the functional counterparts of the uridine bulge in HIV-1 TAR (Carvalho & Derse, 1991). Recent NMR results showed that in HIV-1 TAR, U23, the critical "bulge" nucleotide for Tat binding, remains stacked upon A22 prior to Tat binding (Puglisi et al., 1992). Similarly, U10 and U11 of EIAV TAR which are clearly base paired and incorporated into the helical stem may present a similar target for its respective Tat protein. It would not be surprising for the EIAV and HIV-1 TAR RNAs to share some common structural features, based on their functional similarity and

the sequence homology of their respective Tat proteins. Ultimately, an understanding of the structure and function of the EIAV TAR element may contribute to the understanding of trans-activation in all tat-containing lentiviruses.

A comparison of the mobilities of the EIAV TAR and a designed dimer RNA was used to verify that EIAV TAR forms a monomeric stem–loop structure, and not a dimer, at the conditions of the NMR experiments. It is important to note that all of the observed NOEs or base pairings summarized in Figure 7 are consistent with either a dimer or a stem–loop structure. The result of the gel mobility experiment was, therefore, crucial in accurately interpreting the NMR data.

The NMR results described in this paper provide a description of the major structural features of the RNA stem–loop of the EIAV TAR element. The structural information is derived primarily from proton–proton NOE data. As expected for an RNA molecule of this size, extensive overlap of cross-peaks within the 2D spectra limits the number of resolved NOEs and ultimately limits the resolution of the structural model. This overlap is apparent in the H6/H8–H1′ region of the NOE spectrum (Figure 6) and is even more extensive in the H6/H8–H2′ region. Recently it has been shown that heteronuclear and homonuclear three-dimensional NMR methods can be used to effectively resolve peaks which overlap in 2D NOE spectra of nucleic acids (Nikonowicz & Pardi, 1992; Radhakrishnan et al., 1992). Three-dimensional NMR data should permit the EIAV TAR structure to be determined at higher resolution. Experiments directed toward this goal are currently in progress.

ACKNOWLEDGMENT

We thank Stephanie Porter for purifying the T7 RNA polymerase and Xiaolian Gao for many helpful discussions.

REFERENCES

Ashcroft, J., Live, D. H., Patel, D. J., & Cowburn, D. (1991) *Biopolymers 31*, 45–55.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Bax, A., Griffey, R. H., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301–315.
Berkhout, B. (1992) *Nucleic Acids Res. 20*, 27–31.
Berkhout, B., & Jeang, K.-T. (1989) *J. Virol. 63*, 5501–5504.
Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., & Frankel, A. D. (1991) *Science 252*, 1167–1170.
Carroll, R., Martarano, L., & Derse, D. (1991) *J. Virol. 65*, 3460–3467.
Carvalho, M., & Derse, D. (1991) *J. Virol. 65*, 3468–3474.
Cheevers, W. P., & McGuire, T. C. (1985) *Rev. Infect. Dis. 7*, 83–88.
Cheong, C., Varani, G., & Tinoco, I. (1990) *Nature 346*, 680–682.
Colvin, R. A., & Garcia-Blanco, M. A. (1992) *J. Virol. 66*, 930–935.
Colvin, R. A., White, S. W., Garcia-Blanco, M. A., & Hoffman, D. W. (1993) *Biochemistry* (following paper in this issue).
Crick, F. H. (1966) *J. Mol. Biol. 19*, 548–555.
Cullen, B. R. (1990) *Cell 63*, 655–657.
Cullen, B. R., & Green, W. C. (1989) *Cell 58*, 423–426.
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.
Dorn, P., DaSilva, L., Martarano, L., & Derse, D. (1990) *J. Virol. 64*, 1616–1624.
Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.
Feng, S., & Holland, E. (1988) *Nature 334*, 165–167.
Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7445–7456.
Harper, J. W., & Logsdon, N. J. (1991) *Biochemistry 30*, 8060–8066.
Heus, H. A., & Pardi, A. (1991a) *Science 253*, 191–193.
Heus, H. A., & Pardi, A. (1991b) *J. Am. Chem. Soc. 113*, 4360–4361.
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.
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., Kouchakdjian, M., Li, B. F. L., Swann, P. F., & Patel, D. J. (1988) *Biochemistry 27*, 108–115.
LaPlante, S. R., Boudreau, E. A., Zanatta, N., Levy, G. C., & Borer, P. N. (1988) *Biochemistry 27*, 7902–7909.
Leupin, W., Wagner, G., Denny, W. A., & Wüthrich, K. (1987) *Nucleic Acid Res. 15*, 267–275.
Macura, S., & Ernst, R. R. (1979) *Mol. Phys. 41*, 95–117.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399–416.
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.
Nikonowicz, E. P., & Pardi, A. (1992) *Nature 355*, 184–186.
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.
Quignard, E., Fazakerley, G. V., van der Marel, G., van Boom, J. H., & Guschlbauer, W. (1987) *Nucleic Acids Res. 15*, 3397–3409.
Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry 31*, 2514–2523.
Ramakrishnan, V., & White, S. W. (1992) *Nature 358*, 768–771.
Varani, G., & Tinoco, I. (1991) *J. Am. Chem. Soc. 113*, 9349–9352.
Varani, G., Wimberly, B., & Tinoco, I. (1989) *Biochemistry 28*, 7760–7772.
Varani, G., Cheong, C., & Tinoco, I. (1991) *Biochemistry 30*, 3280–3289.
Wang, A. C., Kim, S. G., Flynn, P. F., Chou, S.-H., Orban, J., & Reid, B. R. (1992) *Biochemistry 31*, 3940–3946.
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.
Wilson, K. S., Appelt, K., Badger, J., Tanaka, I., & White, S. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7251–7255.