Structure of the Autoregulatory Pseudoknot within the Gene *32* Messenger RNA of Bacteriophages T2 and T6: A Model for a Possible Family of Structurally Related RNA Pseudoknots†

Zhihua Du,‡ David P. Giedroc,§ and David W. Hoffman*,‡

Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, and Department of Biochemistry and Biophysics, Center for Macromolecular Design, Institute of Biosciences and Technology, Texas A&M University, College Station, Texas 77843-2128

Received November 16, 1995; Revised Manuscript Received January 16, 1996⊗

ABSTRACT: A 36-nucleotide RNA with a sequence corresponding to the 5′ end region of the gene *32* mRNA of bacteriophages T2 and T6 was analyzed by one- and two-dimensional NMR methods. NMR results provide clear evidence that the RNA is folded into a pseudoknot structure with two coaxial stems connected by two loops, in a classic pseudoknot topology. The pseudoknot is unusual in that one of the loops consists of only one nucleotide, which spans the major groove of a seven base pair helical stem. Imino proton resonances indicate the hydrogen bonding pattern within the pseudoknot, and two-dimensional NOE spectra provide information that describes many of the structural features. The temperature dependence of the UV absorption and imino proton exchange rates provides insight into the stability of the pseudoknot. A three-dimensional model of the pseudoknot that is consistent with our NMR data is presented, and features that may be important for stabilizing the pseudoknot structure are discussed. A substantial number of other putative RNA pseudoknots described in the literature have sequences and topologies that appear to be related to the T2 and T6 pseudoknots. We propose that these RNAs may be members of a family of pseudoknots related by a similar structural motif, which we refer to as "common pseudoknot motif 1" or CPK1. The bacteriophage T2/T6 pseudoknot can be considered a structural model for the CPK1 family. The common features of the CPK1 pseudoknots are a stem 2 with six or seven base pairs, a loop 1 consisting of a single adenosine, and a variable length stem 1 and loop 2. The first "dangling" nucleotide at the 3′ end of the molecule probably stabilizes stem 2. The CPK1 family includes several of the retroviral pseudoknots associated with mRNA frameshifting and readthrough. The work presented here describes the first detailed NMR analysis of an RNA pseudoknot with an entirely natural nucleotide sequence.

RNA pseudoknot structures play an essential role in a variety of biological processes. Pseudoknots were first identified as a component of plant virus RNA that is recognized by tRNA processing enzymes (Pleij et al., 1982; van Belkam et al., 1985). More recently it has been shown that pseudoknots, often located within messenger RNAs, are an essential part of regulatory mechanisms for translational efficiency, frameshifting, and readthrough, among other functions. For example, a pseudoknot structure has been implicated in the regulation of frameshifting and readthrough at the junction of the gag-pol fusion protein in many of the retroviruses (Jacks et al., 1988a,b; Wills et al., 1994; Chen et al., 1995). Pseudoknots may direct the synthesis of selenocysteine-containing proteins (Taylor et al., 1994), and a pseudoknot forms a part of the structure of 16S ribosomal RNA (Moazad & Noller, 1987). An RNA pseudoknot has been identified within the autoregulatory region of the gene *32* mRNA[1] of bacteriophages T2, T4, and T6 (McPheeters et al., 1988).

The protein product of gene *32* of bacteriophages T2, T4, and T6 is a single-stranded nucleic acid binding protein that binds preferentially to single-stranded DNA and with lower affinity to RNA single strands (Newport et al., 1981). Multiple monomers of the gene 32 protein bind cooperatively to single-stranded nucleic acid with little natural sequence selectivity. An important function of gene 32 protein is to protect single-stranded DNA that is transiently produced during DNA replication, recombination, and repair processes. The intracellular concentration of the product of gene 32 is believed to be tightly regulated by exploiting the differences in affinities for its various nucleic acid targets (McPheeters et al., 1988). The highest affinity targets of gene 32 protein are DNA, and these targets are saturated at relatively low protein concentrations. At higher concentrations gene 32

† This work was supported by NSF Grants MCB-9406065 and STI-9413770 (to D.W.H.) and NIH Grant GM42569 (to D.P.G.) and is in partial fulfillment of the requirements for the Ph.D. degree at the University of Texas at Austin (to Z.D.).

* Corresponding author. Telephone: (512) 471-7859. Fax: (512) 471-8696.

‡ University of Texas at Austin.

§ Texas A&M University.

⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1996.

[1] Abbreviations: BaEV, baboon endogenous virus; CPK1, common pseudoknot motif 1; FeLV, feline leukemia virus; GaLV, gibbon ape leukemia virus; HIV-1, human immunodeficiency virus type 1; HMQC, heteronuclear multiple-quantum coherence spectroscopy; HTLV-1, human T-cell leukemia virus type 1; mRNA, messenger RNA; MuLV, murine leukemia virus; MMTV, mouse mammary tumor virus; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NTPs, nucleoside triphosphates; SRV-1, simian retrovirus type 1; SSB, single-stranded binding protein; 2QF-COSY, two-quantum-filtered correlated spectroscopy; TOCSY, total correlated spectroscopy.

0006-2960/96/0435-4187$12.00/0

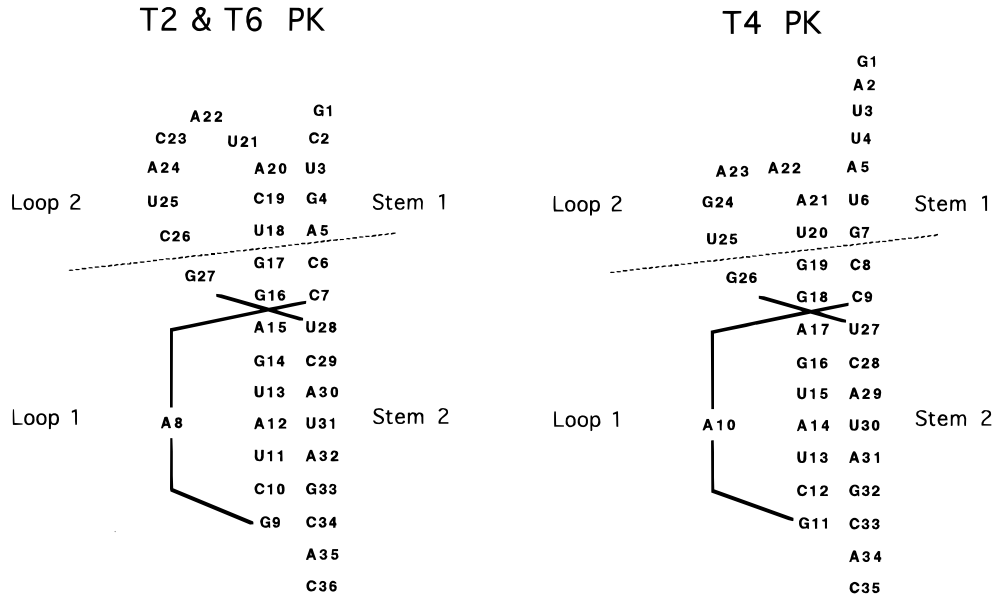

FIGURE 1: Schematic diagrams of the nucleotide sequences of the fragments of the bacteriophage gene *32* mRNAs used in this work. The diagrams are drawn in such a way as to show the base pairing that occurs within the RNAs. The pseudoknots from bacteriophages T2 and T6, shown on the left, have the same sequence. The pseudoknot from the bacteriophage T4 gene *32* mRNA is shown on the right. The pseudoknots have the same sequence in the region of the figures below the dashed line.

protein binds to its own mRNA, thus interfering with translation, and in this way the production of additional gene 32 product is inhibited. The 5′ end of the gene *32* mRNA contains a nucleotide sequence that suggests a pseudoknot structure (McPheeters et al., 1988). This pseudoknot is believed to act as a nucleation site for binding the gene 32 SSB to the mRNA. The pseudoknot within the gene *32* mRNA therefore appears to play a direct role in the regulation of its own gene product.

The nucleotide sequences of the bacteriophage T2 and T6 gene *32* mRNA pseudoknots are identical. The sequence of the bacteriophage T4 pseudoknot is identical to those of T2/T6 in the regions below the dashed line in Figure 1 but differs in stem 1 and loop 2. In the preceding paper, Qiu et al. (1996) used a comparison of imino proton spectra to show that the T4 and T2/T6 RNA pseudoknots have substantially similar structures. Although the autoregulatory process has only been studied in bacteriophage T4, the T2, T4, and T6 gene 32 proteins are 98% identical (McPheeters et al., 1988). This suggests that the mechanism of gene *32* autoregulation is similar in all three bacteriophages. Therefore, investigations of the T2/T6 pseudoknot can likely lead to an understanding of the phage T4 system.

Despite the wide-ranging importance of RNA pseudoknots, there is a relatively small amount of detailed structural information available at this time. Pleij and co-workers (Rietveld et al., 1983, 1984; Pleij et al., 1985) proposed a model that describes the general features of pseudoknot structure [reviewed by Pleij and Bosch (1989) and Puglisi and Tinoco (1989)]. In the idealized model, the pseudoknot consists of two coaxially stacked helical stems, with two loops connecting the stems, as illustrated in Figure 1. Loop 1 crosses the deep major groove of stem 2, and loop 2 crosses the shallower minor groove of stem 1. Puglisi et al. (1990a) used NMR methods to confirm the general features of the model of Pleij and co-workers, and the effects of loop size on pseudoknot stability have also been investigated (Wyatt et al., 1990). More recently, Shen and Tinoco (1995) used NMR methods to determine the structure of an RNA pseudoknot that is similar in sequence to the pseudoknot at the frameshift site in mouse mammary tumor virus mRNA. In the work reported here, the structure and stability of the pseudoknot in the bacteriophage T2 and T6 gene *32* autoregulatory mRNA are investigated. This study is part of our ongoing investigations into the principals that govern RNA pseudoknot structure and function relationships.

MATERIALS AND METHODS

RNA Synthesis and Purification. RNA molecules with the sequences shown in Figure 1 were transcribed using T7 RNA polymerase and synthetic DNA templates, essentially as described by Milligan et al. (1987). The DNA templates consisted of a double-stranded 18 base pair T7 promoter sequence and a single-stranded coding sequence. T7 RNA polymerase was isolated from an overproducing strain of *Escherichia coli* (pAR1219) provided by John Dunn. Preparative-scale transcription reactions typically contained 4 mM NTPs, 10 mM $MgCl_2$, and 300 nM template DNA. Transcriptions were incubated for 4 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 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 electro-eluter. The RNA was further purified by repeated ethanol precipitation and 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/15 mL transcription volume.

NMR Experiments. NMR spectra of the bacteriophage T2/T6 RNA pseudoknot were collected at 500 MHz on a two-channel Bruker AMX spectrometer equipped with inverse and proton-dedicated probes. Spectra in D_2O solvent were obtained using samples with 5−16 mg of RNA dissolved in 550 μL of 10 mM Na/K phosphate buffer at pH 6.5. Presaturation was used for solvent suppression for samples in D_2O. Proton spectra in 90% H_2O/10% D_2O solvent were

obtained using the jump and return method (Guéron et al., 1991). 2D nuclear Overhauser effect (NOE) spectra were acquired in both D_2O and H_2O solvents using a range of mixing times between 100 and 300 ms. 2QF-COSY spectra in D_2O solvent were acquired by the method of Müller et al. (1986). A TOCSY spectra with a mixing time of 80 ms was acquired by the method of Bax and Davis (1985). All 2D NMR were acquired in the phase-sensitive mode using the method of time-proportional phase incrementation (Marion & Wüthrich, 1983). 2D spectra were typically acquired with 750 blocks of either 1024 or 2048 complex points, using sweep widths of 4500−5000 Hz for spectra in D_2O and 12 195 Hz for spectra in H_2O. Two-dimensional data sets were acquired at 1 and 20 °C, resulting in relatively small temperature-dependent shifts in proton resonance frequencies; these shifts were useful in identifying overlapping resonances. A $^{13}C-^{1}H$ correlated heteronuclear multiple-quantum coherence (HMQC) spectrum of the pseudoknot was acquired at 500 MHz proton frequency using the pulse sequence of Bax et al. (1983). A total of 200 blocks of 512 complex points were acquired with 128 scans per block, for a total acquisition time of 18 h. A sweep width of 5000 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 20 °C with a 20 s recycle time between scans, which exceeded four times the T_1 of the slowest relaxing proton in the RNA. A total of 16 scans were acquired for each of 14 inversion−recovery delays between 0 and 5.0 s.

Thermodynamic Studies. The thermal denaturation of the T2/T6 pseudoknot RNA was monitored at 260 nm. Melts were collected on an exhaustively dialyzed RNA in 10 mM sodium phosphate and 50 mM NaCl, pH 6.90, at 4 μM RNA strand exactly as described in the accompanying paper (Qiu et al., 1996). These conditions were chosen to be consistent with our thermodynamic analysis of bacteriophage T4 pseudoknots (Qiu et al., 1996) and differ slightly from the conditions where most of the NMR data were collected. A one-dimensional NMR spectrum collected at the salt and pH conditions of the UV melting experiments appears identical to our other NMR data. The single UV melt shown was subjected to two methods of analysis as described by Laing and Draper (1994) to obtain thermodynamic parameters exactly as described in the accompanying paper (Qiu et al., 1996).

RESULTS

NMR Analysis. NMR methods were used to investigate the 36-nucleotide RNA with a sequence corresponding to the pseudoknot located near the 5′ end of the gene *32* mRNA of bacteriophages T2 and T6. In addition, we also carried out a preliminary NMR investigation of the analogous bacteriophage T4 pseudoknot, the results of which are reported in the preceding paper (Qiu et al., 1996). Although both the T2/T6 and T4 pseudoknots yield high quality NMR spectra with sharp and disperse resonance lines, in our hands the T2/T6 pseudoknot was significantly and consistently more resistant to degradation over the periods of time required for NMR data collection. The T2/T6 RNA was therefore the preferred molecule for our more intensive NMR studies. A schematic diagram of the T2/T6 pseudoknot sequence and topology is shown in Figure 1, and the sequence of the T4 pseudoknot is shown for comparison. The "dangling" nucleotides at the 5′ and 3′ ends of the T2/T6 pseudoknot were included in the molecule used in the detailed NMR analysis. A T2/T6 pseudoknot RNA prepared without nucleotides A35 and C36 was destabilized, as judged by its imino proton resonances. This is consistent with our accompanying work indicating that at least one of these nucleotides is required for stabilizing the pseudoknot structure (Qiu et al., 1996).

Table 1: Assignments of Chemical Shifts (ppm) for the 36 Nucleotide RNA Pseudoknot of the Gene *32* mRNA of Bacteriophages T2 and T6[a]

	H8/H6	H5/H2	C2	H1′	H2′	H3′	*J*(1′−2′)	imino/amino
G1	8.62	na	na	6.26	5.55	5.02	S	
C2	7.64	5.67	na	5.79	4.13	4.21	S	
U3	7.86	5.71	na	5.95	4.46	4.72	S	
G4	8.16	na	na	5.48	4.24	4.47	W	12.13
A5	7.88	7.87	151.62	6.08	4.60		W	
C6	7.49	5.23	na	5.52	4.15	4.53	W	7.00/8.53
C7	7.71	5.49	na	5.50	4.38		W	6.79/8.55
A8	8.22	7.56	151.93	5.78	4.76	4.54	S	
G9	8.09	na	na	5.89	4.80		W	12.61
C10	7.97	5.33	na	5.57	4.60	4.34	W	6.93/8.52
U11	7.86	5.46	na	5.54	4.46	4.73	W	13.20
A12	8.05	6.97	150.32	5.99	4.25		W	
U13	7.49	5.03	na	5.34	4.45	4.33	W	13.33
G14	7.51	na	na	5.68	4.58	4.45	W	11.66
A15	7.55	7.63	150.90	5.77	4.48	4.45	W	
G16	7.19	na	na	5.35	4.44	4.60	W	12.90
G17	7.27	na	na	5.79	4.49	4.43	W	13.38
U18	7.91	5.10	na	5.71	4.60		W	14.44
C19	7.71	5.71	na	5.70	4.46	4.55	W	7.00/8.38
A20	8.24	8.21	153.17	6.04	4.64	4.66	S	
U21	7.62	5.51	na	5.75	4.24	4.41	S	
A22	7.78	8.16	153.24	5.80	4.94	5.20	S	
C23	7.69	6.01	na	5.77	4.17	4.53	S	
A24	8.37	8.12	153.24	6.02	4.92	4.81	S	
U25	7.78	5.72	na	5.87	4.33	4.60	S	
C26	8.00	5.87	na	5.77	4.00		S	
G27	7.49	na	na	5.88	4.40	4.50	S	
U28	7.66	5.94	na	5.92	4.49	4.14	W	14.16
C29	7.94	5.83	na	5.62	4.58		W	7.03/8.14
A30	8.14	7.26	150.77	5.92	4.48	4.34	W	
U31	7.68	5.10	na	5.46	4.21	4.60	W	13.23
A32	7.98	6.83	149.86	5.95	4.52	4.72	W	
G33	7.12	na	na	5.49	4.34	4.28	W	13.49
C34	7.16	5.15	na	5.32	4.49		W	6.81/8.05
A35	7.91	7.32	151.62	5.99	4.32	4.09	W	
C36	7.45	5.37	na	5.47	3.88	4.07	S	

[a] Data for protons are reported at 20 °C, with chemical shifts relative to the residual HDO resonance at 4.84 ppm. Carbon chemical shifts are relative to external TMS. S and W refer to strong and weak H1′ to H2′ cross-peaks in the 2QF-COSY spectrum, respectively, where W means that the cross-peak is either absent or very weak. na = not applicable.

In any NMR investigation of macromolecular structure, detailed structural information can only be derived after multidimensional methods have been used to assign a substantial fraction of the resonances to the specific protons from which they arise. Here we report assignments for a large fraction of the resonances in the T2/T6 pseudoknot, with at least partial assignments for all 36 nucleotides (Table 1). As an aid to obtaining the maximum number of resonance assignments for the T2/T6 RNA pseudoknot, 2D NMR data in both D_2O and H_2O solvents were collected at both 1 and 20 °C. The chemical shifts of many of the protons were found to exhibit small changes with temperature, typically on the order of 0.01−0.05 ppm over the 19 °C temperature range. These shifts were sufficient in many cases to separate peaks that overlap at a particular temper-

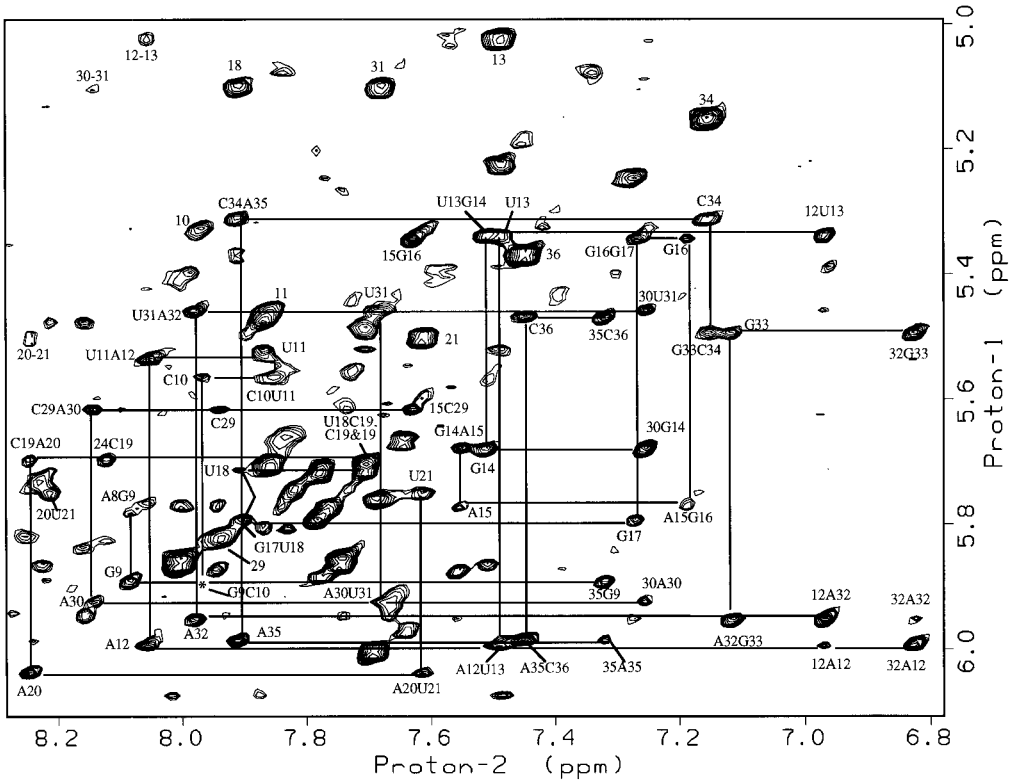

FIGURE 2: Section of a 2D NOE spectrum of the bacteriophage T2/T6 RNA pseudoknot in D_2O solvent and 10 mM phosphate buffer, pH 6, at 20 °C. The spectrum shown was obtained with a mixing time of 280 ms. Essentially all of the visible peaks in this region of the spectrum have been assigned, and some of the assigned peaks are labeled as follows. Intranucleotide H6/H8−H1′ NOEs are labeled by type and number, for example, A12. Sequential H1′(*i*)−H6/H8(*i*+1) NOEs are labeled by the type and number of the two nucleotides, for example, C34A35. Sequential H6/H8(*i*)−H5(*i*+1) NOEs are labeled by the numbers of the two nucleotides separated by a dash, for example, 20−21. NOEs from adenosine H2 to ribose H1′ are labeled by the number of the adenosine and the type and number of the ribose proton, for example, 32G33. Pyrimidine intranucleotide H5−H6 NOEs are labeled by the number of the nucleotide, for example, 31. The sequential NOE peak labeled G9C10 is marked with an asterisk since it is observable at a contour level slightly lower than shown.

ature. Given the small size of the shifts, they are unlikely to be associated with a significant structural change in the RNA and are more likely a consequence of the hypersensitivity of chemical shifts to very small conformational changes. The hypothesis that there is little structural change in the RNA over the investigated temperature range is supported by the patterns of intensities within the 2D NOE spectra, which are quite similar at the two temperatures investigated. These temperatures are each significantly below the melting point of the RNA as indicated by UV absorbance changes and imino proton spectra.

The structure of the T2/T6 RNA pseudoknot can be divided into six distinct regions, and these will be described in turn. These regions are the A-form helical stems (nucleotides 3−6, 9−14, 17−20, and 29−34), the junction of the two stems (nucleotides 7, 15, 16, and 28), loop 1 (nucleotide 8), loop 2 (nucleotides 21−27), the 5′ end of the molecule (nucleotides 1−2), and the 3′ end of the molecule (nucleotides 35−36). Nucleotides are numbered as in Figure 1.

(i) A-Form Helical Stems (Nucleotides 3−6, 9−14, 17−20, and 29−34). Resonance assignments were straightforward for the regions of the molecule that are most similar to ideal A-form helices, and these resonances were assigned first. Convenient starting points for making spectral assignments were provided by the pyrimidine H5−H6 correlations observed in the 2QF-COSY spectrum. It was then possible to use the 2D NOE spectrum to "walk" from one nucleotide to the next via short sequential and intranucleotide NOEs between the H6 or H8 protons on the bases and the ribose H1′, H2′, and H3′ protons. Examples of two such sequential walks are shown in Figure 2. Assignments of the adenosine H2 protons were confirmed by the chemical shifts of their associated carbons, near 151 ppm, identified in a $^1H-^{13}C$ HMQC spectrum (Varani & Tinoco, 1991). The adenosine H2 protons of A5, A12, A15, A20, A30, A32, and A35 exhibit relatively long T_1 relaxation times, in the range of 3−4 s, as is typical for these protons when they are located within A-form helices (Hoffman et al., 1993). Most of the nonexchangeable protons in the RNA molecule have T_1 values on the order of 1.0 s.

NOEs involving the imino and amino protons were particularly valuable in establishing the resonance assignments in the A-form helical part of the RNA molecule (Figure 3). These resonances were assigned using 2D NOE spectra collected in 90% H_2O/10% D_2O solvent at 1 and 20 °C. NOEs from the stem imino protons to the H1′ resonances of the nucleotides in the *i*+1 position relative to each member of the base pair were also observed (Figure 3A). These imino to H1′ NOES are consistent with the A-form helical structure of the stem (Heus & Pardi, 1991a) and confirmed our assignments of the stem H1′ resonances. NOEs were observed between the imino protons of sequential base pairs (Figure 3B). The imino resonance for G9 in the closing base pair of stem 2 is clearly observed and appears to be significantly protected from exchange with the solvent, even at relatively high temperatures. A schematic diagram summarizing many of the observed internucleotide NOEs

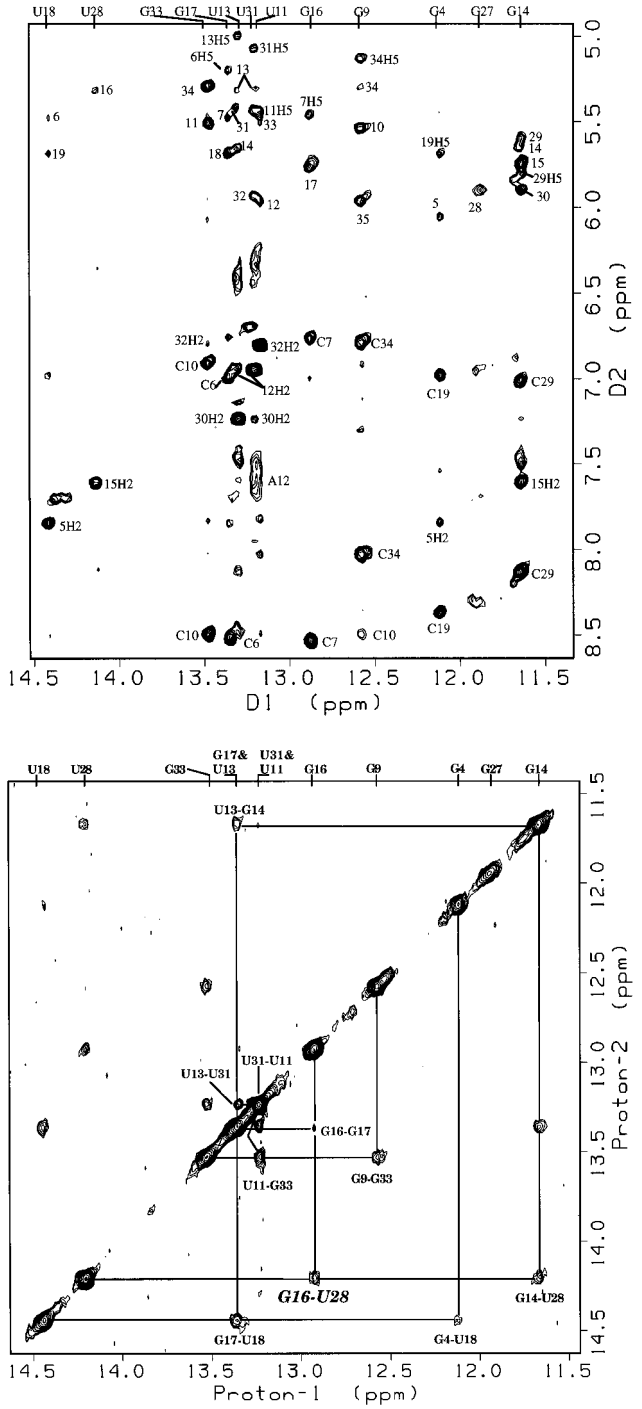

FIGURE 3: (A, top) Section of the 2D NOE spectrum (280 ms mixing time) of the bacteriophage T2/T6 RNA pseudoknot in 90% H_2O/10% D_2O solvent and 10 mM phosphate buffer, pH 6, at 1 °C. The solvent resonance was suppressed by the jump and return method. Chemical shift positions of the imino protons are labeled on the top of the window. Imino to ribose H1′ NOEs are labeled by the number of the nucleotide to which the H1′ belongs; for example, 19 refers to the U18 imino to C19 H1′ NOE. Imino to pyrimidine H5 or adenosine H2 NOEs are labeled by the number of the nucleotide followed by H2 or H5; for example, 5H2 refers to the U18 imino to A5H2 NOE. Imino to amino NOEs are labeled by the type and number of the nucleotide; for example, C19 refers to the G4 imino to C19 amino NOEs. (B, bottom) Section of the 2D NOE spectrum, showing NOEs between imino protons. The assigned cross-peaks are labeled. Of particular interest is the NOE between the imino protons of G16 and U28, across the junction between the two coaxial stems, labeled with a larger font.

for the bacteriophage T2/T6 pseudoknot is displayed in Figure 4.

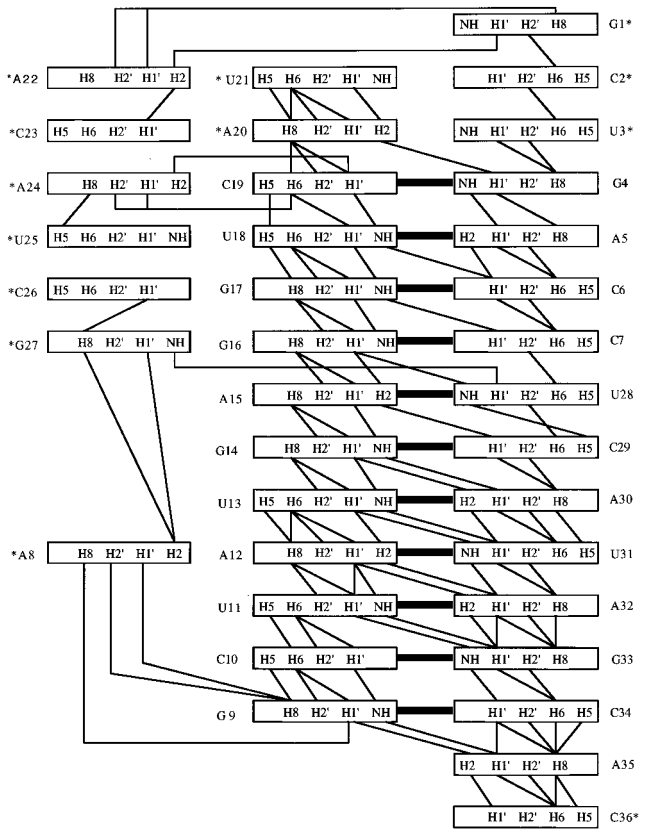

FIGURE 4: Schematic diagram summarizing the observed internucleotide NOEs and hydrogen bonding pattern within the helical stems of the bacteriophage T2/T6 pseudoknot. Assigned internucleotide NOEs are indicated by thin lines; intranucleotide NOEs are not indicated. Nucleotides with riboses predominantly in the C2′ *endo* conformation are indicated by asterisks. Base pairs with Watson−Crick-type hydrogen bonds, confirmed by assigned imino proton resonances, are indicated by thick bars. Additional NOEs are certainly present but remain unassigned at this time due to resonance overlap.

H1′ to H2′ coupling constants are sensitive indicators of ribose conformation. Small coupling constants indicate riboses that are primarily in the C3′ *endo* conformation (Puglisi et al., 1990b). For nucleotides in the helical stems, weak or absent H1′−H2′ correlations in the 2QF-COSY and TOCSY spectra were assumed to indicate small H1′−H2′ coupling constants and therefore C3′ *endo* ribose conformations. Using this assumption, all of the stem nucleotides with the exception of the A20-U3 base pair at the top of stem 1 are C3′ *endo*, as is normally observed in A-form helical RNA. Larger H1′−H2′ couplings for U3 and A20 indicate either C2′ *endo* or mixed conformations, probably due to the influence of the adjacent nucleotides U21 and C2, which are not part of the regular A-helical structure.

Although in most respects the stem nucleotides appear to be ideal A-helical, there are some minor irregularities. For example, the imino resonance for the uridine in the A20-U3 base pair is not observed. Since the NOE intensities involving the nonexchangeable protons for this base pair are typical of A-helical RNA, we think it is most likely that the A20-U3 base pair is actually present. Presumably the imino proton is exchanging with the solvent too rapidly to be observed in the NMR spectra. This rapid exchange is not unusual for imino protons in base pairs at the ends of RNA helices (Hoffman et al., 1993). A slightly unusual feature is observed for the imino resonance of G9. This resonance

is split into a doublet (Figure 6), possibly indicating that an alternative conformation is present, slightly differing from the major form in the vicinity of the G9-C34 base pair. It is interesting that, in the T4 RNA pseudoknot, the G11 resonance (corresponding to G9 in the T2/T6 sequence) is also a doublet (Qiu et al., 1996).

(ii) Junction of the Helical Stems (Nucleotides 7, 15, 16, and 28). Assignment of the resonances of the A15-U28 and G16-C7 base pairs was made using sequential NOEs from the G14-C29 and C6-G17 base pairs. The most striking feature of the junction of the two stems is how little it differs from an ordinary A-form helix. Of particular interest are the NOEs involving the imino protons (Figure 3). It is possible to follow the sequential imino-to-imino NOEs in an unbroken walk through the junction of the two stems (Figure 3B). The sequential NOE peak between the imino protons in the A15-U28 and G16-C7 base pairs is clearly observed in Figure 3. Sequential NOEs involving the nonexchangeable protons also indicate the stacking of base pairs through the junction of the two stems (Figure 4). For example, NOEs to the H1′ proton of G16 from the imino of U28 and the H2 proton of A15 are an indication that the A15-U28 and G16-C7 do not deviate far from the conformations expected in an ordinary A-form helix. An NOE is observed between H2′ of C7 and H6 of U28, as would be expected if C7 and U28 were sequential nucleotides in an A-form helix. There is, however, some evidence of distortion of the A-helical stems at the junction, indicated by sequential NOE intensities that differ somewhat from the rest of the helix. For example, the G16 imino to G17 imino NOE, though present, is weaker than the other imino-to-imino NOEs in the A-helical stems (Figure 3), and some of the usual A-helix type NOEs are not observed for C7. Ribose conformations for all four of the nucleotides in the junction appear to be C3′ *endo*, as is usual for A-helical RNA. We conclude that the two stems are stacked coaxially, with only minor distortion from ideal A-helical geometry at the junction. This is consistent with the result reported by Puglisi et al. (1990a,b) for their designed RNA pseudoknot.

(iii) Nucleotide A8 of Loop 1. The C7-G16 and G9-C34 base pairs are quite clearly defined by our NMR data. This requires that the single nucleotide A8 must span the seven base pair length of stem 2 (Figure 1). At first glance it would seem that it would not be possible to accomplish this without significant distortion of the helix in stem 2. However, the natural twist of the A-form helix is such that the 3′ end of C7 and the 5′ end of G9 are placed relatively close together. With ideal A-helical stems, the distance from O3′ of C7 to the phosphorus atom of G9 is 11 Å. In our model-building studies we find that the gap between C7 and G9 can be spanned by the single A8 nucleotide if the phosphate backbone is slightly moved to decrease the distance from the O3′ of C7 to the phosphorus of G9 to about 8.8 Å. This moving of the phosphate backbone can easily be accomplished without disrupting the C7-G16 and G9-C34 base pairs. A stem 2 with a length of either six or seven nucleotides is optimum for having loop 1 consist of only one nucleotide. A similar conclusion was reached by McPheeters et al. (1988) on the basis of their model-building studies for the T4 gene *32* mRNA pseudoknot. NOEs are observed between the H8 proton of G9 and the ribose protons of A8 and from the H8 proton of A8 to the ribose of G9, as well as from the H2 proton of A8 and H1′ and H8 of G27 (Figure 4). These observed NOEs are consistent with the base of A8 being in the major groove of stem 2 and can be accommodated by our structural model. The ribose of A8 is in the C2′ *endo* conformation.

In a study of loop size requirements for pseudoknots, Wyatt et al. (1990) concluded that a minimum of three nucleotides were required for loop 1 and four nucleotides were required for loop 2 in their particular pseudoknot. This is *not* inconsistent with our results, since the lengths of stems 1 and 2 in Wyatt's study were three and five nucleotides, respectively. Clearly, loop length requirements depend on the lengths of the stems. Although at first glance it seems counterintuitive, in cases where stem 2 contains six or seven base pairs it can spanned by a relatively short loop 1, while a stem 2 consisting of four or five base pairs requires a longer loop 1.

(iv) Nucleotides of Loop 2 (Nucleotides 21−27). Assigning the resonances of loop 2 was significantly more difficult than for other parts of the T2/T6 RNA pseudoknot, and these assignments should be viewed as being somewhat less certain. However, any assignment errors for resonances of loop 2, if they are present, would have only a minimal effect on the structural model of the pseudoknot presented in this work, since the pseudoknot model is primarily determined by the constraints imposed by the coaxial helical stems where assignments are unambiguous. Loop 2 assignments were made using NOE information, combined with rounds of model building to test the validity of the assignments. Since we have established that loop 2 must span the distance from A20 to U28, located at opposite ends of stem 1, real and substantial constraints are imposed on our options for model building and assignments.

The first loop 2 nucleotide, U21, is stacked on A20. This is indicated by A-helix-like NOEs from A20 H8 to U21 H5, A20 H1′ and H2′ to U21 H6, A20 H2 to U21 H1′, and A20 H8 to U21 H6. In addition to NOEs between sequential nucleotides, some longer range NOEs are present. NOEs between A22 and G1 indicate that these nucleotides are relatively close together (Figure 4). Of particular interest are NOEs between the H2 proton on the base of A24 and the H1′ proton of C19 and NOEs from H6 of C19 to H1′ and H2′ of A24. These NOEs are consistent with A24 being in the minor groove of stem 1. The NOE between H2 of A24 and H1′ of C19 is similar to an NOE observed between a single-stranded adenosine H2 proton and an H1′ proton in the minor groove side of a helix in a domain of the *Tetrahymena* group I intron previously studied by NMR (Chastain & Tinoco, 1992). Of the loop 2 nucleotides, only G27 has an observed imino proton. This suggests that G27 is involved in a hydrogen bond, although we have not yet been able to identify its hydrogen bonding partner. The G27 imino proton does have an NOE to the H1′ proton of U28, which was useful in determining the approximate position of G27 in our structural model. The other imino protons of loop 2, belonging to U21 and U25, we assume are exchanging too fast with the solvent to be observed. The riboses of A22 through G27 are in the C2′ *endo* or mixed conformation. Although our NOE information is relatively sparse and less certain for loop 2 than for other parts of the pseudoknot, our assigned data are consistent with the following model: Nucleotides C23 through C26 are stacked with the bases oriented toward the minor groove of stem 1; and the base of G27 is located quite close to the bases of G16 and U28 at

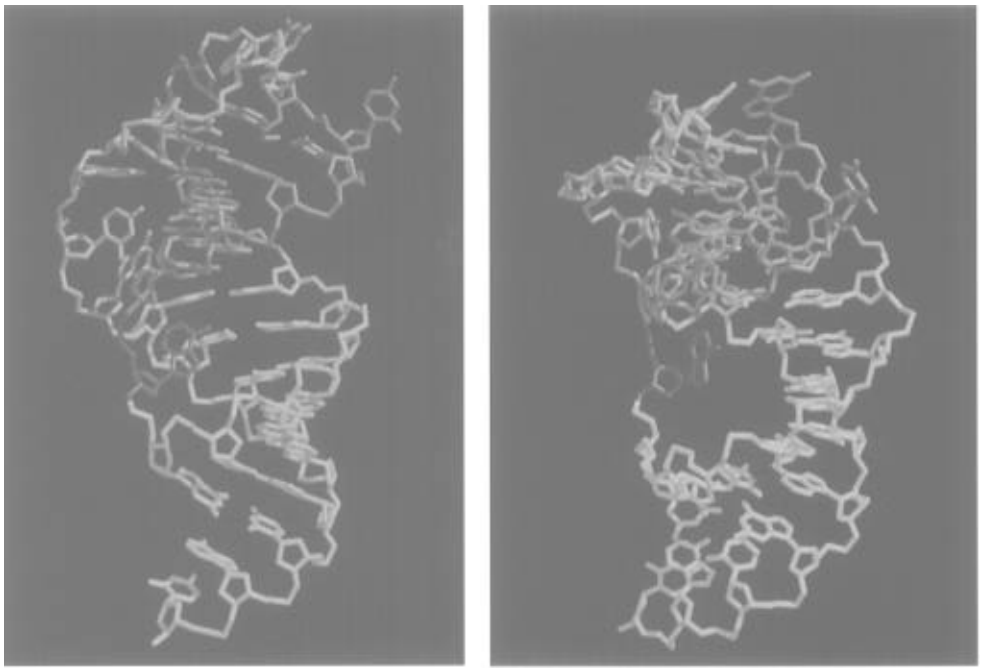

FIGURE 5: (A, left) A view of the T2/T6 gene *32* mRNA pseudoknot, with stem 1 colored green (nucleotides 1−7 and 16−20), stem 2 colored blue (nucleotides 9−15 and 28−36), loop 1 colored red (nucleotide A8), and loop 2 colored purple (nucleotides 21−27). The purine ring of G1 is at the top of the figure, C2 is exposed at the upper right corner of the figure, and C36 is at the bottom of the figure. (B, right) A slightly different view of the pseudoknot, tilted approximately 30° about the horizontal axis to show the deep major groove of stem 2 and the shallower minor groove of stem 1. Nucleotides are color coded as in (A). Nucleotide A8 can be seen spanning the major groove, bridging the relatively short distance between C7 and G9. Loop 2 spans the minor groove of stem 1.

the junction of the two stems; and U21 and A22 form a turn linking stem 1 to loop 2. Analogous to the case with loop 1, there are possible opportunities for interactions between the bases of loop 2 and the nucleotides of stem 1.

(v) 5′ End Nucleotides (G1 and C2). Resonance assignments for U2 and G1 were made using sequential NOEs, starting from nucleotide U3 in the helix of stem 1. The sequential NOEs from G1 to C2 and C2 to U3 are weaker than would be expected for A-helical RNA. This may be a result of disorder at the 5′ end of the molecule, or perhaps C2 is bulged and not directly stacked on U3. A G1-U21 base pair cannot be ruled out by our NMR data. The riboses of G1 and C2 are either C2′ endo or mixed conformations. Somewhat unusual chemical shifts are observed for the ribose of G1 (Table 1), possibly a result of an unusual conformation of C2. No imino proton resonance was observed for G1; apparently it exchanges quickly with the solvent.

(vi) 3′ End Nucleotides (A35 and C36). Near A-helical base stacking continues from stem 2 through A35 and C36, even though these nucleotides are not base paired. This is not unusual for unpaired nucleotides at the 3′ end of an A-helix [see, for example, Colvin et al. (1993)]. Unpaired bases can be stabilized by a hydrogen bond between the 2′-hydroxyl proton of nucleotide *i* and the O4′ oxygen of nucleotide *i*+1 (Saenger, 1984), in addition to the stabilization derived from base-stacking interactions. NOEs from A35 H8 to the ribose protons of C34 and an NOE from the H2 base proton of A35 to the H1′ of G9 are observed, as is typical of A-helical RNA. The A35 H2 to G9 H1′ NOE is particularly valuable in further confirming the existence of the undistorted G9-C34 base pair at the end of stem 2. The near-ideal A-helical stacking of A35 below the G9-C34 base pair probably accounts for the protection of the G9 imino proton from fast exchange with the solvent. The sequential NOE intensities from C36 to A35 NOEs differ from those in the stems, and some unusual sequential NOEs are observed, such as C36 H5 to A35 H8. Apparently the conformation of the "dangling" C36 deviates somewhat from an ideal A-form helix. The ribose of C36 is in either C2′ *endo* or mixed conformations.

In the case of the bacteriophage T4 pseudoknot, it was found that omitting the two "dangling" nucleotides at the 3′ end of the molecule resulted in substantial changes in the melting behavior of the RNA (Qiu et al., 1996) and significantly destabilized the pseudoknot. It was for this reason that the dangling nucleotides were included in our T2/T6 pseudoknot RNA. It is clear that the dangling nucleotides play an essential role in stabilizing stem 2. A35 in particular should be considered to be an integral part of the minimal T2/T6 pseudoknot structure.

Three-Dimensional Model of the Pseudoknot Structure. The T2/T6 pseudoknot described in this work is one of the largest RNAs to be studied by NMR methods at this time. As expected for an RNA of this size, the problems encountered in interpreting the rather complex NMR spectra are formidable, and this ultimately limits the resolution of the structural model we obtain. However, the substantial amount of data that we have interpreted (specifically, assigned NOEs, ribose conformations, and hydrogen bonds identified by slow exchanging imino and amino protons), and in particular the well-defined coaxial helical stems, provide significant constraints on the possible structure of the pseudoknot. We have constructed a model of the T2/T6 RNA pseudoknot that is consistent with our assigned NMR data. Clearly, this model does not represent a finished high-resolution structure. The model is intended to give the reader a sense of the overall shape of the pseudoknot and to convey the approximate relative positions of the structural elements. In particular, the model illustrates how a single

adenosine nucleotide can span the gap crossed by loop 1 without disrupting stem 2.

The structural models shown in Figure 5 were constructed using the X-PLOR version 3.1 program suite (Brünger, 1993). An initial model was constructed as follows: Nucleotides 1−7, 9−20, and 28−36 were placed exactly as they would be in an ideal A-form helix (Arnott et al., 1972). A8 was placed in the major groove of stem 2 so that it spanned from C7 to G9 and was consistent with our observed NOEs from A8 to G9 and G27. U21 through G27 were placed along the minor groove of stem 1 so that they spanned from A20 to U28, while satisfying the constraints imposed by our observed A24-C19 NOEs. This initial model had a relatively high value for the X-PLOR energy function, primarily due to incorrect bond lengths and angles and van der Waals violations in the loops and at the junction of the two stems. The initial model was then subjected to simulated annealing with the following constraints: Interatomic distances between the atoms of nucleotides 1−7, 9−20, and 28−36 were restricted to be within 1 Å of ideal A-helical values; riboses of the 24 stem nucleotides with weak cross-peaks in the 2QF-COSY spectrum were restricted to the C3′ *endo* conformation; χ angles for the loop 2 bases were restricted to be within the range of 140−180° (*anti* conformation); hydrogen bond lengths within the base pairs were restricted to be within 0.2 Å of standard values (Saenger, 1984); experimentally derived NOEs (222 total) were classified as strong, medium, or weak and included as distance constraints corresponding to interproton distances of less than 4 Å, less than 5 Å, and less than 6 Å, respectively. The large majority of the NOE constraints were observed in an NOE spectrum obtained with a 100 ms mixing time; NOEs observed with a 200 ms mixing time were classified as weak. Views of a structure with a low value of the X-PLOR energy function, chemically reasonable values for bond lengths and angles, no significant van der Waals violations, and consistent with our assigned NMR data are shown in Figure 5. Again, we emphasize that this model is "consistent with our NMR data", rather than entirely determined by our NMR data.

The model of the T2/T6 pseudoknot clearly shows the overall shape of the molecule (Figure 5). Nucleotides of loops 1 and 2 are on the same side of the pseudoknot structure. The model confirms that nucleotide A8 can indeed span the gap from C7 to G9, while maintaining chemically reasonable bond lengths and angles, without distortion of the G9-C34 and C7-G16 base pairs and the A-helical geometry of stem 2. The phosphate groups of U28 and A8 are close together, consistent with the findings of Puglisi et al. (1990a).

Stability Studies. (i) Temperature Dependence of Exchangeable Proton Resonances. The temperature dependence of the imino proton resonances provides insight into the relative stabilities of the base pairs in various regions of the pseudoknot structure. As the temperature is increased, the exchange rate of the imino protons with those of the solvent increases, and the resonances broaden and disappear. The exchange rates of the imino protons are related to the frequency at which the base pairs open. This frequency increases as the melting point of the RNA is approached. The T2/T6 RNA pseudoknot appears to be quite stable as judged by the temperature dependence of their imino resonances. Spectra were recorded at 5° temperature intervals up to 55 °C and are shown at 10° intervals in Figure 6.

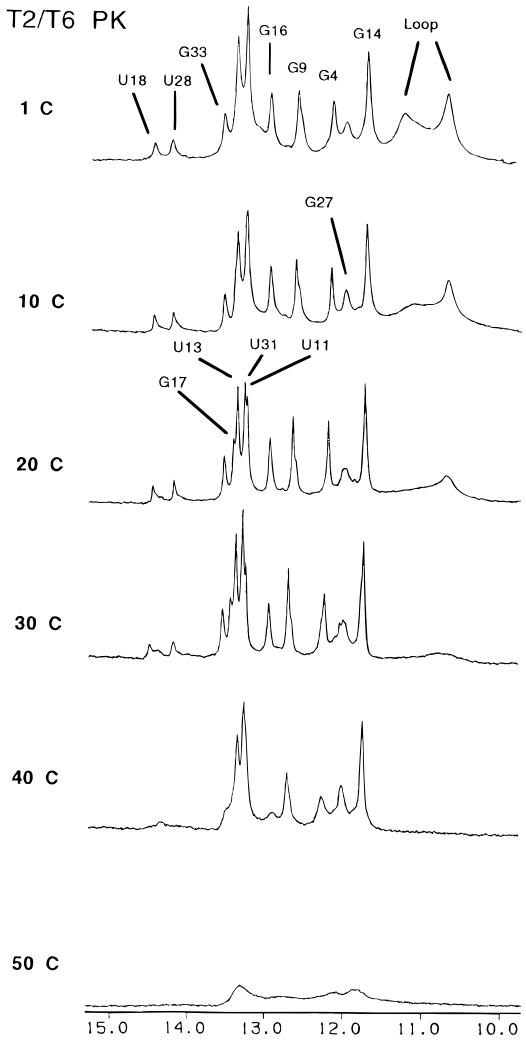

FIGURE 6: NMR spectra of the imino proton resonances of the 36 nucleotide T2/T6 gene *32* mRNA pseudoknot, recorded at temperatures between 1 and 50 °C in 90% H_2O/10% D_2O solvent and 10 mM phosphate buffer, pH 6. Resonance assignments are numbered as in Figure 1. The broad resonances designated Loop are most likely due to the remaining unassigned imino protons: G1, U3, U21, and/or U25.

The first resonances to disappear are the two very broad unassigned resonances with chemical shifts of less than 11.5 ppm. These resonances can be assigned by default to either G1, U3, U21, or U25, all of which are expected to be relatively solvent accessible. The remaining relatively sharp imino resonances are associated with the two stems and G27. These imino resonances disappear in two distinct groups. Imino resonances associated with four base pairs (A5-U18, C6-G17, C7-G16, and A15-U28) exhibit a marked decrease in intensity between 30 and 40 °C. Two of these base pairs are at the junction of the two stems, and two are in the central region of stem 1. The remaining imino resonances, assigned to five of the base pairs in stem 2, the G4-C19 base pair in stem 1, and G27 in loop 2, exhibit a sharp decrease in intensity between 40 and 50 °C. An additional imino resonance appears in the spectrum near the U18 resonance, most visible at 30 °C, before disappearing above 40 °C. This may be associated with a conformational change in the pseudoknot that occurs prior to melting. The rather complex temperature dependence for the T2/T6 imino resonances indicates multistep melting behavior. We conclude that the most stable base pairs of the T2/T6 RNA pseudoknot are in

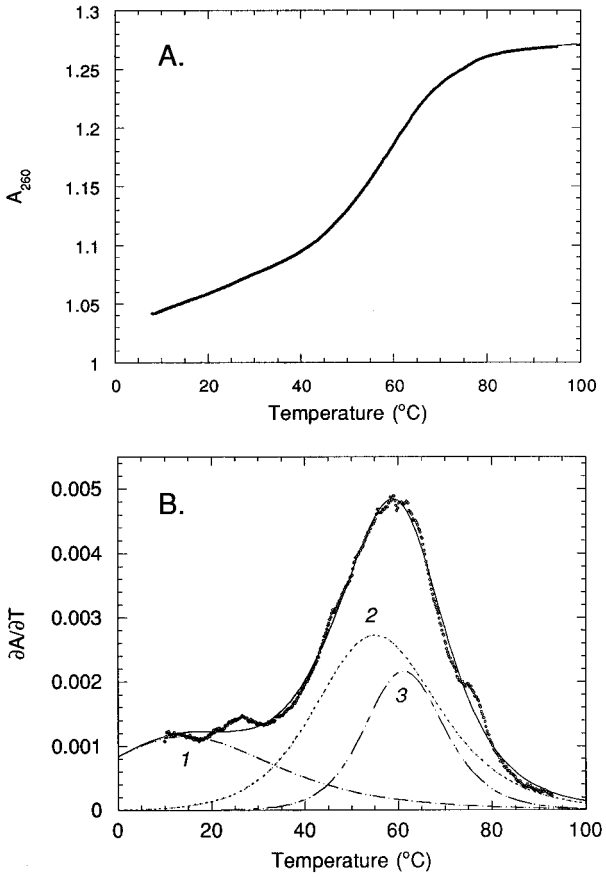

FIGURE 7: (A) Melting curve obtained for the 4.1 μM T2/T6 RNA pseudoknot in 50 mM NaCl, pH 6.90, at a scan rate of 0.2 °C/min. The thin line is a fit to the data with the sloping baseline method to give ΔH = 32 kcal mol^{-1} and t_m = 61.0 °C analyzed over a temperature range of 10−95 °C. (B) First derivative of the melting curve shown in panel A. The solid line through the experimental data is a fit to the three independent transitions indicated, as described in Materials and Methods. Each transition is defined by the following parameters: 1, t_m = 16.4 °C (fixed), ΔH = 13.0 kcal mol^{-1}, A = 0.058; 2, t_m = 56.0 °C, ΔH = 24.5 kcal mol^{-1}, A = 0.095; 3, t_m = 61.8 °C, ΔH = 39.1 kcal mol^{-1}, A = 0.049.

stem 2 and G4-C19 in stem 1. As temperature is increased, melting of the pseudoknot initiates in the region of the junction of the two stems. It is interesting that the G9-C34 base pair at the end of stem 2 is one of the most stable base pairs in the pseudoknot, as judged by imino proton exchange rates (Figure 6). This is consistent with our other evidence indicating that the closing base pair of stem 2 is stabilized by its stacking upon A35.

(ii) Thermodynamic Analysis of the Thermal Denaturation of the T2/T6 Pseudoknot. We used UV absorption spectroscopy to monitor the unfolding of the T2/T6 pseudoknot at 10 mM sodium phosphate and 50 mM NaCl, pH 6.90, as a function of temperature. A representative thermal melt obtained with the T2/T6 pseudoknot is shown in Figure 7. In panel A, a nonlinear least squares fit of these data by the sloping baseline method to a single two-state unimolecular transition gives ΔH of 32 kcal mol^{-1} and a t_m of 61.0 °C. While providing an acceptable fit to the data, this ΔH is far too small to be reasonable. For example, if one considers only the coaxial stacking of stems 1 and 2 to give a 12 base pair pseudocontinuous helix, the predicted enthalpy change in 1 M NaCl derived from nearest neighbor thermodynamics (Turner et al., 1988) is 113 kcal mol^{-1}. Therefore, the anomalously low enthalpy change suggests that the melting of the T2/T6 pseudoknot is not well described as a single, cooperative two-state transition, in apparent contrast to the T4 pseudoknot (Qiu et al., 1996). This is borne out by the quantitative analysis of the absorbance derivative data (Figure 7B). The derivative plot is characterized by significant nonzero $\partial A/\partial T$ at low temperature, precluding a good fit to a single melting transition. If the data are fitted by two independent transitions, the total enthalpy recovered is only marginally higher (~45 kcal mol^{-1}) than that obtained from the single transition analysis. However, when three independent transitions are used to fit the data, the goodness of fit improves modestly and the total ΔH approaches 90−100 kcal mol^{-1} (the fit is shown in Figure 7B). The t_m of the first melting transition cannot be adequately determined by the data; however, its enthalpy is fairly low (15−20 kcal mol^{-1}), depending on t_m. The major transition is fitted as a near superposition of two transitions, 2 and 3. Transition 2 has an enthalpy of 25 kcal mol^{-1} and a t_m of 56.0 °C, while transition 3 has ΔH = 39 kcal mol^{-1} and a t_m of 61.6 °C. If the temperature dependence of imino proton exchange is used as a guide to interpret the equilibrium unfolding enthalpies and t_m's of the three transitions in structural terms, then transitions 1−3 may correspond to the melting of the stem 1−stem 2 junction region, followed by most of stem 1 and finally by the five terminal base pairs in stem 2. Melting experiments carried out in the presence of 2 mM Mg^{2+} support the notion of three independent transitions in the T2/T6 pseudoknot with a total recovered enthalpy of 115 kcal mol^{-1} (data not shown).

DISCUSSION

Our NMR analysis provides strong evidence in support of the pseudoknot topology and base pairing schemes shown in Figure 1, with two coaxial A-form helical stems connected by two loops. Our NOE data show that the stems contain only minimal distortion from ideal A-form helical geometry. In this respect, the T2/T6 pseudoknot is similar to a pseudoknot previously studied by NMR methods (Puglisi et al., 1990a,b). The T2/T6 pseudoknot structure is novel in that it establishes that one of the connecting loops contains only one nucleotide (A8) spanning a seven base pair stem. This feature of the T2/T6 pseudoknot is particularly clear, since each of the adjacent nucleotides (C7 and G9) participate in well-defined Watson−Crick base pairs separated by the six well-defined Watson−Crick base pairs of stem 2. Previous to this NMR study, there had been no direct evidence as to whether a stable pseudoknot could be formed using only a single nucleotide in loop 1 to span a seven nucleotide stem 2. This observation supports models of other naturally occurring pseudoknots that may use a loop 1 of only one nucleotide in spanning a six or seven base pair stem 2.

In addition to the Watson−Crick base-pairing interactions within stem 1 and stem 2, RNA pseudoknots may contain additional tertiary structural features, such as specific interactions between the nucleotides of loop 2 and the minor groove side of stem 1 and between loop 1 and the major groove side of stem 2. NOEs between the base of A24 and protons of C19, for example (Figure 4), provide hints of these tertiary interactions. Nucleotide A8, being in the major groove of stem 2, has the potential to form specific interactions with the stem 2 nucleotides. Isotope labeling of the T2/T6 RNA with carbon-13 and nitrogen-15, combined with multidi-

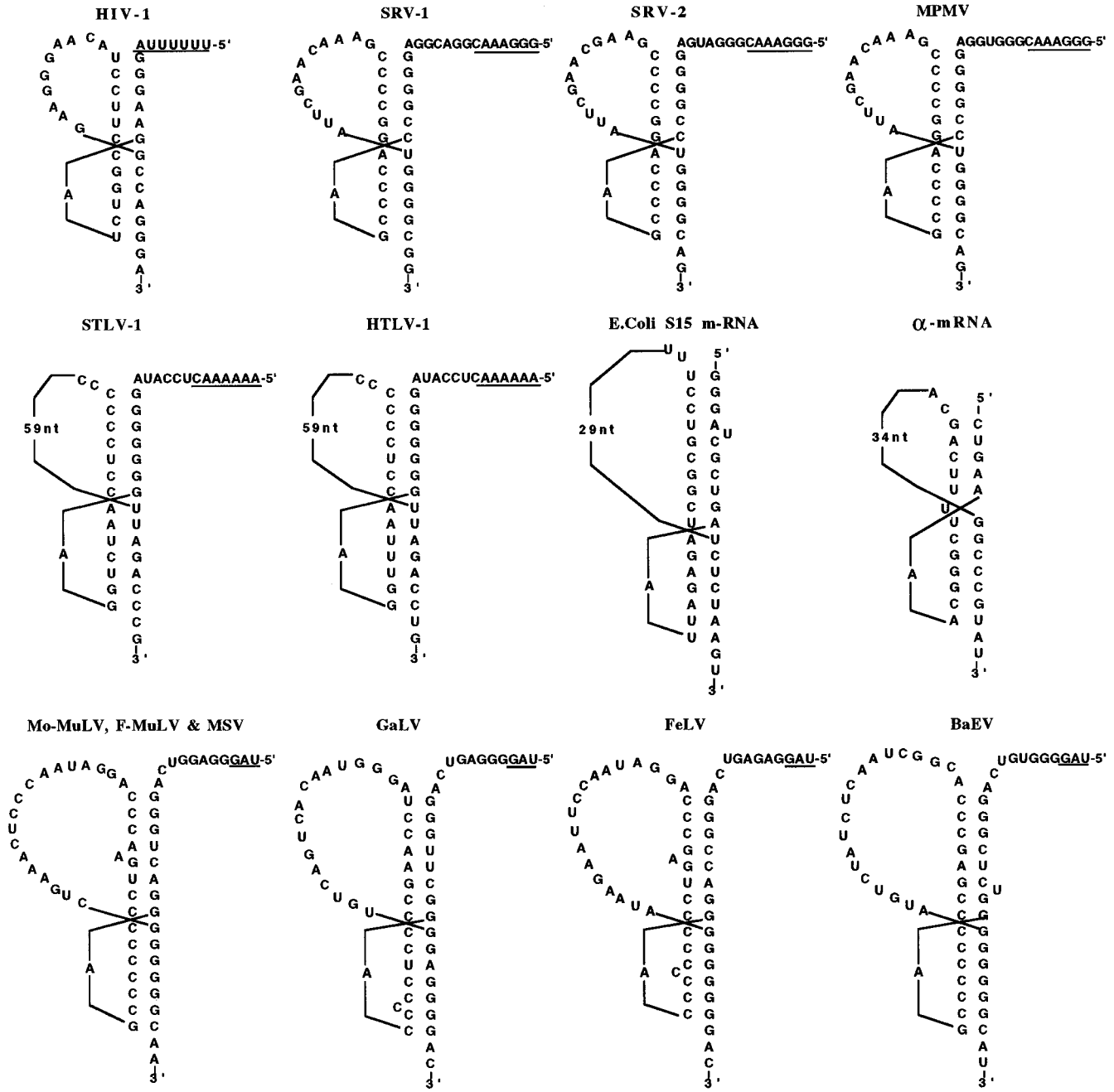

FIGURE 8: Diagrams of possible pseudoknots belonging to the CPK1 family. Top row: Examples of CPK1 pseudoknots with six base pairs in stem 2. Shown are RNAs from HIV-1 strain HXB2R, simian retrovirus 1 (SRV-1), SRV-2, and Mason−Pfizor monkey virus (MPMV). Each of these putative pseudoknots is associated with frameshifting at the viral *gag-pol* or *gag-pro* junction. The slippery sequence at each frameshift site is underlined. Second row: Examples of CPK1 pseudoknots with seven base pairs in stem 2. Shown are RNAs from the *gag-pro* frameshift site in simian T-cell leukemia virus (STLV-1) and human T-cell leukemia virus type 1 (HTLV-1), *E. coli* S15 mRNA, and the *E. coli* α-operon mRNA. Third row: Examples of CPK1 pseudoknots with seven base pairs in stem 2 and associated with readthrough in retroviral mRNAs. Stop codons at the readthrough sites are underlined. Shown are putative pseudoknots from Moloney murine leukemia virus (Mo-MuLV), Friend murine leukemia virus (F-MuLV) and murine sarcoma virus (MSV, same sequence), gibbon ape leukemia virus (GaLV), feline leukemia virus (FeLV), and baboon endogenous virus (BaEV). Other viral RNA pseudoknots that conform to the CPK1 motif occur near frameshift or readthrough sites in mouse hepatitis virus (MHV-JHV) and *Saccharomyces cerevisiae* double-stranded RNA virus (ScV).

mensional heteronuclear NMR experiments, may provide further details of the pseudoknot structure. These efforts have recently been initiated in our laboratories.

A comparison of imino proton resonances of the bacteriophage T2 and T6 pseudoknot with the imino resonances of the corresponding bacteriophage T4 pseudoknot (Qiu et al., 1996) indicates that these RNAs have similar structures, despite substantial differences in their nucleotide sequences in stem 1 and loop 2. This implies that many of our conclusions regarding the structural features of the T2/T6 pseudoknot can be applied to the bacteriophage T4 pseudoknot as well. As shown in Figure 1, the T2/T6 and T4 pseudoknots are identical in stem 2, loop 1, and the junction of the two stems.

The CPK1 Family of RNA Pseudoknots. A substantial number of putative RNA pseudoknots described in the literature have sequences and topologies that appear to be remarkably similar to the bacteriophage T2/T6 and T4 pseudoknots. Several examples of these pseudoknots are shown in Figure 8, drawn in such a way as to be easily

compared with the T2/T6 and T4 pseudoknots. We propose that the T2/T6 pseudoknot has structural features in common with these other putative RNA pseudoknots and that all of these pseudoknots are representatives of a "family" related by their similar structural features. We refer to this family of RNA pseudoknots as "CPK1" for "common pseudoknot motif 1", and propose that the members of the CPK1 family contain two coaxial stems, a stem 2 with a length of six or seven base pairs and a loop 1 consisting of a single nucleotide (usually adenosine), and variable lengths for stem 1 and loop 2.

The base pairings implied by the renditions of the pseudoknots shown in Figure 8 are similar to, but in some cases not identical to, the base-pairing schemes proposed by other authors. For example, models presented for a pseudoknot within the α operon mRNA of *E. coli* (Tang & Draper, 1989) and for a regulatory pseudoknot within the *E. coli* ribosomal protein S15 operon (Philippe et al., 1995) are the same as shown in Figure 8. The model of the SRV-1 pseudoknot presented by ten Dam et al. (1994) is the same as shown in Figure 8. The RNA sequence near the HTLV-1 *gag-pro* frameshift site is often presented as a stem−loop structure (Rice et al., 1985). The RNA sequence associated with the HIV-1 *gag-pol* frameshift site is usually presented as either a stem−loop or as a pseudoknot with different base pairings than we propose [for example, see Jacks et al. (1988b; Hatfield et al. (1992), and Taylor et al. (1994)]. In a figure shown by Wills et al. (1994) the MuLV, FeLV, BaEV, and GaLV *gag-pol* readthrough pseudoknots are drawn with two nucleotides in loop 1, although a model with one nucleotide in loop 1 would also be consistent with the available data.

It is intriguing that when the frameshift- and readthrough-directing pseudoknots are drawn in the CPK1 motif, these pseudoknots with similar functions have strikingly similar features (Figure 8). For example, most of the frameshift- or readthrough-directing pseudoknots begin with three guanosines at the 5′ end of stem 1, and each has a distinctly purine-rich side to their stems. A purine is usually the first dangling nucleotide after stem 2. These pseudoknots all occur slightly downstream of the frameshift or readthrough site (underlined in Figure 8), so that the CPK1 motif will have entered the ribosome when the frameshift or readthrough occurs. When the known frameshift-associated "XXX-YYYZ" slippery sequence (Brierley et al., 1989) is considered together with the features common to the CPK1 pseudoknot family (stem 2 of six or seven base pairs, variable stem 1 and loop 2, loop 1 consisting of an adenosine), the combination could be considered to be a consensus sequence for a frameshift or readthrough signal. The existence of an apparent consensus sequence lends support to the existence of the CPK1 family. The conserved adenosine in loop 1 is particularly intriguing. Perhaps it is involved in stabilizing the pseudoknot through an essential tertiary interaction in the major groove of stem 2 or has an essential interaction with the translational apparatus related to frameshifting and readthrough. Since the functions of the pseudoknots in the broad CPK1 structural family are divergent, we speculate that some of the common features may be conserved for structural reasons. Perhaps this structural arrangement represents a particularly stable motif, analogous to the "UNCG" and "GNRA" stable tetraloop motifs observed in RNA hairpins (Cheong et al., 1990; Heus & Pardi, 1991b). Of course, the existence of the CPK1 family of structurally related pseudoknots will remain hypothetical until structures are determined for some of the RNAs shown in Figure 8.

Recently, the structure of an RNA pseudoknot (named VPK) that is similar in sequence to the MMTV frameshifting pseudoknot was analyzed using NMR methods (Shen & Tinoco, 1995). In contrast to the base-pairing scheme proposed for the CPK1 pseudoknots, one or two base pairs were reported to be absent from stem 2 in VPK, loop 1 was found to contain two or three nucleotides and to be outside of the major groove, and an adenosine was found to be intercalated in the junction of the stems to form a flexible hinge (Shen & Tinoco, 1995). However, the VPK molecule lacked the dangling nucleotides at the 3′ end of stem 2, had three G-C base pairs switched to C-G in stem 2, and had one G-C pair switched to C-G in stem 1. As a result of these sequence changes, it is possible that the VPK and wild-type MMTV pseudoknots may differ in structure. In our studies of the bacteriophage T2/T6 and T4 pseudoknots we found that the 3′ dangling nucleotides stabilized the pseudoknot structure (Qiu et al., 1996). Perhaps at least one dangling nucleotide is important for stabilizing the MMTV frameshifting pseudoknot as well. If so, the wild-type sequence MMTV pseudoknot may closely resemble the other members of the CPK1 family.

The possible existence of the CPK1 family of structurally related pseudoknots clearly deserves further investigation. The structural and functional consequences of mutating the conserved adenosine of loop 1, making conservative mutations in the stems, or systematically changing the 3′ dangling nucleotides in the T2/T6 and proposed frameshifting and readthrough pseudoknots are intriguing. Mutational investigations are most informative when combined with an analysis of pseudoknot structure and functional viability. Such investigations are currently underway in our laboratories.

REFERENCES

Arnott, S., Hukins, D. W. L., & Dover, S. D. (1972) *Biochem. Biophys. Res. Commun. 48*, 1392−1398.

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.

Brierley, I., Digard, P., & Inglis, S. C. (1989) *Cell 57*, 537−547.

Brünger, A. T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Chastain, M., & Tinoco, I. (1992) *Biochemistry 31*, 12733−12741.

Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., & Varmus, H. E. (1995) *EMBO J. 14*, 842−852.

Cheong, C., Varani, G., & Tinoco, I. (1990) *Nature 346*, 680−682.

Colvin, R. A., White, S. W., Garcia-Blanco, M. A., & Hoffman, D. W. (1993) *Biochemistry 32*, 1105−1112.

Guéron, M., Plateau, P., & Decorps, M. (1991) *Prog. Nucl. Magn. Reson. Spectrosc. 23*, 135−219.

Hatfield, D. L., Levin, J. G., Rein, A., & Oroszlan, S. (1992) *Adv. Virus Res. 41*, 193−239.

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., Colvin, R. A., Garcia-Blanco, M. A., of White, S. W. (1993) *Biochemistry 32*, 1096−1104.

Jacks, T., Madhani, H. D., Masiarz, F. R., & Varmus, H. E. (1988a) *Cell 55*, 447−458.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., & Varmus, H. E. (1988b) *Nature 331*, 280−283.

Laing, L. G., & Draper, D. E. (1994) *J. Mol. Bol. 237*, 560−576.
Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
McPheeters, D. S., Stormo, G. D., & Gold, G. (1988) *J. Mol. Biol. 201*, 517−535.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
Moazad, D., & Noller, H. (1987) *Nature 327*, 389−394.
Müller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482− 6492.
Newport, J. W., Lonberg, N., & Kowalczykowski, S. C. (1981) *J. Mol. Biol. 145*, 105.
Philippe, C., Bénard, L., Portier, C., Westof, E., Ehresmann, B., & Ehresmann, C. (1995) *Nucleic Acids Res. 23*, 18−28.
Pleij, C. W. A., & Bosch, L. (1989) *Methods Enzymol. 180*, 289−303.
Pleij, C. W. A., Rietveld, K., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 1717−1731.
Puglisi, J. D., & Tinoco, I. (1989) *Methods Enzymol. 180*, 304.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I. (1990a) *J. Mol. Biol. 214*, 437−453.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I. (1990b) *Biochemistry 29*, 4215−4226.
Puglisi, J. D., Chen, L., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76−80.
Qiu, H., Kaluarachchi, K., Du, Z., Hoffman, D. W., & Giedroc, D. P. (1996) *Biochemistry 35*, 4176−4186.
Rice, N. R., Stephens, R. M., Burney, A., & Gilden, R. V. (1985) *Virology 142*, 357−377.
Rietveld, K., van Poelgeest, R., Pleij, C. W. A., van Boom, J. H., & Bosch, L. (1982) *Nucleic Acids Res. 10*, 1929−1946.
Rietveld, K., Pleij, C. W. A., & Bosch, L. (1983) *EMBO J. 2*, 1079−1085.
Rietveld, K., Linschooten, K., Pleij, C. W. A., & Bosch, L. (1984) *EMBO J. 3*, 2613−2619.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Shen, L. X., & Tinoco, I. (1995) *J. Mol. Biol., 247*, 963−978.
Tang, C. K., & Draper, D. E. (1989) *Cell 57*, 531−536.
Taylor, E. W., Ramanathan, C. S., Jalluri, R. K., & Nadimpalli, R. G. (1994) *J. Med. Chem. 37*, 2637−2653.
ten Dam, E., Brierley, I., Inglis, S., & Pleij, C. (1994) *Nucleic Acids Res. 22*, 2304−2310.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
van Belkum, A., Abrahams, J. P., Pleij, C. W., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 7673−7686.
Varani, G., & Tinoco, I. (1991) *J. Am. Chem. Soc. 113*, 9349−9352.
Willis, N. M., Gesteland, R. F., & Atkins, J. F. (1994) *EMBO J. 13*, 4137−4144.
Wyatt, J. R., Puglisi, J. D., & Tinoco, I. (1990) *J. Mol. Biol. 214*, 455−470.

BI9527350