NMR Structure and Dynamics of an RNA Motif Common to the Spliceosome Branch-point Helix and the RNA-Binding Site for Phage GA Coat Protein[†,‡]

Jeff S. Smith and Edward P. Nikonowicz*

Department of Biochemistry and Cell Biology, Rice University, Houston, Texas 77251-1892

Received June 30, 1998; Revised Manuscript Received August 3, 1998

ABSTRACT: The RNA molecules that make up the spliceosome branch-point helix and the binding site for phage GA coat protein share a secondary structure motif in which two consecutive adenine residues occupy the strand opposite a single uridine, creating the potential to form one of two different A·U base pairs while leaving the other adenine unpaired or bulged. During the splicing of introns out of pre-mRNA, the 2′-OH of the bulged adenine participates in the transesterification reaction at the 5′-exon and forms the branch-point residue of the lariat intermediate. Either adenine may act as the branch-point residue in mammals, but the 3′-proximal adenine does so preferentially. When bound to phage GA coat protein, the bulged adenine loops out of the helix and occupies a binding pocket on the surface of the protein, forming a nucleation complex for phage assembly. The coat protein can bind helices with bulged adenines at either position, but the 3′-proximal site binds with greater affinity. We have studied this RNA motif in a 21 nucleotide hairpin containing a GA coat protein-binding site whose four nucleotide loop has been replaced by a more stable loop from the related phage Ms2. Using heteronuclear NMR spectroscopy, we have determined the structure of this hairpin to an overall precision of 2.0 Å. Both adenine bases stack into the helix, and while all available NOE and coupling constant data are consistent with both possible A·U base pairs, the base pair involving the 5′-proximal adenine appears to be the major conformation. The 3′-proximal bulged adenine protonates at unusually high pH, and to account for this, we propose a model in which the protonated adenine is stabilized by a hydrogen bond to the uridine O2 of the A·U base pair. The 2′-OH of the bulged adenine adopts a regular A-form helical geometry, suggesting that in order to participate in the splicing reaction, the conformation of the branch-point helix in the active spliceosome may change from the conformation described here. Thus, while the adenine site preferences of the spliceosome and of phage GA may be due to protein factors, the preferred adenine is predisposed in the free RNA to conformational rearrangement involved in formation of the active complexes.

Unpaired or bulged nucleotides are a class of RNA secondary structural motifs that serve as sites of protein recognition and binding (reviewed in ref *1*), participate in RNA cleavage and ligation (*2*, *3*), and may play a role in the folding of large RNAs (*4*). The unique three-dimensional structures of RNAs with unpaired bases presumably contribute to their biological functions. The conformations of bulged bases in DNA oligonucleotides have been extensively studied using nuclear magnetic resonance (NMR)[1] spectroscopy. In general, bulged purine bases tend to stack into the helix (*5*) while bulged pyrimidine bases either stack into or loop out of the helix depending upon temperature and the identity of neighboring nucleotides (*6*, *7*). Fewer bulged bases in RNA oligonucleotides have been studied, but the conformation of a bulged purine base may depend on whether its flanking regions form Watson−Crick base pairs (*8*−*10*).

Genetic and biochemical investigations have revealed that a bulged base is necessary for the splicing of introns out of pre-mRNA (*3*, *11*). The first cleavage step of nuclear pre-mRNA splicing involves nucleophilic attack of the phosphate at the 5′-splice site by the 2′-OH group of the branch-point adenine nucleotide to produce the 5′-exon and the lariat intermediate (*12*). The branch-point nucleotide is located within a bimolecular helix formed by pairing between the U2 snRNA and the pre-mRNA within the spliceosome. The primary structure of the branch-point helix is highly conserved in yeast, with adenine serving as the branch-point residue and an A·U base pair occupying the 5′-flanking position (Figure 1) (*11*). Thus, the branch-point residue is the 3′-proximal adenine nucleotide of an RNA bulge motif defined by two adjacent adenine nucleotides opposite a uridine nucleotide, referred to in this report as the (A−A)·U motif. The primary structure of the branch-point helix is less conserved in mammals, but the optimal nucleotide sequence for splicing remains the (A−A)·U motif (*13*, *14*). In addition, both adenines can act as the branch-point nucleotide in mammals, but the 3′-proximal site preferentially does so ∼95% of the time (*3*).

[†] This work was supported by National Institutes of Health Grant (GM52115) to E.P.N.

[‡] Atomic coordinates for the refined structures have been deposited with the Brookhaven Protein Data Bank under accession 17ra.

* Address correspondence to this author. E-mail: edn@bioc.rice.edu. Phone: (713) 527-4912. Fax; (713) 285−5154.

[1] Abbreviations: NMR, nuclear magnetic resonance; NTP, nucleoside triphosphate; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; DQF-COSY, double quantum filtered correlated spectroscopy; 2D, two-dimensional; 3D, three-dimensional; HetCor, heteronuclear correlation; HMQC, heteronuclear multiple quantum coherence; HSQC, heteronuclear single quantum coherence; rMD, restrained molecular dynamics; RMSD, root-mean-square deviation; NH, imino; NH_2, amino.

S0006-2960(98)01558-X CCC: $15.00
Published on Web 09/11/1998

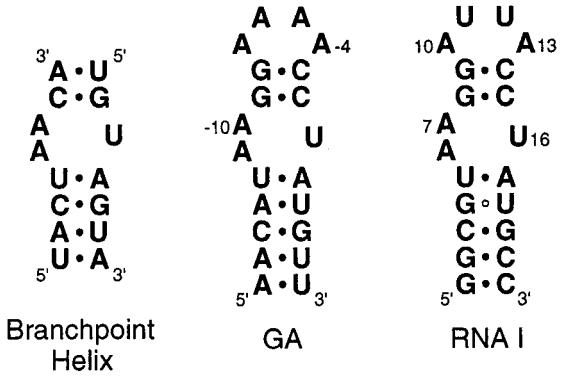

FIGURE 1: Sequence and secondary structures of the conserved branch-point helix from yeast, the RNA operator from phage GA, and the RNA molecule used in this study, RNA I. Residues in the GA operator are numbered relative to the replicase start codon. Each secondary structure contains the (A−A)•U secondary structure motif.

The bulged base motif is also important for protein recognition in bacteriophage. Ms2 and GA are single-stranded positive sense RNA phage that infect *Escherichia coli* by adsorbing to conjugative F pili. Their coat proteins bind a 21 nucleotide region of their genome containing the ribosome-binding site and start codon of the replicase gene, preventing translation of the replicase and promoting the encapsidation of the RNA into phage particles (*15*, *16*). The coat protein binding site consists of a base-paired stem with a single unpaired adenosine and a four nucleotide loop (Figure 1). Mutagenesis studies have shown that, provided the overall secondary structure is maintained, there are only a few nucleotides whose identity is necessary for binding to Ms2 coat protein: an adenine at position −4, a pyrimidine at position −5, and purines at positions −7 and −10 (*16*). GA coat protein is less restrictive: it does not require a pyrimidine at −5 and is able to bind hairpins with purines at either position −10 or position −11 (*17*). The wild-type GA operator contains the (A−A)•U motif so that adenines occupy both the −10 and −11 positions, but biochemical investigations have demonstrated that the GA coat protein binds hairpins with single bulged adenines with 3-fold greater affinity if the bulged base occupies the 3′-proximal rather than the 5′-proximal position (*17*).

The (A−A)•U motif can thus contribute to a variety of biological functions in both eukaryotic and prokaryotic systems, but its three-dimensional molecular structure is not known. While it is possible that other protein or RNA factors permit the motif to function in a variety of roles, it is also possible that the motif has unique structural or dynamic characteristics that are important for the various functions in which it participates. The motif may adopt a static conformation in which either the 5′- or 3′-proximal adenine is preferentially bulged. Alternatively, there may be no preference for either adenine to bulge, inducing the uridine to rapidly switch between the two adenine positions and result in a dynamic and perhaps disordered region of the helix.

We have used heteronuclear NMR spectroscopy to study the solution structure and dynamics of the (A−A)•U motif in a 21 nucleotide RNA hairpin containing a phage GA coat protein-binding site whose four nucleotide loop has been replaced by the more stable AUUA loop from the related phage Ms2 (Figure 1). Our results indicate that the 3′-proximal adenine preferentially bulges but remains stacked into the helix and is predisposed to conformational rearrangement. The structural and dynamic features of the (A−A)•U motif appear to play a significant role in pre-mRNA splicing and GA coat protein binding.

MATERIALS AND METHODS

All enzymes were purchased from Sigma Chemical (St. Louis, MO), except for T7 RNA polymerase, which was prepared as described (*18*). Deoxyribonuclease I type II, pyruvate kinase, adenylate kinase, and nucleotide monophosphate kinase were obtained as powders, dissolved in solutions of 15% glycerol, 1 mM dithiothreitol, and 10 mM Tris-HCl, pH 7.4, and then stored at −20 °C. Guanylate kinase and nuclease P1 were obtained as solutions and stored at −20 °C. Unlabeled 5′-nucleoside triphosphates (5′-NTPs) were purchased from Sigma, phosphoenolpyruvate (potassium salt) was purchased from Bachem, and 99% [^{15}N]-ammonium sulfate and 99% [^{13}C]glucose were purchased from Cambridge Isotope Labs.

Preparation of RNA Samples. The hybrid phage Ms2/GA RNA sequence, RNA I, shown in Figure 1 was synthesized in vitro using T7 RNA polymerase and a synthetic DNA template (*19*, *20*). The nucleotide sequence of the stem was modified from the wild-type GA sequence to promote efficient transcription and to simplify the resonance assignments. Unlabeled RNA molecules were prepared from 20 mL transcription reactions using 4 mM 5′-NTPs. Isotopically labeled RNA molecules were prepared from 14 mL transcription reactions using 3 mM uniform ^{15}N- and ^{13}C/^{15}N-enriched 5′-NTPs as described (*21*, *22*). The RNA molecules were purified by passage through 20% (w/v) preparative polyacrylamide gels, electroluted (Schleicher and Schuell), and precipitated with ethanol. The purified oligonucleotides were dissolved in 1.0 M NaCl, 20 mM potassium phosphate, pH 6.8, and 2.0 mM EDTA and dialyzed extensively against 10 mM NaCl, 10 mM potassium phosphate, pH 6.8, and 0.05 mM EDTA, using a Centricon-3 concentrator (Amicon Inc.). The samples were diluted with buffer to a volume of 210 μL and lyophilized to a powder. For experiments involving the nonexchangeable protons, the samples were exchanged twice with 99.9% ^2H_2O and then resuspended in 210 μL of 99.96% ^2H_2O. For experiments involving the exchangeable protons, the samples were resuspended in 210 μL of 90% H_2O/10% ^2H_2O. The samples contained ∼110−130 A_{260} OD units in 210 μL (∼2.3−2.7 mM).

Preparation of Ms2 Coat Protein. The Ms2 coat protein variant was expressed in *E. coli* BL21(DE3)pLysS cells containing the plasmid for the V75E;A81G Ms2 coat protein variant (*23*). The protein was purified as reported (*23*) and stored at 4 °C. Protein−RNA complex was prepared on ice by diluting RNA hairpin (∼1.2 mM) with coat protein (0.22 mM) in a buffer of 10 mM NaCl, 100 mM Na_2SO_4, and 10 mM potassium phosphate, pH 6.8, in 90% H_2O/10% ^2H_2O. The final concentration of complex was ∼0.18 mM. The noncompetitive and competitive RNA hairpins, 5′-(GGGAUACUGCUUCGGUAAGUCCC)-3′ and 5′-(GGGAUCACCAUUAGGGAUCUC)-3′, respectively, were added to the complex in molar ratios of 1:1 and 1:0.5, complex:hairpin.

NMR Spectroscopy. All NMR spectra were acquired on a Bruker AMX-500 spectrometer equipped with a ^{1}H-[^{13}C, ^{15}N] triple resonance probe, except for the ^{31}P-decoupled ^{13}C-^{1}H constant time (ct) HSQC experiment, which was collected on a Varian Unity Plus 600 MHz spectrometer equipped with a ^{1}H-[^{13}C, ^{31}P] triple resonance probe. Broad-band decoupling of the carbon and nitrogen resonances was achieved using GARP with $\gamma B_2 = 3125$ Hz for carbon and $\gamma B_2 = 1570$ Hz for nitrogen. $^{1}H_2O$ spectra were collected at 10° C with solvent suppression using either spin lock pulses or binomial $1\bar{1}$ or $1\bar{3}3\bar{1}$ read pulses with maximum excitation at 12.5 ppm. $^{2}H_2O$ spectra were collected at 22 °C with presaturation or spin lock pulses to suppress the residual HDO peak. Quadrature detection was achieved using the States-TPPI method, and acquisition was delayed by a half-dwell in all indirectly detected dimensions, unless otherwise noted. Typically, the data points were extended by 33% using linear prediction for the indirectly detected dimensions and the data were apodized using 1 Hz line broadening and 65° shifted sinebell functions. ^{1}H spectra were referenced to the internal HDO resonance at 4.80 ppm at 20 °C. The ^{13}C, ^{15}N, and ^{31}P spectra were referenced to external standards of TSP, NH_4OH, and TMP, respectively, which resonate at 0.00 ppm. All spectra were processed and analyzed with Felix 95.0 (Molecular Simulations, Inc.). Further details of the acquisition and processing parameters are provided as Supporting Information.

2D ^{13}C-^{1}H HMQC and HSQC spectra were collected to identify ^{13}C-^{1}H chemical shift correlations. 3D HCCH-COSY and HCCH-TOCSY (24 ms DIPSI-3 spin lock) experiments optimized for polarization transfer through the ribose carbons and a 2D ^{13}C-^{1}H HCCH-TOCSY (52 ms DIPSI-3 spin lock) optimized for polarization transfer through the adenine bases were collected in $^{2}H_2O$ to identify ribose spin systems and H8−H2 correlations, respectively (*24*, *25*). To identify intraresidue base−sugar correlations, a 2D ^{15}N-^{1}H HSQC experiment was acquired in $^{2}H_2O$ and optimized for two- and three-bond correlations as reported (*26*), except that the ^{15}N-^{1}H antiphase magnetization was refocused after the t_1 evolution period to achieve cross-peaks that were *in-phase* in ω_1 and ω_2 and to permit application of broad-band ^{15}N decoupling during the t_2 acquisition period.

Distance constraints for the nonexchangeable resonances were derived from 2D ^{1}H-^{1}H NOESY spectra (65, 100, 160, and 400 ms mixing times), a ^{13}C-edited 3D NOESY-HMQC spectrum (420 ms mixing time), and ^{13}C-edited 3D NOESY-ctHSQC spectra (75, 160, 210, and 420 ms mixing time) optimized for the ribose resonances in ω_2 and ω_3. For the exchangeable resonances, 2D ^{15}N-^{1}H HSQC spectra were collected to identify ^{15}N-^{1}H chemical shift correlations. 2D ^{1}H-^{1}H NOESY and ^{13}C-edited 2D NOESY-HMQC spectra optimized for imino (NH) proton resonances in ω_2 were acquired at 380 ms mixing time in $^{1}H_2O$ to obtain distance restraints involving the exchangeable protons.

Backbone dihedral angle constraints were derived from spectra collected to measure ^{1}H-^{1}H, ^{31}P-^{1}H, and ^{13}C-^{31}P coupling constants. A ^{31}P-decoupled DQF-COSY experiment and a 2D HetCor experiment were acquired in $^{2}H_2O$ with an unlabeled RNA sample. Three-bond ^{31}P-^{13}C coupling constants were measured using a pair of 2D ^{13}C-^{1}H ctHSQC experiments with and without ^{31}P decoupling in t_1 (*27*).

^{13}C $T_{1\rho}$ relaxation times were measured using 2D ^{13}C-^{1}H ctHSQC based experiments (*28*) separately optimized for C6/8 and C1′ regions by centering the ^{13}C carrier at 135.2 and 91.7 ppm, respectively. A 2.3 kHz ^{13}C spin lock field was used with delays of 4, 8, 12, 20, 24, 32, 40, 64, and 80 ms and 4, 8, 16, 24, 32, 36, 44, 52, 80, and 104 ms for the C1′ and C6/8 measurements, respectively. The 12 ms experiment was collected twice to provide an estimate of the error of the measured intensities. The ^{13}C-^{1}H cross-peak volumes were fit to a single-exponential decay.

Interproton Distance Constraints. Semiquantitative distance constraints between nonexchangeable protons were estimated from cross-peak intensities in 2D NOESY spectra and in 3D ^{13}C-edited NOESY spectra. Using the covalently fixed pyrimidine H5−H6 distance (∼2.4 Å) and the conformationally restricted sugar H1′−H2′ distance (2.8−3.0 Å) as references, peak intensities were classified as strong, medium, weak, or very weak and their corresponding proton pairs given upper bound distance constraints of 3.0, 4.0, 5.0, or 6.0 Å, respectively. Cross-peaks observed only at mixing times greater than 210 ms were classified as extremely weak and given 7.0 Å upper bound distance constraints to account for the possibility of spin diffusion. To improve convergence of the calculations, all distance constraints were given lower bounds of 0.0 Å. Distance constraints involving exchangeable protons were estimated from 2D ^{1}H-^{1}H and ^{13}C-^{1}H NOESY spectra and were classified as either weak, very weak, or extremely weak, except for the intra-base-pair distances A·U H2−NH, G·C NH−NH_2, and G·U NH−NH, which were classified as strong constraints. Intraresidue sugar-to-sugar constraints were not included in the calculations.

An initial set of structures was calculated using a shortened version of the simulated annealing protocol (described below) with a model-built A-form helix as a starting structure. A list of all proton pairs in the calculated structures closer than 5.0 Å (representing expected NOEs) was compared to the list of constraints. The NOESY spectra were then reexamined for predicted NOEs absent from the constraint list. In some cases, this allowed the unambiguous assignment of previously unidentified NOEs, but in other cases, the predicted NOEs were unobservable due to spectral overlap or the broadening of resonances by exchange with solvent. After the final calculations, virtually all predicted NOEs not in the list could be accounted for by spectral overlap or exchange broadening. No NOEs could be observed, however, for five proton pairs involving loop residue A_{13} that were consistently found to be closer than 5.0 Å in calculated structures (see Discussion).

Hydrogen-Bonding Constraints. Watson−Crick base pairs were identified using two criteria: the observation of a significantly downfield shifted NH or NH_2 proton resonance and the observation of strong G·C NH−NH_2 or A·U H2−NH NOEs. The wobble G·U pair was identified by observation of a strong NOE between guanine and uridine NH resonances. Hydrogen bonds were introduced as distance restraints of 2.9 ± 0.3 Å between donor and acceptor heavy atoms and 2.0 ± 0.2 Å between acceptor and hydrogen atoms. Constraints identified in this way were included in the calculations for base pairs G_1·C_{21}, G_2·C_{20}, C_3·G_{19}, G_4·U_{18}, U_5·A_{17}, G_8·C_{15}, and G_9·C_{14}.

Dihedral Angle Constraints. Constraints on the ribose ring and backbone dihedral angles were derived from semiquantitative measurements of $^3J_{H-H}$, $^3J_{H-P}$, and $^3J_{C-P}$ couplings (*27*, *29*). Sugar pucker conformations were determined from $^3J_{H1'-H2'}$ couplings in ^{31}P-decoupled 2D DQF-COSY spectra. Residues with couplings >7 Hz were constrained to the C2′-endo conformation through three of the torsion angles in the ribose sugar ring (*30*). Independent confirmation of sugar pucker conformation was provided by the observation of weak (<5 Hz) $^3J_{H3'-H4'}$ couplings, C3′ resonances shifted downfield to 76−80 ppm from the main cluster at 70−72 ppm, and C4′ resonances shifted downfield to 85−86 ppm from the main cluster at 82−84 ppm (*31*). Residues with weak (<5 Hz) $^3J_{H1'-H2'}$ couplings and large (>5 Hz) $^3J_{H3'-H4'}$ couplings were constrained to the C3′-endo conformation. Residues with intermediate $^3J_{H1'-H2'}$ couplings were left unconstrained to reflect the possibility of conformational averaging.

Dihedral angle constraints for the γ-torsion angle were derived from $^3J_{H4'-H5'}$ and $^3J_{H4'-H5''}$ couplings in the DQF-COSY spectrum and intraresidue H4′−H5′ and H4′−H5″ cross-peak intensities in the 75 ms mixing time 3D NOESY-ctHSQC spectrum. For residues in which H4′−H5′ and H4′−H5″ peaks in the DQF-COSY spectra were clearly absent, representing couplings of <5 Hz, γ was constrained to the gauche$^+$ conformation (60 ± 20°) (*30*, *32*). For residues with clear $^3J_{H4'-H5'}$ or $^3J_{H4'-H5''}$ couplings of >5 Hz and unequal H4′−H5′ and H4′−H5″ NOE intensities, γ was constrained to include both the trans and gauche$^-$ conformations (−120 ± 120°), reflecting the lack of stereospecific assignments for the H5′ and H5″ resonances. For residues with only weak or unobservable $^3J_{H4'-H5'}$ or $^3J_{H4'-H5''}$ couplings and unequal H4′−H5′ and H4′−H5″ NOE intensities, γ was left unconstrained to reflect the possibility of conformational averaging.

Dihedral angle restraints for the β- and ϵ-torsion angles were derived from $^3J_{P-H5'}$, $^3J_{P-H5''}$, and $^3J_{P-H3'}$ couplings measured in 2D ^{31}P-^{1}H HetCor spectra and $^3J_{P-C2'}$ couplings measured in 2D ctHSQC spectra. β was constrained to the trans conformation (180 ± 20°) for residues in which P−H5′ and P−H5″ peaks in the HetCor spectra were clearly absent, representing couplings of <5 Hz (*30*, *32*). For residues in which P−H5′ and P−H5″ peaks could be observed, β was left unconstrained to reflect the lack of stereospecific assignments and the possibility of conformational averaging. ϵ was constrained to the trans conformation (−145 ± 20°) for residues with $^3J_{P-H3'} > 5$ Hz and $^3J_{P-C2'} < 5$ Hz and to the gauche$^-$ conformation (−100 ± 25°) for residues with $^3J_{P-H3'} > 5$ Hz and $^3J_{P-C2'} > 5$ Hz (*30*, *32*). Residues with $^3J_{P-H3'} > 5$ Hz but for which $^3J_{P-C2'}$ could not be measured were loosely constrained to include both conformations (−120 ± 45°).

Dihedral angle restraints for α and ζ were derived from the observation that a trans conformation of either dihedral angle is generally associated with a large downfield shift of the bridging ^{31}P resonance (*33*). Because no such shift is observed for any of the ^{31}P resonances in the hybrid hairpin, α and ζ were loosely constrained to exclude the trans conformation (0 ± 120°) for all residues except those in the bulge or loop regions, which were left unconstrained. No dihedral angle constraints were used for the glycosidic angle χ. A total of 69 dihedral angle restraints (10 α, 12 β, 17 γ, 20 ϵ, and 10 ζ) were used in the calculations.

Structure Calculations. All calculations were carried out on Silicon Graphics Indy work stations using X-PLOR 3.1 (*34*). The dihedral angles of a model structure (generated using Insight II, Molecular Simulations, Inc.) were randomized to generate 50 starting structures which were used in a simulated annealing/restrained molecular dynamics (rMD) routine (*30*). The calculation protocol was divided into three stages: global fold, refinement, and final minimization. The global fold step consisted of 1000 cycles of unconstrained energy minimization, 15 ps of rMD at 1000 K using only hydrogen bond and NOE constraints, 24 ps of rMD at 1000 K during which repulsive van der Waals forces were introduced, 14 ps of rMD while cooling to 300 K, and 1000 cycles of constrained minimization. The structures were then refined with 500 cycles of constrained minimization; 3 ps of rMD at 1000 K, during which the β, γ, ϵ, and sugar ring dihedral constraints were slowly introduced followed by another 3 ps, during which the α and ζ constraints were introduced; 3.5 ps of rMD while cooling to 300 K; and 1000 cycles of constrained minimization. The final step consisted of 1000 cycles of conjugate gradient energy minimization using all constraints and repulsive van der Waals potentials. To determine the consistency of the NMR data with different possible base-pairing schemes in the bulge, three additional sets of calculations were performed: one with A_6•U_{16} base pair constraints, one with A_7•U_{16} base pair constraints, and one with A_6•U_{16} base pair constraints and an additional U_{16} O2-A_7 N1 constraint. These calculations were performed by taking the global-fold-stage structures of converged calculations and re-refining and reminimizing them with the appropriate base constraints. Calculated structures were viewed using MidasPlus (UCSF) and Insight II.

RESULTS

Chemical Shift Assignments. The nonexchangeable ^{1}H and ^{13}C resonances of RNA I (Figure 1) were assigned using standard heteronuclear techniques (*35*, *36*). Most of the base and ribose ^{1}H-^{13}C correlations are resolved and only the A_6 H2 and C_{14} H1′ resonances have spectral characteristics indicative of intermediate exchange. All 21 ribose spin systems, except for the incompletely labeled 5′-terminal nucleotide, were identified using 3D HCCH-COSY and 3D HCCH-TOCSY experiments (*24*, *25*). Four of the five adenine intrabase H8−H2 correlations were identified using similar experiments optimized for magnetization transfer through the adenine bases. Intraresidue base-to-sugar correlations were identified using 2D ^{15}N-^{1}H HSQC experiments optimized to yield the multiple bond correlations H5−N1, H8−N9, and H1′−N1/N9 (*26*). Eight of the 10 pyrimidine correlations were identified in these spectra, but the degeneracy of the N9 resonances permitted only 2 of the 11 purine correlations to be unambiguously identified. Interestingly, H2′−N1/N9 correlations were observed for U_{11}, U_{12}, and A_{13}, suggesting these nucleotides do not have C3′-endo ribose ring puckers or that they have unusual T_2 relaxation properties.

Sequential assignments for the nonexchangeable resonances were made using 3D ^{13}C-edited NOESY experiments

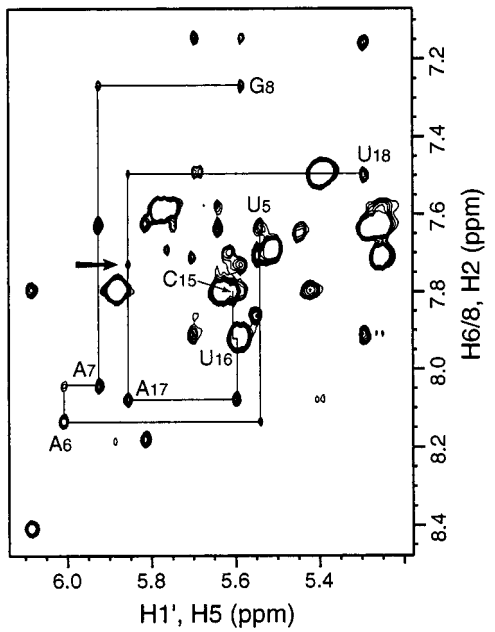

FIGURE 2: Base H6/8-to-H1′ region of the 2D 160 ms NOESY spectrum. The sequential walk is traced through the bulge region (residues U_5−G_8 and C_{15}−U_{18}) with intraresidue peaks labeled. The arrow points to the long-range NOE between A_7 H2 and A_{17} H1′.

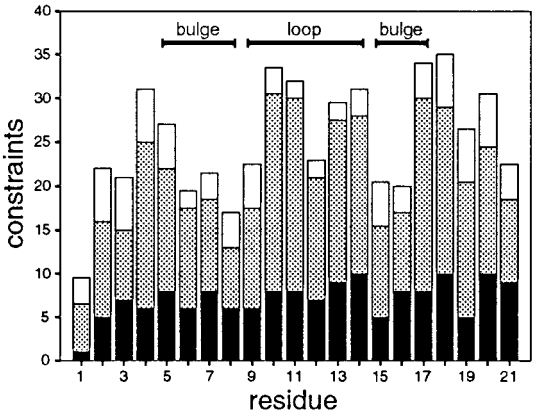

FIGURE 3: Distribution of experimental constraints per residue. Gray and black bars indicate the number of interresidue and conformationally restrictive intraresidue distance constraints, respectively, used in the calculations. Interresidue constraints are counted as one-half constraint for each residue involved. White bars indicate the number of backbone dihedral angle constraints. Base pair and sugar pucker constraints are not included.

to identify sequential H6/8−H1′ NOE connectivities (*35*). The intraresidue ^{15}N-^{1}H base-to-sugar correlations simplified the assignment process by permitting intraresidue cross-peaks to be distinguished from interresidue cross-peaks. The sequential base-1′ NOE connectivities are continuous through all 21 nucleotides in the 420 ms NOESY spectrum, but in the 160 ms spectrum, the interresidue connectivities are interrupted between loop nucleotides A_{10}, U_{11}, U_{12}, and A_{13}. The H6/8−H1′ sequential walk is continuous through the bulge region (Figure 2). The H6/8−H2′ interresidue connectivities are continuous through most of the molecule, but in the 160 ms spectrum the U_{12}−A_{13} and A_{13}−C_{14} connectivities are weak and the U_{11}−U_{12} connectivity is absent, consistent with disruption of the helix suggested by the C2′-endo ribose ring conformations of U_{11}, U_{12}, and A_{13}.

The exchangeable NH and NH_2 resonances were assigned using 2D NOESY and 2D ^{15}N-edited NOESY-HMQC experiments. Briefly, the NH resonances of the G·U base pair were first identified by their characteristically strong NH−NH NOE cross-peak. The remaining NH resonances were assigned using the weaker NOE connectivities between NH proton resonances of neighboring base pairs. These connectivities are continuous in the helix from G_2 to U_5. The cytidine NH_2 resonances were assigned using the strong intrabase-pair C NH_2 to G NH NOE cross-peaks. Independent confirmation of the resonance assignments was provided by strong cytidine intrabase NH_2 to H5 NOE cross-peaks in G·C base pairs, by strong uridine NH to adenine H2 NOE cross-peaks in A·U base pairs, and by uridine H5 to NH nitrogen correlations in long-range HMQC spectra. The NH proton resonances of U_{11}, and U_{12} and the NH_2 resonances of A_6, A_7, C_{15}, and all guanine nucleotides except G_4 were not observed and could not be assigned. Two adenine NH_2 resonances that are present in HSQC spectra but do not give rise to cross-peaks in NOESY spectra have been tentatively assigned to A_{10} and A_{13}. All other exchangeable proton and protonated nitrogen resonances were assigned. Because divalent metal ion is required for the splicing reaction, a 2D ^{15}N-^{1}H HMQC spectrum was collected for RNA I after addition of 10 mM Mg^{2+}. No chemical shift changes are observed, but the U_5 NH resonance is slightly weakened. This suggests that Mg^{2+} may associate with the RNA helix, but does not induce significant structural changes.

All internucleotide ^{31}P resonances are dispersed between −3.4 and −4.6 ppm. Eighteen out of 20 ^{31}P resonances were able to be assigned using the P−H3′ correlations from 2D ^{31}P-^{1}H HetCor spectra. Several P−H4′ and P−H5′/H5″ correlations also were present in the HetCor spectra and provided independent confirmation of the ^{31}P assignments.

Structure Calculations. The structure of RNA I was calculated using a restrained molecular dynamics routine starting from 50 structures with completely random backbone dihedral angles. The calculations used a total of 481 distance constraints and 69 dihedral angle constraints (Figure 3) to produce 12 converged structures (Figure 4). Structures were classified as converged if they were consistent with the NMR data and maintained correct stereochemistry. The converged structures had an average of 12 distance constraint violations between 0.1 and 0.3 Å, most of these involving the loop region. All converged structures violated less than three NOE constraints by more than 0.3 Å and none by more than 0.4 Å, and when the structures were arranged in order of increasing overall energy, the converged structures formed a plateau with similarly low overall and constraint violation energies (*30*). The convergence rate was very sensitive to the initial strength with which repulsive van der Waals forces were first introduced. To achieve a 25% convergence rate, van der Waals forces had to be introduced more slowly than in other protocols (*30*). In addition to the converged structures, the calculations produced a second class of structures that had either A_6 or A_7 unstacked from the helix and positioned in the minor groove. These structures were not classified as converged since A_6 and A_7 give rise to only one set of chemical shifts, and it is unlikely that such dissimilar structures could exchange rapidly on the chemical shift time scale. The heavy atoms of all 21 nucleotides of the final converged structures superimpose on the average structure with an average RMSD of 2.00 Å, consistent with

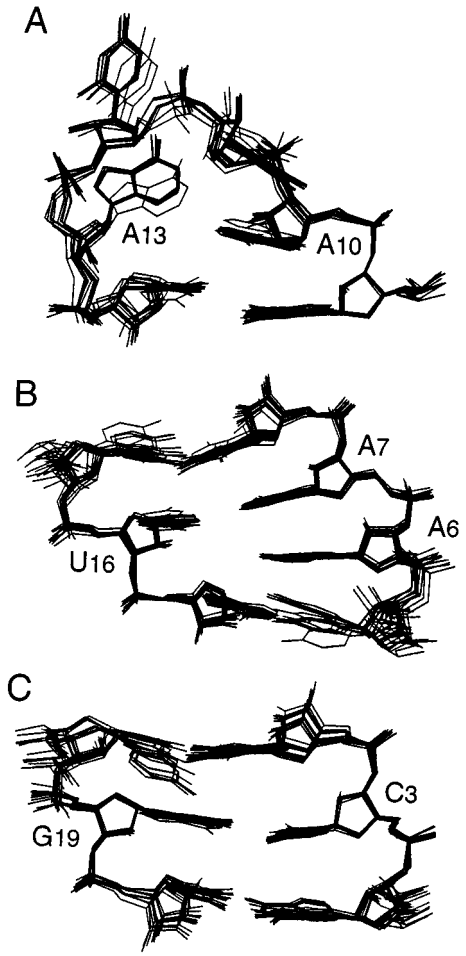

FIGURE 4: Local superposition of all 12 converged structures for the (A) loop, (B) bulge, and (C) stem regions. All views are into the major groove. The RMSDs between the individual structures and the average structure are listed in Table 2. Although the loop region is well defined, the $T_{1\rho}$ relaxation data show that the loop is more dynamic than the stem and bulge regions.

observations from other studies (*30*) that NMR methods define the long-range structure (such as bending) of helical RNAs with less precision than they define local structure. Additional structure calculation statistics are provided in Tables 1 and 2.

Structure of the Loop Region. The abundance of distance constraints in the loop region (residues G_9-C_{14}) defines the conformation of these nucleotides very precisely (Figure 3 and Table 2), and several of the structurally important NOEs in the loop involve the adenine H2 resonances. NOEs from A_{10} H2 sequentially to U_{11} H1′ and across the strand to C_{14} H1′ and C_{14} NH_2 indicate that the A_{10} base stacks on top of the $G_9{\cdot}C_{14}$ base pair in a largely A-form geometry. NOEs from A_{13} H2 to U_{11} H5′ and H5″ indicate that the A_{13} base points into the helix, but no NOEs could be identified that support stacking of A_{13} on the $G_9{\cdot}C_{14}$ base pair. Unusual interresidue sugar-to-sugar NOEs also play a key role in defining the structure of the loop. These include NOEs from U_{12} H4′ to U_{11} H5′ and H5″, from A_{13} H3′ to C_{14} H5′/H5″, and exceptionally strong NOEs from A_{10} H2′ to U_{11} H5′ and H5″. These NOEs indicate that the sugar rings of the loop pack together tightly in an arrangement distinctly different from standard helices.

Several nucleotides in the loop have unusual sugar−phosphate backbone conformations. The large H1′−H2′ couplings of residues U_{11}, U_{12}, and A_{13} indicate that their ribose sugar rings occupy the C2′-endo conformation, and the >5 Hz P−C2′ coupling of residue U_{11} indicates that its ϵ-backbone angle occupies the nonstandard gauche$^-$ conformation. The ϵ-angle of A_{13} adopts the trans conformation associated with A-form geometry, and that of U_{12} was constrained to include trans and gauche$^-$ conformations but is trans in all converged structures (Figure 5). H4′−H5′/H5″ couplings of >5 Hz in the DQF-COSY spectrum and H4′−H5′/H5″ cross-peaks of unequal intensity in the 75 ms 3D NOESY spectrum indicate that residues A_{13} and U_{12} have nonstandard γ-backbone angles. These angles were constrained to include both the trans and gauche$^-$ conformations but occupy the trans range in all converged structures (Figure 5). A nonstandard trans conformation for the α- and ζ-backbone angles is predicted to cause a downfield shift of the corresponding ^{31}P resonance (*33*). Since the ^{31}P resonances of RNA I all cluster around −4.0 ppm, it is unlikely any of these angles adopt the trans conformation. The α-,

Table 1: Summary of Experimental Distance and Dihedral Angle Constraints and Refinement Statistics

constraint	
NOE distance constraints	
intraresidue[a]	150
interresidue	293
mean number per residue	21
NOE constraints by category	
strong (0.0−3.0 Å)	15
medium (0.0−4.0 Å)	38
weak (0.0−5.0 Å)	145
very weak (0.0−6.0 Å)	101
extremely weak (0.0−7.0 Å)	144
base pair constraints	
total[b]	38
dihedral angle constraints	
ribose ring[c]	20
backbone	69
mean number per residue	4.2
violations	
average distance constraints > 0.3 Å[d]	1.0
RMSDs for distance constraints (Å)	0.031
average dihedral constraints > 0.5°[e]	3.4
RMSDs for distance constraints (deg)	0.158
RMSD from ideal geometry[f]	
bonds (Å)	0.006
angles (deg)	1.442

[a] Only conformationally restrictive constraints are included. [b] The number of base pair constraints does not include $A_6{\cdot}U_{16}$ or $A_7{\cdot}U_{16}$ base pair constraints. [c] Three torsion angles within each ribose ring were used to constrain the ring to either the C2′-endo or C3′-endo conformation. The ring pucker of residue A_6 was not constrained. [d] A distance violation of 0.3 Å corresponds to 5.0 kcal energy penalty. [e] A dihedral angle violation of 0.5° corresponds to 0.05 kcal energy penalty. [f] Calculated for the minimized average structure.

Table 2: Structure Calculation Statistics for Three Bulge Base Pair Constraints[a]

RMSDs	none	$A_6{\cdot}U_{16}$	$A_7{\cdot}U_{16}$
total	2.00	2.15	2.10
stem	0.54	0.54	0.54
bulge	0.56	1.09	0.81
loop	0.53	0.49	0.49

[a] RMSDs represent the average deviation between individual minimized structures and the average structure (heavy atoms only). Local structural features are precisely defined by NMR data, but long-range structural features (such as bending) are less well defined (*30*).

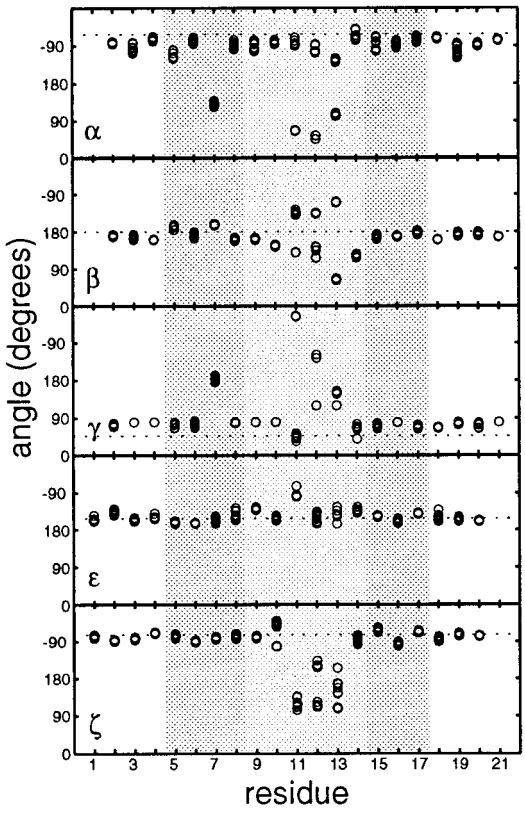

FIGURE 5: Phosphate backbone angles for the 12 converged structures. The bulge region is shaded dark gray and the tetraloop region is shaded light gray. Dotted lines indicate values for standard A-form helices.

ζ-, and β-angles were left unconstrained for loop residues A_{10}−C_{14}. In the converged structures, these angles tend to cluster into a few distinct conformations (Figure 5). All scalar coupling measurements indicate a standard A-form geometry for the sugar rings and the γ- and ϵ-backbone angles of residues G_9, A_{10}, and C_{14}.

A superposition of the loop regions from the 12 converged structures is shown in Figure 4 and the minimized average structure is shown in Figure 6. The helical base stack continues up the 5′-side of the loop, with U_{11} stacking on top of A_{10}, which in turn stacks on top of the G_9•C_{14} base pair. Residues U_{12} and A_{13}, however, do not participate in either the 5′- or 3′-base stack. The U_{12} base points out into solution, while the A_{13} base fits under the U_{12} sugar, pointed at A_{10} but twisted ∼60° relative to the plane of its base. None of the 2′-OH resonances were observed in any spectra, but the 2′-OH groups of U_{11}, U_{12}, and A_{13} point into the major groove in all converged structures.

Structure of the Bulge Region. Many of the exchangeable resonances in the bulge region (residues U_5−G_8 and C_{15}−A_{17}) are weak or not observable due to broadening caused by mobility or exchange with solvent, including the NH resonance of G_8 and the NH_2 resonances of A_6, A_7, and C_{15}. In addition, the U_{16} NH resonance has no observable cross-peaks in 2D NOESY spectra, preventing the direct identification of an A_6•U_{16} or A_7•U_{16} base pair. Because no distance constraints could be obtained for these exchangeable protons, most of the key NOE interactions that define the structure of the bulge involve the adenine H2 resonances. All three adenines in the bulge region (A_6, A_7, and A_{17}) have the weak sequential $H2_i$ to $H1'_{i+1}$ NOEs typical of A-form RNA helices. There are also cross-strand NOEs between A_{17} H2 and A_6 H1′ and between A_6 H2 and A_{17} H1′, indicating that the A_6 base stacks into the helix on top of the U_5•A_{17} base pair. The A_7 H2 has a cross-strand NOE to U_{16} H1′, indicating that it stacks into the helix as well, but it also has unusual long-range NOE cross-peaks to the A_{17} H1′ and U_{16} H2′ resonances (Figure 2) that suggest A_7 may be partially displaced toward the minor groove. Nevertheless, the pattern of intra- and interresidue H6/8−H1′ and H6/8−H2′ NOEs in the bulge region does not significantly differ from that observed in the surrounding stem regions.

The sugar and phosphate backbone torsion angles in the bulge region also lie within the limits of A-form RNA helices, with only a few exceptions (Figure 5). The H1′−H2′ coupling constant of A_6 is ∼4.7 Hz, intermediate between the values characteristic of C3′-endo or C2′-endo sugar conformations, suggesting that the A6 sugar is in dynamic equilibrium. H1′−H2′ and H3′−H4′ couplings indicate that all of the other sugars in the bulge region occupy the 3′-endo conformation. In addition, the H4′−H5′ and H4′−H5″ NOE cross-peaks of A_6 or A_7 (or both) are of unequal intensity in the 75 ms 3D NOESY spectrum, indicating a nonstandard conformation for the γ-backbone angle. Because of spectral overlap, γ was left unconstrained for both of these residues. Small H4′−H5′/H5″ couplings for the other nucleotides in the bulge region indicate that their γ-angles occupy the A-form gauche$^+$ conformation. In the original calculations, the A_6 sugar conformation was left unconstrained, and in all converged structures, A_7 γ is in the trans conformation while A_6 γ is in the normal gauche$^+$ conformation. Because the molecular dynamics parameters only included the 3′-endo sugar pucker as a low-energy conformation, a second set of structures was calculated using explicit 2′-endo constraints. In these calculations, all converged structures have both γ-angles trans. P−C2′ couplings indicate that the ϵ-angles of U_5, A_6, A_7, and C_{15} occupy the A-form trans conformation. The other ϵ-angles in the bulge region were constrained to include both the trans and gauche$^-$ conformations, but these angles are trans in all of the converged structures. Although the lack of downfield-shifted ^{31}P resonances in RNA I suggests that none of the α- or ζ-angles occupy the trans conformation, these angles in the bulge region were left unconstrained in the calculations and adopt the A-form gauche$^-$ conformation in all converged structures, except for A_7 α, which adopts the gauche$^+$ conformation. All of the β-backbone angles in the bulge region were also left unconstrained and occupy the A-form trans conformation in all converged structures.

Although the U_{16} NH resonance has no NOEs that would permit the direct identification of a base pair partner, the U_{16} NH nitrogen and hydrogen nuclei resonate near those of U_5 in the downfield region of the spectrum characteristic of Watson−Crick A•U base pairs. To identify the nature of this base pair, three sets of structures were calculated: one without any base pair constraints involving U_{16}, one with A_6•U_{16} base pair constraints, and one with A_7•U_{16} base pair constraints. None of the converged structures violated any of the bulge region distance constraints by more than 0.3 Å or any bulge region dihedral angle constraints by more than 0.5°, indicating that the available NOE and coupling constant data are consistent with both possible base pairs. Figure 7

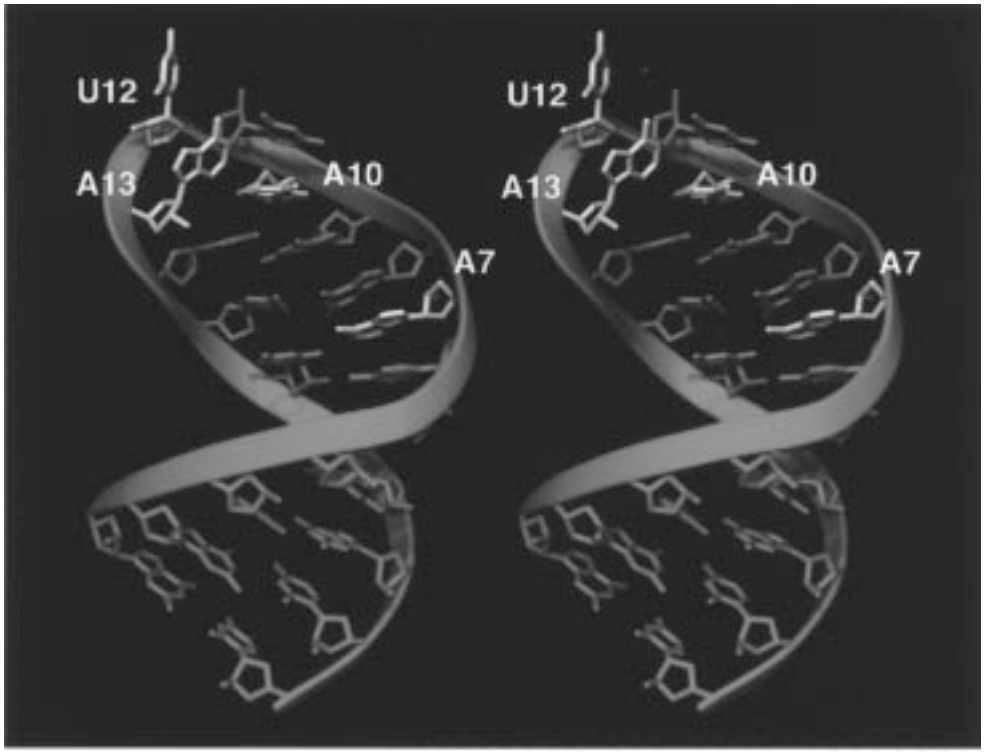

FIGURE 6: Stereoview of the minimized average structure of the hairpin RNA I. Nucleotides that are specifically required for Ms2 coat protein binding are colored white. The sugar−phosphate backbone is not distorted in the bulge region and the 2′-OH of A_7, the branch-point nucleotide, is not unusually positioned within the helix. In the loop, A_{10} stacks on the G_9•C_{14} base pair and U_{11} stacks on top of A_{10}. Although A_{13} points into the helix, there is no evidence of hydrogen bonds between A_{10} and A_{13} at pH 6.7.

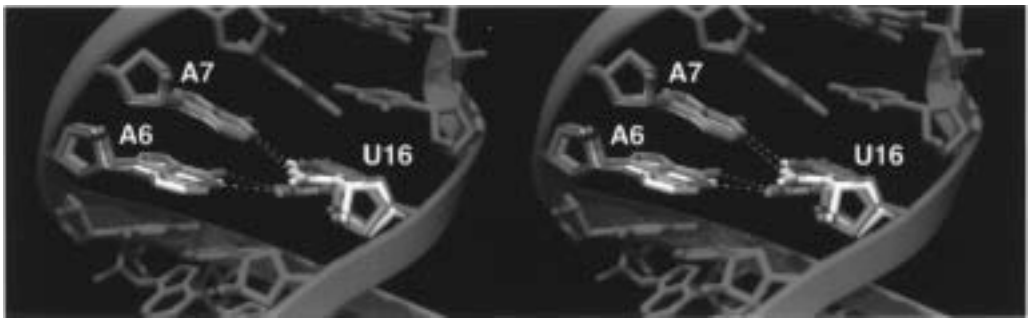

FIGURE 7: Stereoview superposition of the bulge regions of the minimized average structures from calculations with different base pair constraints (dashed lines) involving U_{16}: no base pair constraints (red), A_6•U_{16} base pair constraints (green), and A_7•U_{16} base pair constraints (white). The view is into the minor groove. All three structures equally satisfy the experimental NMR data.

shows a superposition of the bulge regions of the minimized average structures from each of the three sets of calculations. In the absence of base pair constraints, both A_6 and A_7 are stacked into the helix. The base of U_{16} is coplanar with A_7 but is slightly displaced from the normal Watson−Crick orientation toward the major groove in a geometry similar to a G•U wobble pair. The U_{16} NH proton is not in a position to form a hydrogen bond with A_6 N1, A_7 N1, or any other nitrogen atom acceptor. When base pair constraints are included in the calculations, U_{16} is no longer oriented toward the major groove and can form a base pair of standard geometry with either A_6 or A_7 without significant additional NOE violations. In all of the calculated structures, the base of A_7 is pointed toward the minor groove, consistent with the observed NOEs. The greatest difference between the three structures is the orientation of the U_{16} base. The bulge region of the average structure generated without base pair constraints superimposes on the average structures generated using A_6•U_{16} and A_7•U_{16} base pair constraints with RMSDs of 0.62 and 0.92 Å, respectively. In the absence of constraints on the A_6 sugar conformation, the 2′-OH groups of all residues in the bulge region point into the minor groove like those of the stem. When the A6 sugar is constrained to the 2′-endo conformation, the A_6 2′-OH group points at A7 O4′ and is less accessible to solvent from the minor groove. The geometry of the A_6, A_7, and U_{16} bases, however, do not appear to change significantly.

Protonation of A_7 and the Identification of an A_6•U_{16} Base Pair. The C2H2 correlations of A_6 and A_7 are weak in HMQC spectra and are broadened in the ^{1}H and ^{13}C dimensions, respectively (Figure 8). The chemical shift of A_7 C2 is strongly pH dependent, ranging from 153.8 ppm at pH 7.1 to 147.6 ppm at pH 5.45, indicating that the purine ring is being protonated at its N1 position with a p$K_a \approx 6.1$. A_{13} has a somewhat elevated pK_a of ∼5.4, but A_6, A_{11}, and A_{17} have pK_a values of <5.0.

A high pK_a does not indicate by itself, though, that an adenine bulges at neutral pH. At low pH, the ^{1}H and ^{15}N

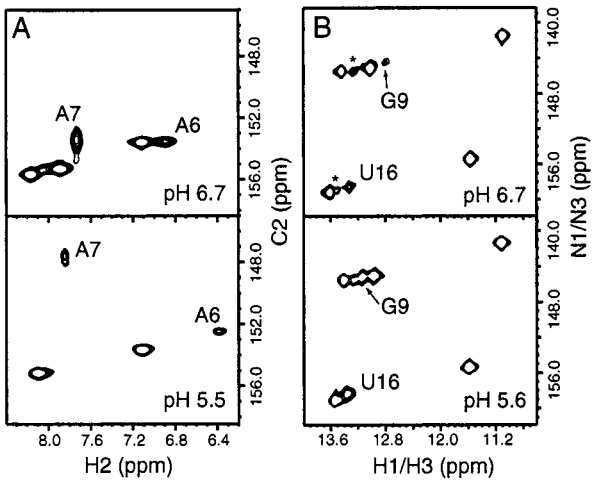

FIGURE 8: HMQC spectra at high and low pH showing (A) the C2−H2 region and (B) the NH region. The H2 resonance from residue A_6 is slightly broadened due to conformational flexibility at pH 6.7. As the pH is decreased, the C2 of A_7 migrates ~6 ppm upfield, indicating that its N1 position protonates with a pK_a of 6.1. The lack of change in the U_{16} NH resonance indicates that the A_6•U_{16} base pair is the major conformation at pH 6.7. A_{13} has the second highest pK_a, is partially protonated at pH 5.5, and is in intermediate exchange. The G_9 NH shifts downfield and intensifies at low pH, possibly due to formation of an A_{10}-A_{13} hydrogen bond. Minor conformations of the G_2 and U_5 NH resonances are indicated (*) in panel B.

chemical shifts of the U_{16} NH resonance change only slightly and remain in the region characteristic of Watson−Crick A•U base pairs (Figure 8). Since protonated adenines cannot form standard A•U base pairs, at low pH, the base pair partner of U_{16} must be A_6. If the U_{16} was rapidly exchanging between the A_6 and A_7 base pairs at pH 6.8, its NH chemical shifts would be the weighted average of these states. The lack of significant change of the U_{16} NH chemical shifts with pH suggests, then, that the A_6•U_{16} base pair and A_7 bulge is the major conformation at pH 6.8.

To account for the stabilization of the protonated adenine, we propose a model in which the A_6•U_{16} base pair forms an additional hydrogen bond between U_{16} O2 and A_7 NH^+, analogous to those seen in AH^+•C base pairs (*37*, *38*). This model was tested by calculating a set of structures using explicit hydrogen bond constraints between A_6, A_7, and U_{16}. None of the calculated structures violated any bulge region distance constraint by more than 0.3 Å or any bulge region dihedral angle constraints by more than 0.5°, indicating that the model is consistent with the available NMR data. All NOEs predicted by the model were confirmed except for those obscured by resonance boadening or spectral overlap. A converged structure from these calculations is shown in Figure 9.

Table 3: C6/8 and C1′ $T_{1\rho}$ Relaxation Times for RNA I

$T_{1\rho}$ (ms)			$T_{1\rho}$ (ms)		
residue[a]	C1′	C6/C8	residue	C1′	C6/C8
G_2	53[b]	38	U_{12}	77	79
C_3		46	A_{13}	59	75
G_4	50	43	C_{14}	47	48
U_5	50		C_{15}	55	44
A_6	53	51	U_{16}		43
A_7	48	38	A_{17}	56	48
G_8	45	23	U_{18}	52	41
G_9		27	G_{19}		49
A_{10}	43	43	C_{20}	58	43
U_{11}	72	56	C_{21}	68	47

[a] The transcription reactions were primed with 5′-GMP, resulting in incomplete labeling of the 5′-terminal nucleotide and preventing the G_1 C8 and C1′ relaxation rates from being measured. [b] The uncertainty in the measured relaxation times is ±5%.

In the Watson−Crick A•U region of HMQC spectra, there is an additional unassigned uridine NH correlation that is distinct but weak (~5% the intensity of other NH resonances). Since none of the loop uridines are in a position to form base pairs, it is likely that there are minor conformations in the bulge region that persist even at low pH. Weak additional C2H2 and C8H8 correlations are also present at pH 6.8 in HMQC spectra near the resonances of adenines in the bulge region, but these have not been assigned.

^{13}C Relaxation Measurements. The reorientation of a ^{13}C−^{1}H bond vector on the picosecond time scale can be assessed through its carbon $T_{1\rho}$ relaxation: the longer the relaxation time, the more mobile the ^{13}C−^{1}H pair (*39*). The $T_{1\rho}$ relaxation times for the base 6/8 and ribose 1′-carbon positions are listed in Table 3. The A_{13} C8 resonance has a relaxation time of 75 ms, while the other guanine and adenine resonances have relaxation times of 38−50 ms. The C6 relaxation times of the pyrimidines are between 41 and 48 ms, except for those of U_{11} and U_{12} which are, respectively, 56 and 78 ms. Loop nucleotides U_{11} and U_{12} also have the

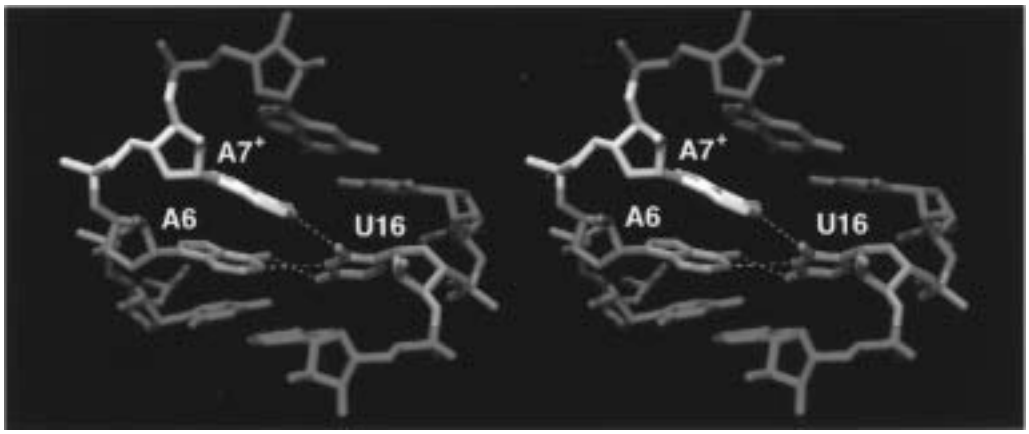

FIGURE 9: Stereo depiction of the model proposed for the (A−A)•U motif, containing the protonated A_7 (white) viewed into the minor groove. A_6 and U_{16} (green) form a Watson−Crick base pair with a propellar twist of ~30°. The A_7 NH^+ group forms a hydrogen bond with the O2 of U_{16} and stabilizes the A_7NH^+. The distance between A_7 N1 and U_{16} O2 is ~2.7 Å.

longest C1′ $T_{1\rho}$ values, 72 and 77 ms, respectively, followed by that of the 3′-terminal nucleotide at 68 ms. The increased mobilities of residues U_{11}, U_{12}, and A_{13} indicated by their long relaxation times are consistent with the decreased stacking in the 3′-portion of the loop. The C8 and C1′ relaxation times of A_6 and A_7 and the C1′ relaxation time of U_{16} are comparable to those of other nucleotides in the stem that participate in Watson−Crick or G•U wobble base pairs, indicating that the bulge region is not especially dynamic on the picosecond time scale and suggesting that the structure of the bulge is stable.

Ms2 Coat Protein Dimer Binds RNA I Specifically. Since there has been no study of the binding affinity of RNA I for phage GA or Ms2 coat protein, there is the formal possibility that the hairpin does not bind or does so nonspecifically. To better understand our results in terms of their implications for the coat protein−RNA complex, we performed binding competition experiments to assess the binding specificity of Ms2 coat protein for RNA I. Figure 10A shows the NH region of ^{15}N-^{1}H HMQC spectra of RNA I, in the presence of a 1:1 mixture of RNA I and Ms2 coat protein dimer. In the presence of the coat protein, Figure 10B, the U_5 NH resonance shifts downfield in both the ^{15}N and ^{1}H dimensions, while the U_{16} NH resonance shifts upfield in both dimensions and becomes more intense. Two additional NH correlations appear that have been tentatively assigned to U_{11} and G_8. The NH resonances of the stem do not noticeably change in position or intensity. Figure 10, panels C and D, shows the NH spectra of the RNA I-Ms2 coat protein mixture in the presence of noncompetitor and competitor RNA hairpins, respectively. Peaks in these spectra correspond only to RNA I since neither the competitor nor noncompetitor hairpins are ^{15}N enriched. The addition of an equal molar amount of noncompetitor hairpin lacking the sequence elements necessary for coat protein binding produces no change in the spectrum. The addition of a half mole ratio of competitor hairpin containing a phage Ms2 coat protein-binding site, however, results in a weakening of the protein bound U_5 and U_{16} resonances and a reappearance of the U_5 resonance from the free RNA I. These results indicate that the changes in the NMR spectra of RNA I caused by the addition of Ms2 coat protein result from the formation of a specific RNA−protein bound complex. Furthermore, the lack of change in the stem NH resonances upon binding is consistent with a bound RNA conformation essentially the same as that observed in X-ray crystal structures of the Ms2 protein−RNA complex (*40*).

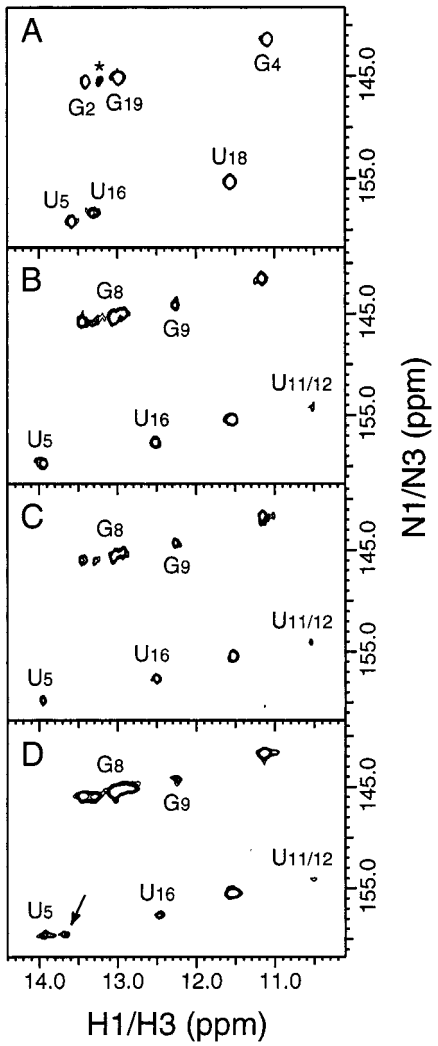

FIGURE 10: NH region of HMQC spectra of (A) free RNA I, (B) the Ms2 coat protein−RNA I complex alone, (C) after addition of noncompetitor RNA, and (D) after addition of competitor RNA. All samples were at pH 6.8 in a buffer of 10 mM NaCl, 100 mM Na_2SO_4 and 10 mM potassium phosphate. Cross-peaks correspond only to RNA I since the competitor and noncompetitor RNAs are not ^{15}N enriched. A minor conformation of the G_2 imino resonance is indicated (*) in panel A. Only the resonances that change in intesity or position are labeled in panels B−D. The U_5 cross-peak corresponding to the free form of RNA I is indicated by an arrow in panel D. The appearance of the protein free U_5 NH resonance only after addition of competitor RNA to the coat protein-RNA I complex demonstrates that Ms2 coat protein dimer specifically binds RNA I.

DISCUSSION

The coliphage Ms2 and GA coat protein dimers bind RNA hairpins composed of four-nucleotide loops and stems containing single bulged adenine bases. The position of the unpaired base relative to the hairpin loop is important for binding since the coat protein forms specific contacts to both elements (*40*, *41*). Bulged nucleotides also play a pivotal role in the splicing of nuclear pre-mRNA transcripts and group II introns (*2*, *3*, *11*). Although the spliceosome employs a variety of protein and RNA factors to accomplish this complex task accurately and efficienciently, the principal catalytic steps are likely to be RNA mediated. The structure and dynamics of the RNA hairpin investigated here suggest a common pathway for the formation of active complexes involving the (A−A)•U bulge motif.

Structure of the RNA Hairpin Loop. Despite many unusual non-A-form structural characteristics, the AUUA hairpin tetraloop has a very defined tertiary structure (Figure 6). The 5′-base stack continues into the loop, with the A_{10} base stacking on the G_9•C_{14} base pair in a normal A-form conformation, while the U_{11} base stacks on A_{10}. The U_{12} base does not stack at all but instead points out into solution approximately along the helix axis. The A_{13} base points into the helix but has a propeller twist of ~60° relative to the base of A_{10} and does not appear to stack on the G_9•C_{14} base pair. Since there is no spectral evidence for any of the proposed hydrogen bonds between the bases of A_{10} and A_{13} (*42*), the preciseness of the structure calculations suggests

that the loop may be effectively stabilized by stacking and hydrophobic interactions alone. The long $T_{1\rho}$ relaxation times of the U_{11}, U_{12}, and A_{13} C6/8 and C1′ atoms (Table 3) indicate that these residues are very mobile on the picosecond time scale and may occupy a greater region of space than indicated by the structure calculations (Figure 4A). However, the $T_{1\rho}$ relaxation times do not provide direct information on the amplitudes of the nucleotide motions, which may be small. Nevertheless, local conformational mobility on the 3′-side of the loop is consistent the broadness of the C_{14} H1′ resonance and the lack of NOEs between five pairs of atoms involving A_{13} that are consistently <5.0 Å apart in converged structures. The overall structure of the loop is similar to that of the Ms2 variant AUCA loop reported by Borer and co-workers (*9*, *42*).

Structure of the Adenine Bulge. The stem of RNA I contains a distinctive structural motif: two adjacent adenine nucleotides on one strand and a single uridine nucleotide on the complementary strand. In principle, this (A−A)•U motif can adopt either of two secondary structures, the A_6•U_{16} or the A_7•U_{16} base pair, but the spectral data indicate that the A_6•U_{16} base pair is the predominant form below pH 7. Importantly, the experimental NOE constraints do not restrict this region to a unique secondary structure. Indeed, the use of explicit A_6•U_{16} or A_7•U_{16} base pair constraints in the refinement procedure resulted in structures that equally satisfied all experimental distance and dihedral angle constraints (Figure 7).

The secondary structure of the bulge region is largely revealed by the protonation of A_7 at an unusually high pH. Protonated adenines have been identified in RNA molecules, but typically these bases participate in noncanonical AH^+•C base pairs (*37*, *38*). The cytidine stabilizes the protonated form of the adenine base by providing a hydrogen bond acceptor, resulting in an increase of the adenine pK_a. The positions of A_7 and U_{16} suggest a tertiary interaction that may facilitate protonation of the A_7 base. The O2 of U_{16} points into the minor groove when the A_6 and U_{16} nucleotides form a Watson−Crick base pair. The NH nitrogen of A_7 points toward the major groove face of the A_6•U_{16} base pair, allowing the A_7 NH^+ to form a hydrogen bond to the U_{16} O2. The A_7−U_{16} interaction is analogous to that found in the AH^+•C base pair scheme where the NH proton of AH^+ forms a hydrogen bond to the cytidine O2. It is unlikely that the high pK_a of A_7 is simply due to its lack of a base pair partner, since a single unpaired adenine flanked by G•C base pairs has a pK_a ~0.7 units lower than that of A_7 here (data not shown). While nucleotides A_7 and U_{16} are not excluded from adopting a Watson−Crick base pair near neutral pH, the orientation of the A_6 base within the helix does not permit the formation of additional stabilizing interactions beyond the stacking of the A_6 base into the helix. This model (Figure 9) is consistent with all of the available NMR data.

Comparison to Other Bulge-Containing Stems. Single and multiple bulged bases are a recurring theme in RNA secondary structures. Structural investigations of single bulged nucleotides in RNA suggest that the conformation adopted is dependent upon the nature of the base pairs of the flanking helices. A single bulged adenine flanked by Watson−Crick base-paired G and C residues adopts the expected intrahelical stacked conformation (*9*), but a bulged adenine flanked on one side by a non-Watson−Crick base pair adopts an extra-helical conformation and occupies the minor groove (*10*). Comparative analysis studies reveal the presence of (A−A)•U motifs in 16S and 23S rRNAs, but do not suggest that a common A•U base pair scheme exists (*43*). The phylogenetically conserved internal loop of the RNA-binding site for ribosomal protein S8 contains the (A−A)•U motif flanked by G•C and U•G base pairs in addition to two Mg^{2+}-binding sites. In this context, the 5′-proximal adenine base is extrahelical and forms a hydrogen bond to the 3′-proximal adenine which participates in the A•U base pair (*44*). This differs from the structure presented here in which both adenine bases are intrahelical and the 5′-adenine preferentially forms the A•U base pair.

A characteristic feature of the ^{31}P spectra of DNA duplexes containing unpaired bases is the downfield shift of a resonance corresponding to a position proximal to the bulged base (*5*, *6*). Interestingly, all the phosphorus nuclei proximal to A_6 and A_7 resonate in the center of the cluster of ^{31}P peaks. The γ-torsion angle of A_7 may alleviate any strain on the backbone imposed by the bulged nucleotide and thus permit the α- and ζ-torsion angles to adopt the preferred gauche configurations.

Significance of the (A−A)•U Bulge for the Branch-Point Helix. The first transesterification reaction in the splicing of introns out of pre-mRNA involves the nucleophilic attack of the phosphate at the 5′-splice site by the 2′-OH of a bulged adenine within the (A−A)•U motif of the branch-point helix (*12*). The (A−A)•U motif of RNA I is flanked on the 3′-side by a G•C base pair, but the conserved sequences of yeast and mammalian pre-mRNAs have a C•G pair at this site (Figure 1). While this substitution interferes with spliceosomal assembly events prior to the formation of the branch-point helix, it still allows accurate splicing and does not appear to affect the first catalytic step (*45*). The structure of RNA I, then, is likely to be representative of the free branch-point helix.

It has been proposed that the pairing of U2 snRNA with the branch-point region of pre-mRNA selectively bulges the branch-point nucleotide and aligns the 2′-OH of that nucleotide with the phosphodiester bond at the 5′-splice site (*3*). Our study of the (A−A)•U motif found in branch-point helices indicates that the 3′-proximal adenine does in fact preferentially bulge in the free RNA, consistent with the known preference of the branch-point nucleotide in yeast and mammals (*3*, *46*). The 2′-OH of the bulged adenine, however, adopts a standard geometry and points into the minor groove like all of the other base-paired residues. The sugar−phosphate backbone of the bulge region is also very regular, with the only departures from normal A-form geometry being the trans conformation of the bulged adenine γ-backbone angle and the mixture of C2′-endo and C3′-endo conformations in the 5′-proximal adenine ribose sugar ring. Since the conformational differences between the (A−A)•U motif and a regular A-form helix are minor, it is likely that the presentation of the bulged 2′-OH to the phosphate at the 5′-splice site requires a local structural rearrangement.

Such rearrangement could be assisted by a protein which binds to the branch-point helix and allows the branch-point adenine sugar to loop out from the helix, analogous to the rearrangement that occurs with binding of the Ms2 coat protein to its RNA operator sequence (*40*). Indeed, photo-

chemical cross-link studies have identified proteins that associate with the branch-point helix within the spliceosome (reviewed in ref *47*). This looped out conformation would permit the close approach of the 5′-splice site by extending the 2′-OH away from the negatively charged phosphate backbone of the branch-point helix. Alternatively, the mechanistic similarities between pre-mRNA processing within the spliceosome and the self-splicing of group II introns suggest that rearrangement could be assisted by RNA factors that would allow a more conservative movement, such as destacking of the bulged base out of the helix into the minor groove. Such RNA factors could include the helices formed by the pairing between U6 and U2 snRNA or between U6 snRNA and the 5′-splice site, both of which are close to the branch-point helix in the catalytic core of the spliceosome (*46*).

In yeast, only the 3′-site of the (A−A)•U motif can be used as a branch-point residue, while in mammals, either site can be used but the 3′-site is preferred (*3*). The existence of protein or RNA factors that assist conformational rearrangement of the branch-point helix could explain these differences, in a manner analogous to the (still poorly understood) way that Ms2 coat protein will only bind bulged adenines at the 3′-site of the (A−A)•U motif, while GA coat protein can bind bulged adenines at either site but binds the 3′-site with greater affinity (*17*). In this view, the branch-point helix selects the branch-point residue by bulging it in the free RNA and thus predisposing it to the protein- or RNA-assisted conformational rearrangements likely to be required for the first catalytic step of splicing.

Significance of the Structure of RNA I for the Ms2 Protein−RNA Complex. Comparison of the free and protein-bound structures of the Ms2 AUUA tetraloop reveals that significant conformational rearrangement must take place upon binding. In the crystal structure of the capsid−RNA complex, the base of A_{13} points out of the helix into the coat protein, U_{12} stacks on top of A_{10}, and U_{11} lies in the minor groove (*40*). This bound conformation requires that the base that stacks on A_{10} switch from U_{11} in the free RNA to U_{12} in the complex, although the position and sugar−phosphate backbone geometry of A_{10} remains largely unperturbed. In both the free and protein-bound states, the ribose rings of loop nucleotides U_{11}, U_{12}, and A_{13} adopt the C2′-endo conformation. The 2′-OH of U_{12}, the only ribose functional group that participates in a direct hydrogen bond to the protein, is pointed toward the major groove in both the free and bound conformations. The coat protein also recognizes the phosphate backbone in binding and makes direct hydrogen bond contacts to phosphoryl oxygens 5′ to A_{10}, U_{12}, and A_{13}. The experimentally determined γ-backbone torsion angles of U_{12}, A_{13}, and C_{14} and the ϵ-backbone angle of A_{13} are different in the free RNA from those observed in the structure of the protein bound RNA. The loop nucleotides that undergo major conformational change, residues U_{11}, U_{12}, and A_{13}, are also quite mobile on the picosecond time scale, presumably facilitating their structural rearrangement.

The fast dynamics of A_6 and A_7, however, do not indicate that either base is poised to loop out of the helix. In addition, the functional groups of the loop nucleotides that form direct protein contacts are exposed in the free RNA while those of A_6 or A_7 that would form direct protein contacts are buried deep in the major groove (Figure 6). Protonation of A_7 would be expected to inhibit complex formation by stabilizing the A_7-U_{16} interaction and preventing specific protein−base contacts, but the effect of pH on the affinity of GA coat protein−RNA binding is small—a factor of 4 less at pH 5 than at pH 7 (*17*). These observations suggest that the first step of complex formation is recognition and conformational rearrangement of the hairpin loop. The process of rearrangement in the loop or the structure of the bound loop itself may then destabilize the stem and initiate destacking of the bulged adenine. This mechanism is supported by the observation that the poorly ordered AAAA four nucleotide loop native to phage GA destabilizes the upper portion of the hairpin stem (E. DeJong, and E.P.N., unpublished data). Thus, the 3′-proximal bulged adenine is predisposed to protein-induced conformational rearrangement and is therefore likely to be preferentially associated with phage GA coat protein.

CONCLUSIONS

Our study of the RNA secondary structure motif (A−A)•U indicates that the 3′-proximal adenine preferentially bulges, predisposing it to structural rearrangement important for the motif's diverse functional roles in splicing of pre-mRNA and the binding of phage GA coat protein. Although the conformation of the branch-point helix in the catalytically active spliceosome is not known, the sugar−phosphate backbone geometry of the free helix is largely regular and the branch-point adenine 2′-OH group does not appear uniquely positioned to carrying out the splicing reaction. The 2′-OH group could be presented for attack at the 5′-splice site phosphate by protein- or RNA-mediated destacking of the bulged adenine. Similarly, the formation of a phage GA coat protein−RNA complex is also expected to require a conformational rearrangement where the unpaired adenine loops out from the helix. The fast dynamic properties of the loop nucleotides that change conformation upon coat protein binding indicate that these residues are less stabilized by hydrophobic stacking or hydrogen bonds in the free RNA than are the nucleotides of the (A−A)•U motif, which appear to be more rigid. Thus, the first step of complex formation is likely to be rearrangement of the loop nucleotides, which then destabilizes the stem and induces the bulged adenine to loop out. The 3′ proximal adenine requires the disruption of no more than one hydrogen bond to loop out of the helix and, thus, should be the adenine favored to interact with the GA coat protein.

ACKNOWLEDGMENT

We thank M. Michnicka for preparation of the T7 RNA polymerase and synthesis of the labeled oligoribonucleotides, Dr. D. G. Gorenstein for providing access to the Magnetic Resonance Facility at UTMB in Galveston in order to perform the ^{1}H-{^{13}C, ^{31}P} triple resonance experiments, and Dr. O. C. Uhlenbeck for providing the plasmid construct for the V75E;A81G Ms2 coat protein variant. We also thank E. DeJong for helpful discussions and critical reading of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Experimental NMR acquisition and processing parameters, chemical shifts of the ^{1}H, ^{13}C, ^{15}N, and ^{31}P resonances and

a plot of the backbone dihedral angles as functions of the corresponding ^3J scalar coupling constants (4 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Wyatt, J. R., and Tinoco, I., Jr. (1993) in *The RNA World* (Gesteland, R. F., and Atkins, J. F., Eds.) pp 465−496, Cold Spring Harbor Laboratory Press, New York.
2. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene 82*, 5−30.
3. Query, C. C., Moore, M. J., and Sharp, P. (1994) *Genes Dev. 8*, 587−597.
4. Tang, C. K., and Draper, D. E. (1990) *Biochemistry 29*, 5232−5237.
5. Rosen, M. A., Live, D., and Patel, D. J. (1992) *Biochemistry 31*, 4004−4014.
6. Chastain, M., and Tinoco, I., Jr. (1991) *Prog. Nucleic Acid Res. Mol. Biol. 41*, 131−177.
7. Morden, K. M., and Maskos, K. (1993) *Bioploymers 33*, 27−36.
8. Wimberly, B., Varani, G., and Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078−1087.
9. Borer, P. N., Lin, Y., Wang, S., Ruggenbuck, M. W., Gott, J. M., Uhlenbeck, O. C., and Pelczer, I. (1995) *Biochemistry*, 6488−6503.
10. Greenbaum, N. L., Radhakrishnan, I., Patel, D. J., and Hirsh, D. (1996) *Structure 4*, 725−733.
11. Parker, R., Siliciano, P. G., and Guthrie, C. (1987) *Cell 49*, 229−239.
12. Sharp, P. A. (1987) *Science 235*, 766−771.
13. Wu, J., and Manley, J. L. (1989) *Genes Dev. 3*, 1553−1561.
14. Zhuang, Y., and Weiner, A. M. (1989) *Genes Dev. 3*, 1545−1552.
15. Bernardi, A., and Spahr, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3033−3037.
16. Witherell, G. W., Gott, J. M., and Uhlenbeck, O. C. (1991) *Prog. Nucl. Acid Res. 40*, 185−220.
17. Gott, J. M., Wilhelm, L. J., and Uhlenbeck, O. C. (1991) *Nucleic Acids Res. 19*, 6499−6503.
18. Davanloo, P., Rosenburg, A. H., Dunn, J. J., and Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035−2039.
19. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8789.
20. Heus, H. A., and Pardi, A. (1991) *J. Mol. Biol. 217*, 113−124.
21. Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., and Williamson, J. R. (1992) *Nucleic Acids Res. 20*, 4515−4523.
22. Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., and Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507−4513.
23. LeCuyer, K. A., Behlen, L. S., and Uhlenbeck, O. C. (1995) *Biochemistry 34*, 10600−10606.
24. Pardi, A., and Nikonowicz, E. P. (1992) *J. Am. Chem. Soc. 114*, 9202−9203.
25. Legault, P., Farmer, B. T., II, Mueller, L., and Pardi, A. (1994) *J. Am. Chem. Soc. 116*, 2203−2204.
26. Sklenar, V., Peterson, R. D., Rejante, M. R., and Feigon, J. (1994) *J. Biomol. NMR 4*, 117−122.
27. Legault, P., Jucker, F. M., and Pardi, A. (1995) *FEBS Lett. 362*, 156−160.
28. Yamazaki, T., Muhandiram, R., and Kay, L. E. (1994) *J. Am. Chem. Soc. 116*, 8266−8278.
29. Williamson, D., and Bax, A. (1988) *J. Magn. Reson. 76*, 174−177.
30. Varani, G., Aboul-ela, F., and Allain, F. H.-T. (1996) *Prog. NMR Spectrosc. 29*, 51−127.
31. Santos, R. A., Tang, P., and Harbison, G. S. (1989) *Biochemistry 28*, 9372−9378.
32. Altona, C. (1982) *Recueil Rev. 101*, 413−433.
33. Gorenstein, D. G. (1984) *Phosphorus-31 NMR: Principles and Applications*, Academic Press, New York.
34. Brünger, A. T. (1992) *X-PLOR Version 3.1 Manual*, Yale University, New Haven.
35. Pardi, A. (1995) *Methods Enzymol. 261*, 350−380.
36. Dieckmann, T., and Feigon, J. (1997) *J. Biomol. NMR 9*, 259−272.
37. Legault, P., and Pardi, A. (1994) *J. Am. Chem. Soc. 116*, 8390−8391.
38. Cai, Z., and Tinoco, I. J. (1996) *Biochemistry 35*, 6026−6036.
39. Peng, J. W., and Wagner, G. (1994) *Methods Enzymol. 239*, 563−596.
40. Valegård, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G., and Liljas, L. (1997) *J. Mol. Biol. 270*, 724−738.
41. Wu, H.-N., and Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221−8227.
42. Kerwood, D. J., and Borer, P. N. (1996) *Magn. Reson. Chem. 34*, 146.
43. Gutell, R. R. (1993) *Nucleic Acids Res. 21*, 3051−3054.
44. Kalurachchi, K., Uma, K., Zimmermann, R. A., and Nikonowicz, E. P. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2139−2144.
45. Pascolo, E., and Seraphin, B. (1997) *Mol. Cell. Biol. 17*, 3469−3476.
46. Madhani, H. D., and Guthrie, C. (1994) *Annu. Rev. Genet. 28*, 1−26.
47. Krämer, A. (1996) *Annu. Rev. Biochem. 65*, 367−409.

BI981558A