A Model of the Iron Responsive Element RNA Hairpin Loop Structure Determined from NMR and Thermodynamic Data†

Lance G. Laing‡ and Kathleen B. Hall*

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received June 3, 1996; Revised Manuscript Received August 13, 1996⊗

ABSTRACT: The iron responsive element (IRE) is a conserved RNA structure that is found in the 5′ UTR of ferritin mRNA and in the 3′ UTR of transferrin receptor mRNA. It is the binding site of the iron responsive protein (IRP), and the interaction is part of the regulation of cellular iron metabolism. The IRE six-nucleotide hairpin loop, 5′$C_1A_2G_3U_4G_5N_6$, is conserved in sequence, and mutations have shown that it is required for IRP binding. On the basis of the thermodynamic and NMR experiments utilized here, the IRE loop structure 5′$C_1A_2G_3U_4G_5C_6$, is described in detail. Measurements of loop stability show that it has 2.9 kcal/mol more free energy than predicted. NMR data suggest that there is hydrogen bonding between C1 and G5 in a tertiary interaction across the loop. A model structure, produced by MC-SYM/energy minimization, illustrates the conformational flexibility of U4 and C6, which appear to exhibit considerable local motion in solution. NMR data indicate that the position of G3 is not well defined, leading to two families of loop structures.

Translational regulation of protein synthesis mediated through mRNA sequence or structure is certain to be a frequent and powerful mechanism; however, only a few examples have been described in detail. Perhaps most familiar is the posttranscriptional regulation of the proteins that participate in cellular iron metabolism. In this system, the mRNAs of ferritin and the transferrin receptor contain one or several copies of a conserved stem-loop structure, the iron response element, or IRE (Klausner et al., 1993; Theil, 1994). The IRE is recognized by a cellular protein, the iron regulatory protein, or IRP, IRF, IRE-BP, FRP, or P90 (Mullner et al., 1989; Rouault et al., 1988; Leibold & Munro, 1988), resulting in either inhibition or enhancement of translation. In the ferritin 5′ UTR, the single IRE is insufficient to block translation, but when bound by the IRP, protein synthesis is inhibited (Aziz & Munro, 1987; Hentze et al., 1987; Walden et al., 1988; Gray et al., 1993), possibly through steric exclusion of assembly of the ribosomal preinitiation complex on the mRNA (Gray & Hentze, 1994). In the 3′ UTR of the transferrin receptor, five copies of the IRE are sites for IRP binding; in the bound form, translation of receptor mRNA is presumably enhanced through protection of the mRNA from degradation (Casey et al., 1989). The IRP itself is apparently regulated by the presence or absence of an iron−sulfur cluster: in the presence of the cluster, the protein cannot bind the IRE, while in its absence, it binds specifically to the RNA (Philpott et al., 1994; Hirling et al., 1994).

The IRE contains two regions that are both necessary for IRP binding; one element is a six-nucleotide loop at the end of a five base pair stem, and the second is the contiguous region containing a bulged cytosine nucleotide, followed by another flanking stem. Phylogenetic comparison and mutation (Barton et al., 1990; Harrell et al., 1991; Bettany et al., 1992; Jaffrey et al., 1993) of this RNA have resulted in the identification of a consensus sequence and secondary structure (Kuhn & Hentze, 1992) shown in Figure 1. The loop sequence and the bulged nucleotide are conserved, while the sequence of the stems does not appear to contribute to protein recognition. Binding experiments with the IRP and RNAs containing mutations in the six-nucleotide loop, 5′CAGUGN, seemed to suggest that no single mutation significantly disrupts IRP binding (Leibold et al., 1990; Barton et al., 1990). Larger effects were noted by Jaffrey et al. (1993) with single substitutions in the loop showing at most a 70-fold loss of binding affinity. Deletion of single nucleotides had larger effects (Hentze et al., 1988; Haile et al., 1989; Jaffrey et al., 1993); the interpretation of these results is difficult, since alternative RNA structures could form that prevented binding. Deletion of the bulged cytosine led to a much greater reduction in IRP binding, suggesting that the structure at this position may be recognized by the protein. However, the structure around this site is variously depicted as a single base bulge for IREs in the 3′ UTR of transferrin receptor or as an asymmetric internal loop for the 5′ UTR of ferritin, and thus the structural consequences of the deletion are not always clear. The structural context of the six-nucleotide loop and the unpaired cytosine is critical for IRP binding, as shown by Kikinis et al. (1995) in experiments that altered the loop and stem length and integrity.

Additional data in support of the consensus IRE and its role in IRP binding were obtained by Henderson et al. (1994) in selection experiments. The six nucleotides of the loop were randomized, as was the single nucleotide at the bulge. Preferentially bound RNAs contained the loop sequence shown in Figure 1, and an unpaired "bulged" C. Most unexpected was the discovery of a minor loop species with the sequence 5′UAGUAN. These data led to the suggestion that there is a conserved interaction between nucleotides 1

† This work was supported by the NIH (GM46318 to K.B.H.; GM16891 to L.G.L.) and the Lucille P. Markey Charitable Trust (90-47).

* Address correspondence to this author. Telephone: (314)-362-4196. FAX: (314)-362−7183.

‡ Current address: Scriptgen Pharmaceuticals, Inc., Medford, MA.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0006-2960(96)01310-4 CCC: $12.00

and 5 in this loop that is somehow important for IRP binding (Henderson et al., 1994; Sierzputowska-Gracz et al., 1995). Finally, photochemical cross-linking of the IRE to the IRP was mapped to two sites in the RNA: in the loop near the fifth nucleotide and two nucleotides upstream from the conserved unpaired cytosine (Basilion et al., 1994).

The conservation of the IRE sequence and of its secondary structure, supported by the recent biochemical data, indicates that both are important features of its recognition by the IRP. Because the RNA loop might have an unusual structure, and because the structure around the unpaired cytosine seemed to be functionally significant, the IRE from the human ferritin heavy chain 5′ UTR was chosen for a detailed investigation into its thermodynamic stability and structural determination by NMR. Results indicate that the 5′CAGUGC loop is more stable than predicted from thermodynamic rules, undoubtedly due to the formation of a tertiary base pair formed within the loop. A model is presented, on the basis of the NMR data, to describe the loop structure.

MATERIALS AND METHODS

RNA was synthesized by one of two methods for this work. Duplex RNA strands were synthesized by solid-phase chemistry on an Applied Biosystems Model 391 PCRMate using 2′-*O*-methyl cyanoethyl phosphoramidites (Glenn Research) as described (Scaringe et al., 1990; Usman et al., 1992). The deblocked oligomers were then ethanol precipitated and readied for gel electrophoresis or column chromatography as described below. All other RNA was synthesized *in vitro* using SP6 (Stump & Hall, 1993) or T7 RNA polymerase (Milligan et al., 1987) from DNA oligonucleotide templates, typically in 10 mL transcription reactions. Following addition of 25 mM EDTA and ethanol precipitation from the transcription reaction, the RNAs were separated on an 8 M urea/20% polyacrylamide gel or applied to a Zorbax anion-exchange column (MAC-MOD Analytical, Inc.). Polyacrylamide gels were UV-shadowed; RNA bands were removed and then electroeluted using a Schleicher and Schuell Elutrap device where eluted RNA was removed from the trap at 20 min intervals to reduce RNA loss through the final membrane. Apparently due to the RNA structure, the traditional dialysis with selective molecular weight cutoff membrane resulted in significant loss of material, even with 1000 or 500 molecular weight cutoff. Instead, the RNA was ethanol precipitated in 0.3 M sodium acetate. The RNA was then resuspended twice in 3 mL of 0.4 M NaCl, renatured at 75 °C, and then precipitated again with 3 volumes ice-cold 100% ethanol. The RNA was resuspended in H_2O and lyophilized 2−3 times to remove trace amounts of ethanol prior to a 1 h dialysis using a Spectrapore Spectrum 500 MWCO dispodialyzer. This procedure typically yielded 0.5 mg of pure RNA/mL of transcription. RNAs containing 5−deutero-UTP or -CTP were prepared (Brush et al., 1988; Hayatsu, 1976) using $(N^2H_4)_2SO_3$ in ^2H_2O and incorporated enzymatically by *in vitro* transcription. IRE-G10 RNA sequences were synthesized with ^{15}N-UTP or -GTP (Nikonowicz et al., 1992) in *in vitro* transcription reactions.

Melting Experiments. Thermodynamic stability of the various RNA structures was determined from thermal denaturation of the purified RNA in a Gilford 260 spectrophotometer equipped with a 2527 Gilford thermoprogrammer. Absorbance measurements were taken at 260 nm with heating at a rate of 1 °C/min in 1 M NaCl and 10 mM cacodylate buffer, pH 6. Thermodynamic values for two-state transitions were obtained from global fitting of the first derivative curves as previously described (Laing & Draper, 1994) and represent a minimum of four independent experiments.

Melting experiments were also used to determine monomolecularity in NMR buffer at high RNA concentrations. The various RNA hairpin species were melted over an 87-fold concentration range (up to 0.85 mM RNA) with no detectable variation in the melting temperature.

NMR Experiments. RNA samples for NMR experiments were generally 1.5 mM RNA in 10 mM sodium phosphate, pH 5.8, and 30 mM NaCl. Spectra were acquired on a Varian Unity 500 MHz NMR spectrometer equipped with a Z-Spec IDTG500−5 5 mm or Z-Spec MIDTG-3 3 mm triple resonance gradient probe (Nalorac Cryogenic Corp., Martinez, CA). One NOESY (300 ms τ_m) and one TOCSY (150 ms) experiment was acquired on a Varian Unity 600 MHz spectrometer to provide increased resolution. All NMR data were collected using the States hypercomplex method (States et al., 1982) at 20 °C unless noted otherwise. NMR data were processed with VNMR 5.1 (Varian) or NMRPipe/NMRDraw (Molecular Simulations, Inc.) software. NMRCompass (Molecular Simulations, Inc.) was used to analyze spectra on a Silicon Graphics Indigo workstation.

1D NOE experiments were acquired at 4 °C with a 10K spectral width using a 1−3−3−1 binomial pulse with the offset centered near 10 ppm. 2D SS-NOESY (Smallcombe, 1993) spectra were collected at 4 and 20 °C with τ_m from 200 to 400 ms and a 10 K Hz spectral width with a short gradient pulse to remove the remaining solvent signal. Spectra were zero filled to 4K by 1K real points. 2D ^{1}H NOESY spectra were collected for nonexchangeable protons at various temperatures and 80, 100, 200, 300, and 400 ms mixing times. 2D ROESY experiments (Bax & Davis, 1985a) were acquired at 30 and 50 ms mixing times at 10 and 20 °C. 2D ^{1}H TOCSY spectra were collected at various temperatures for 60, 80, 100, 120 ms mixing times with a MLEV-17 spin-lock (Bax & Davis, 1985b). 3D ^{1}H TOCSY-NOESY spectra (Wijmenga et al., 1994) were collected using both a 100 ms spin-lock, 400 ms τ_m and a 60 ms spin-lock, 250 ms τ_m, with spectral widths of 2900 Hz in both dimensions. DQF-COSY spectra (Rance et al., 1983) with spectral widths of 2280 Hz in t_1 and t_2 were zero filled to 4K by 1K and weighted with a shifted sine bell. ^{1}H/^{15}N sweep widths were 4000 Hz/6000 Hz for ^{15}N/^{1}H HMQC in D_2O (Bax et al., 1990) with ^{15}N centered at 180 ppm, J = 20; and 3000 Hz/1600 Hz for the pfgHMQC in H_2O (Bax & Potchapsky, 1992) with the ^{15}N centered at 150 ppm, J = 90 ($^{15}NHCl_2$ at 0.0 ppm) with Garp decoupling during acquisition. ^{31}P−^{1}H HETCOR experiments (Frey et al., 1985; Sklenár et al., 1986; Sklenar & Bax, 1987) used a Nalorac ID500−5SS 5 mm indirect detection probe, with a 1300 Hz ^{1}H/820 Hz ^{31}P spectral width. Data were zero filled to 2K by 512. ^{31}P−^{1}H heteroTOCSY experiments (Kellogg, 1992; Kellogg & Schweitzer, 1993) were collected with 1300 Hz ^{1}H/560 Hz ^{31}P spectral windows with a 110 ms spin-lock. Data were zero filled to 4K by 512. Phosphorus was referenced to the internal phosphate at 0.0 ppm.

Structure Modeling and Refinement. The program MC-SYM (Macromolecular Conformation by SYMbolic generation) (Major et al., 1991; Gautheret et al., 1993) was used

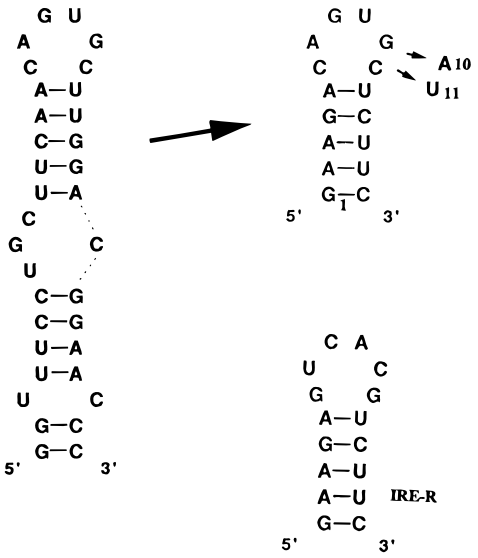

FIGURE 1: Sequence and secondary structure of the 34-nucleotide IRE element, based on the human ferritin heavy chain 5′ IRE. The sequences of the hairpins are IRE-G10, IRE-U11, and IRE-A10, all derived from the original hairpin loop sequence, and IRE-R, a random sequence RNA loop.

to generate a family of model RNA structures. MC-SYM uses a constraint satisfaction algorithm to search conformational space defined by a structural database. The nucleotide conformational database is constructed from all nucleic acid structures determined by X-ray and NMR. Structures from MC-SYM are refined using energy minimization in the INSIGHT 2.3.5 Discover 2.9 package (Biosym/Molecular Simulations, Inc.). Parameters for refinement used the Amber all-atom force field and included van der Waals repulsive/attractive forces and bond angle, dihedral angle, and bond length optimization, with hydrogen bonding terms on but Coulombic terms off. Conjugate gradient minimization was used only to eliminate high energy atom overlap/approach; after 15 iterations minimization was halted. These structures were examined for violations of NMR constraints and bad contacts and used for more extended analysis. All computations were run on Silicon Graphics Indy or Indigo II.

RESULTS

The IRE used for these experiments, shown in Figure 1, is based on the sequence of the human ferritin heavy chain 5′ UTR. The 16-nucleotide RNA stem−loop, with the mutations indicated, was used in both NMR and thermal denaturation experiments.

RNA Thermodynamics

The Hairpins. In addition to the normal IRE sequence, IRE-G10, several mutations were introduced into the loop for comparative purposes (Figure 1). The IRE-A10 mutant replaces the G10 with A10; results from the selection experiments (Henderson et al., 1994) suggested that a base pair formed between nucleotides 1 (C6) and 5 (G10) of the loop, and the A substitution would disrupt this interaction. This mutation reduced binding to the IRP about 70−fold (Jaffrey et al., 1993). The IRE-U11 loop is a naturally occurring one, found for example in the human transferrin 3′ UTR (Casey et al., 1988); a loop with this sequence bound 10-fold weaker to the protein (Jaffrey et al., 1993). Finally, the IRE-R loop is a random sequence six-nucleotide loop that is the control for the thermodynamic experiments.

Table 1: Thermodynamic Parameters for RNA Molecules

RNA	$\Delta H°$ (kcal/mol)	T_M (°C)	$\Delta S°$ (eu)	$\Delta G°$ (kcal/mol)
hairpin[a]	34.8	64.4	103.1	2.5
IRE-U11	53.9 ± 0.6	72.3 ± 0.3	156	5.5
IRE-G10	53.2 ± 0.8	71.5 ± 0.3	154	5.3
IRE-A10	55.3 ± 0.7	64.6 ± 0.5	164	4.5
IRE-R	46.6 ± 0.3	68.6 ± 0.2	136	4.3
IRE34.1	42.5 ± 3.5	61 ± 1	127	3.1
IRE34.2	66.0 ± 0.9	68 ± 1	193	6.0

[a] Calculated from nearest neighbor approximation (Serra & Turner, 1995) for the 16-nucleotide hairpin. IRE34.1/.2 is the first/second transition of the 34-nucleotide IRE RNA (Figure 1) in low salt. All values are the average of at least four independent measurements; errors in $\Delta S°$ and $\Delta G°$ are 20% and 10%, respectively.

Each RNA stem−loop was thermally denatured as a function of concentration, to look for evidence of dimer formation. In 1 M NaCl buffer used for thermodynamic analysis, as well as in low salt concentrations similar to NMR conditions (50 mM NaCl, 10 mM sodium cacodylate, pH 7), the melting temperature of the RNAs was not a function of the RNA concentration, indicating that the RNAs formed hairpins. All unfolding transitions were analyzed assuming a two-state model; thermodynamic parameters describing these RNAs are summarized in Table 1.

A hairpin loop with this particular five base pair stem is predicted to have a folding free energy ($\Delta G°_{37}$) of −2.5 kcal/mol, with an enthalpy of −34.8 kcal and an entropy of −103 eu, based on thermodynamic parameters for RNA structures using the nearest neighbor approximation (Turner et al., 1988; Serra et al., 1994; Serra & Turner, 1995). The IRE-R sequence most closely approximates these theoretical values, as shown in Table 1, although it is more stable than predicted. The IRE-G10 and IRE-U11 RNAs have nearly identical thermodynamic parameters, suggesting their interactions are similar, and both are significantly more stable than the IRE-R RNA. The IRE-A10 RNA, however, has a melting temperature 10 °C lower than either IRE-G10 or IRE-U11 and is less stable by about 1 kcal/mol in free energy. This behavior is consistent with a change in loop interactions.

The IRE. The melting of the entire 34-nucleotide RNA in low salt (50 mM NaCl) has two transitions, which are likely to correspond to the melting of the lower duplex and the melting of the upper stem and loop. Thermodynamic parameters describing these transitions are given in Table 1. However, in higher salt concentrations, as in 1 M NaCl, there is only a single discernable transition, suggesting that the IRE melts as a single species.

Hairpin Structure

Resonance Assignments. The 1D imino proton spectra of the IRE-G10 and IRE-A10 RNAs are shown in Figure 2. The assignments are indicated on the basis of 1D NOE walks between adjacent base pairs. For the IRE-G10 molecule, the G and U imino proton assignments were confirmed from [^{15}N]-uridine or [^{15}N]-guanosine RNA using ^{15}N/^{1}H pfgHMQC experiments. Of particular interest is the broad resonance at 10.6 ppm in the spectra of IRE-G10, which is a G imino proton as identified from spectra of [^{15}N]guanosine IRE-G10. The chemical shift of this resonance suggests it could be involved in a hydrogen bond to a carbonyl oxygen.

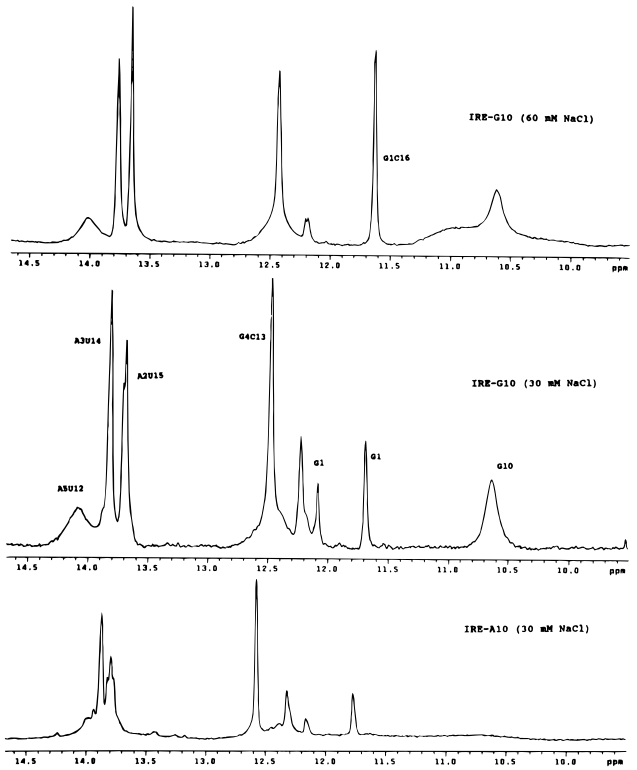

FIGURE 2: 1D imino proton spectra of IRE-G10 (in 30 mM NaCl and 60 mM NaCl) and IRE-A10 with the assignments indicated. In the higher salt, the stem is stabilized, reducing the number of G1 resonances. IRE-A10 is missing not only the G10 imino proton but also the resonance from A5U12, suggesting that this loop-closing base pair is destabilized. RNAs are approximately 1 mM in 30 mM NaCl and 10 mM sodium phosphate, pH 5.8, at 4 °C.

The A5U12 imino proton resonance is broad and much weaker than the other two AU resonances in the IRE-G10 spectrum; its intensity is weaker in the spectrum of IRE-U11 (data not shown) and absent in that of the IRE-A10 molecule. It shows a weak NOE to an AH2 proton and also to water. There are unfortunately several peaks assigned to the G1 imino proton. In 1D NOE experiments, interresidue NOEs from G1 to and from U15 and from G4 to and from U14 were clearly observed. However, no NOEs were observed from the resonance at 10.6 ppm, even in conditions where Mg^{2+} was added to hopefully stabilize the structure. This resonance, which in the IRE-G10 RNA is present from 4 to 30 °C, is also present in the IRE-U11 imino proton spectrum at 4 °C, but notably absent in the IRE-A10 and IRE-R spectra. Also shown in Figure 2 is the imino proton spectrum of IRE-G10 at higher sodium chloride concentration, where the stem is more stable, leading to one dominant peak for the G1C16 imino proton; no other changes in the spectrum or in the NOEs were observed.

SS-NOESY experiments were also used to confirm assignment and for structural information; for example, NOEs from the guanosine imino proton to cytosine amino protons are typical of a G·C base pair. Several features of the SS-NOESY spectrum are important and are shown in Figure 3. The peak at 12.58 ppm on the diagonal is assigned to the G4 imino proton. There are many NOEs corresponding to this peak, which have been assigned to exocyclic amino protons of C13 (interior proton at 8.3 ppm and exterior at 6.9 ppm); there are weak NOEs to amino protons of G4 (at 6.36 ppm) at low temperature and long mixing times. The

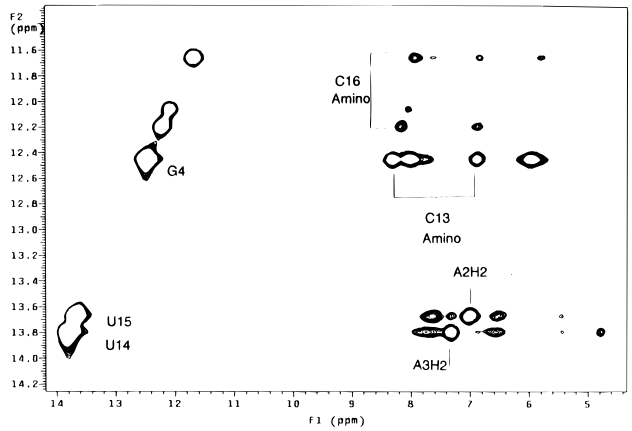

FIGURE 3: 2D SS-NOESY spectrum of IRE-G10 at 10 °C, 250 ms. Neither the imino proton of the A5U12 base pair nor the G10 at 10.6 ppm is observed in this experiment. The three imino protons from 11.8 to 12.5 ppm are assigned to G1 and result from heterogeneity of its position.

loop closing the A5U12 base pair shows a broad imino proton resonance at 14.1 ppm at 10 °C (pH 5.8) in 1D spectra, suggesting that it may be fraying or not forming a typical Watson−Crick pair; it is not visible in the SS-NOESY spectrum (Figure 3), although there is an NOE to water observed at lower contour levels. The G imino resonance at 10.6 ppm also does not appear in the SS-NOESY spectrum, although again there is a weak cross peak to water that can be observed at 4 °C and low contour levels. Finally, a peculiar property of the terminal G1·C16 base pair is that there are several imino protons that are associated with it, which are in the region from 12.2 to 11.6 ppm. These imino proton resonances are identical in spectra of the IRE-G10, IRE-A10, IRE-U11, and IRE-R RNAs. In the SS-NOESY experiments, there are few NOEs from this proton other than to cytosine amino protons. It is possible that this heterogeneity (which is also seen in spectra of the nonexchangeable protons) arises through populations of the RNAs in which G1 is paired to C17 (the $n + 1$ nucleotide added by the polymerase). At higher salt concentrations (>50 mM NaCl), these resonances coalesce into one at 11.6 ppm (Figure 2), leaving other NOE patterns unchanged.

The nonexchangeable aromatic proton assignments were determined by several methods. The cytosine and uridine H5/H6 assignments were confirmed using RNAs containing 5-deuterouracil or -cytosine in 2D TOCSY experiments. [^{15}N]Guanosine RNA was used in ^{15}N−^{1}H HMQC experiments optimized for two-bond couplings to confirm the identity of the four GH8 protons (Michnicka et al., 1993); the line width of the G1H8 proton is significantly greater than the others. All four guanosine ^{15}N7 resonances had similar chemical shifts, suggesting that their environments are similar; in particular, there is no evidence of the upfield shift expected from hydrogen bonding. The adenosine H2 protons could be identified on the basis of their long T_1 relaxation time (Wuthrich, 1986). All were confirmed by their NOEs to the 3′ intra- and interstrand ribose H1′ protons characteristic of A-form helices (Wuthrich, 1986). Sequential H8/H6 assignments were determined by the NOE walk from aromatic proton to ribose sugars observed in 2D NOESY spectra. The aromatic/ribose walk on the 5′ side of the hairpin from G1 at the base of the stem to G8 in the loop is shown in Figure 4. A similar walk can be done from

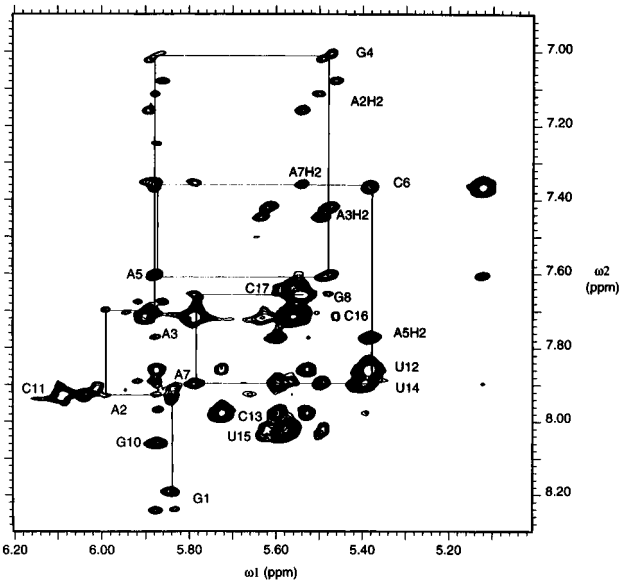

FIGURE 4: H1′ to aromatic region of a 2D NOESY experiment at 20 °C, 300 ms, and 600 MHz. The sequential assignments from G1 to G8 are indicated. There are several resonances assigned to A2H2 and A3H2. RNA is 1.5 mM in 30 mM NaCl and 10 mM sodium phosphate, pH 5.8.

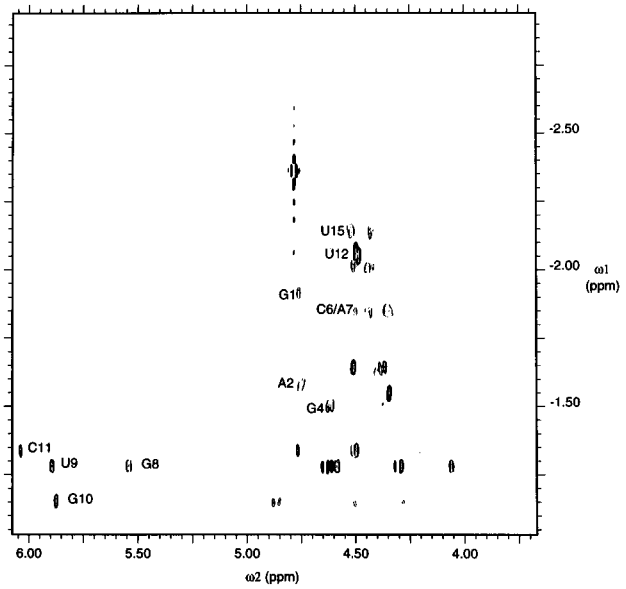

FIGURE 5: $^{31}P-^{1}H$ heteroTOCSY experiment at 20 °C, 110 ms, with assignments of ribose systems indicated.

the 3′ terminus, beginning at C16 and proceeding to the U12 H1′ at the top of the strand (data not shown). From there, the walk jumps to G10H1′ in the loop. To further confirm these assignments, an NOE walk using the aromatic/ribose 2′ protons was mapped, since these protons are closest in a canonical A-form helix. From the intensities of the aromatic/H1′ cross peaks, it is clear that all nucleotides have an *anti* conformation.

Spectra of IRE-G10 were also acquired in buffer containing 5 mM $MgCl_2$. The presence of the Mg^{2+} stabilized the stem structure somewhat, although it did not completely eliminate the heterogeneity at the terminus. The most significant features of the spectra were large changes in the chemical shifts of the ribose protons of G8 and U9. These two bases occupy pivotal positions in the loop structure, and their sensitivity to Mg^{2+} may indicate that there is a divalent ion binding site in the loop.

A notable and confusing feature of these spectra is the presence of several cross peaks corresponding to resonances from the first three base pairs of the stem. At the low salt concentration in these NMR experiments, as well as experiments with added 5 mM $MgCl_2$, the 5′GAA/3′CUU region of the stem is not stable, with the result that the bases and sugars sample several different environments. In the NMR spectra, this heterogeneity is observed in the number of imino protons associated with the G1·C16 pair, in the number of G1H8 to ribose H1′ cross peaks, and in the number of AH2 resonances observed for the A2·U15 and A3·U14 pairs. The identification of these AH2 resonances through their anomalously long T_1 relaxation times was crucial to deciphering these spectra. This heterogeneity could be due to the presence of an extra nucleotide at the 3′ end (usually a C17) which was not entirely removed in the purification; G1 could therefore pair with either C16 or C17. In addition, this sequence may be particularly susceptible to fraying, especially in low salt conditions. In any case, the extra cross peaks add complexity to the spectra.

To assign the ribose protons, 2D TOCSY and 2D ROESY experiments were used to identify the individual spin systems, as well as $^{31}P-^{1}H$ heteroTOCSY experiments which used the phosphorus as a filter for the ribose spin systems (Figure 5). Twelve of 16 ribose spin systems can be discerned in the $^{31}P-^{1}H$ heteroTOCSY spectrum, but the complete 1′ to 5′/5″ connectivities are not observed for most nucleotides, due to the small coupling constants in the ribose ring that limit the relay (and thus confirm sugar pucker descriptions). As shown in Table 2, assignments for loop riboses are nearly complete, although the 5′/5″ protons are not distinguished. The overlap in the proton spectra in the ribose region, despite 3D TOCSY-NOESY experiments and several NOESY experiments at 600 Mz, was too great to completely assign the sugars without ^{13}C-labeling.

Sugar Puckers. The DQF-COSY spectrum (Figure 6) indicates that the riboses in the stem, as well as the first C6 nucleotide on the 5′ side of the loop, are predominantly C3′-*endo*, typical of A-form duplex RNA. The next loop nucleotide, A7, has a ribose pucker that appears to be in equilibrium between C3′-*endo* and C2′-*endo*, based on the 1′/2′ coupling. The G8 ribose showed similar 1′/2′ and 3′/4′ couplings in DQF-COSY experiments, indicating that it was in equilibrium between C3′-*endo* and C2′-*endo* puckers. No 3′/4′ cross peak was observed for the sugars of U9 and G10, indicating they are best described by C2′-*endo*. C11 also has an intense 1′/2′ cross peak and is described by C2′-*endo* although overlap in the data precludes identification of its 3′/4′ cross peak.

Backbone Torsion Angles. As the $^{31}P-^{1}H$ HETCOR spectrum shown in Figure 7 indicates, 12 of the 16 phosphorus resonances can be correlated to their own and 3′ ribose. The multiplet structure of the cross peaks to the H3′, H4′, and H5′/5″ protons could be analyzed for the coupling constants, but the 10 Hz line width limits the resolution, and the measured splitting is quite similar. The chemical shifts of the ^{31}P appear to be loosely clustered around those from loop nucleotides (G8, U9, G10, and C11) and those from the stem. This clustering is also apparent in the $^{1}H-^{31}P$ heteroTOCSY spectrum (Figure 5), where four

Table 2: Chemical Shift Assignments for IRE-G10 (at 20 °C)[a]

nucleotide	H6/H8	H2/H5	H1′	H2′	H3′	H4′	H5′/H5″	^{31}P	imino/amino
G1	8.20*		5.84*	4.90*	4.76*	4.56*	4.49/4.28	−1.92	11.5/*
A2	7.92*	7.10*	5.99*	4.38*	4.75*	4.48*		−1.57	
A3	7.68*	7.42*	5.87*	4.62*	4.61*		4.314		
G4	7.00		5.48	4.44	4.37			−1.83	12.58/6.4
A5	7.59	7.76	5.88	4.51	4.45				
C6	7.36	5.12	5.38	4.42	4.35			−1.54	
A7	7.89	7.35	5.78	4.51	4.38	4.3	4.13/4.04	−1.55	
G8	7.66		5.54	4.64	4.59	4.36	4.05	−1.27	no
U9	7.71	5.79	5.90	4.31	4.62	4.07		−1.27	no
G10	8.06		5.87	4.87	4.83	4.51	4.26/4.11	−1.15	10.6
C11	7.93	6.09	6.04	4.50	4.77	4.50	4.35/4.27	−1.33	no
U12	7.86	5.38	5.53	4.61	4.50	4.38	4.2/4.44	−2.01	14.16
C13	7.98	5.73	5.60	4.49	4.62	4.15		−1.50	6.89/8.3
U14	7.89	5.39	5.49	4.51					13.90
U15	8.03	5.58	5.62	4.49	4.51			−2.12	13.76
C16	7.71	5.56	5.55	4.62					6.84/8.15*

[a] no, not observed. *, more than one peak seen due to circumstances elaborated in the text.

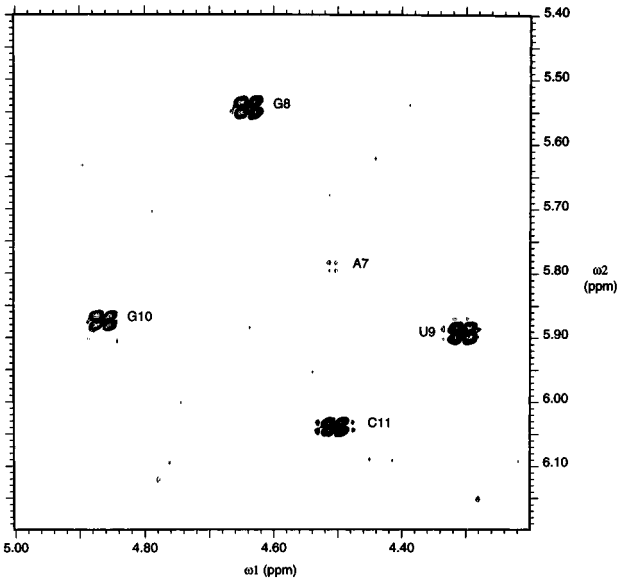

FIGURE 6: H1′/H2′ region of a DQF-COSY experiment at 20 °C (^{31}P decoupled).

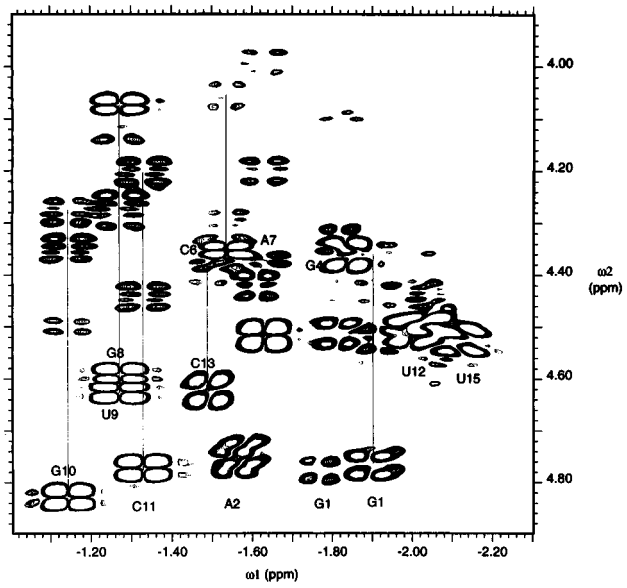

FIGURE 7: ^{31}P−^{1}H HETCOR experiment at 20 °C, with assignments indicated.

of the loop nucleotides are found together. The most unusual ^{31}P shift is assigned to G10 in the loop (−1.15 ppm). As noted by Legault and Pardi (1994), however, the phosphorus chemical shift cannot be easily correlated to the structure of the backbone; the chemical environment of the phosphorus can influence the variations in observed chemical shifts.

Loop Nucleotides. The sequential NOEs proceeded through the stem, as expected, and continued into the loop on the 5′ side, through the C6A7G8 nucleotides. In contrast, there are no NOEs observed to U9 or C11 aromatic resonances from the neighboring ribose protons; these H5/H6 cross peaks have a broad irregular shape in many spectra, especially at low contour levels. These data are interpreted to mean that these nucleotides have dynamic motion and are not part of a regular structure.

NOEs that define the loop are summarized in the representation of the RNA in Figure 8. The stem A5H8 to loop C6H5 NOE is weak; the C6H5 NOE to the following A7H8 is only seen at longer mixing times (300−400 ms) and may come from spin diffusion. These aromatic/aromatic NOEs suggest that C6 is stacked between A5 and A7. Supporting evidence comes from NOEs between A5H2 and C6H1′, as well as between A5H2 and the interstrand C13H1′, in the pattern typical of an A-form duplex. The first loop nucleotide, C6, has NOEs from its H6 to the preceding A5H1′; however, the NOE from C6H1′ to A7H8 can only be distinguished clearly from the overlapping U14 H5/H6 cross peak in the 5-deuterouridine RNA sample, and thus its intensity is in doubt. However, A7H8 has a strong NOE to C6H2′. There is a strong NOE from A7H8 to its own A7H1′, which is not typical of an N-type (C3′-*endo*) sugar; DQF-COSY data show evidence of some C2′-*endo* character of the A7 ribose. The sequential NOE walk from aromatic to sugar includes the next G8H8/A7H1′ step, but this cross peak overlaps with the U9 H5/H6 cross peak and is thus not reliably quantifiable. An NOE from G8H8 to A7H2′ is readily observable, although there are no NOEs from G8H8 to aromatic protons of A7. Again, there is a strong NOE from G8H8 to its own ribose H1′ and a relatively weaker NOE to G8H2′; the DQF-COSY spectrum shows that the G8 sugar pucker is a mixture of C2′-*endo*/C3′-*endo*.

On the 3′ side of the loop, there are NOEs observed between G10H8 and U12 H1′ and H2′, as well as from G10H8 to U12H6, providing strong evidence that G10 is stacked over the loop-closing base pair. The G10 ribose also has C2′-*endo* character. On the basis of these NOE patterns,

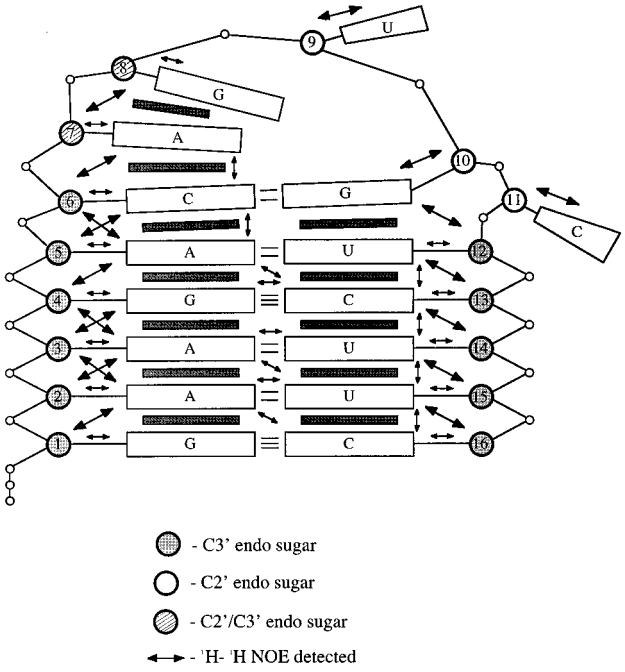

FIGURE 8: Schematic representation of the IRE-G10 hairpin structure, indicating critical NOEs observed and sugar puckers. Two hydrogen bonds are shown for the C6/G10 interaction.

and the comparison of 1D imino proton spectra of the three RNA species, we assign the broad guanosine imino proton at 10.6 ppm to G10.

The NOEs from U9H6 and C11H6 to their own ribose H1′ are the only strong NOEs observed for these bases; the wide line widths of these aromatic protons suggest that they have conformational motion.

Constraints and Structure Modeling

Descriptions of constraints given here are designed to be used with MC-SYM (Major et al., 1991; Gautheret et al., 1993), which is a constraint−satisfaction computational algorithm for modeling 3D RNA structures. This method of generating structures uses the existing experimental database of nucleotide conformations in combination with unique experimentally determined constraints. While MC-SYM has been used for modeling tRNA structures (Major et al., 1993) and RNA aptamers (Leclerc et al., 1994), it has not been used with *de novo* NMR data. The constraints required by this algorithm obviously differ from those required by distance geometry algorithms, since the starting geometries are not completely random.

The constraints given in Table 3 were used to model the IRE loop structure using MC-SYM. Distance constraints were obtained from 2D NOESY experiments at 80, 200, and 300 ms mixing times and are set very loosely. Sugar puckers were based on data from DQF-COSY experiments. The basic script for MC-SYM is given in Table 4. The first four nucleotides at the 5′ end of the RNA are allowed to sample N-type sugar puckers but restricted to describe an A-form geometry ("type A"). The final four nucleotides are also restricted to type A but have the additional constraint that they form Watson−Crick base pairs. These constraints produce an A-form helix that includes nucleotides 1−4 and 12−16. The A5U12 base pair which closes the loop was described by Watson−Crick base pairing with C3′-*endo* sugars and antiglycosidic bonds, but with variations in the

Table 3: NMR Constraints for MC-SYM

<4.5 Å constraints	<6 Å constraints	<7 Å constraints
A5 H8-G4 H2′	G4 H1′-A3 H2	G1 H8-G1 H1′
A5 H8-A5 H2′	G4 H8-G4 H2′	G4 H8-G4 H1′
C6 H6-A5 H3′	G4 H8-G4 H3′	A5 H8-G4 H1′
C6 H6-C6 H2′	A5 H8-G4 H3′	C6 H5-A5 H3′
C6 H6-C6 H3′	A5 H8-A5 H3′	C6 H5-A5 H8
A7 H8-C6 H2′	C6 H1′-A5 H2	C6 H6-A5 H1′
A7 H8-C6 H3′	C6 H6-A5 H2′	C6 H5-C6 H2′
A7 H8-A7 H2′	C6 H6-C6 H1′	C6 H5-C6 H3′
G8 H8-G8 H1′	A7 H8-A7 H1′	G8 H1′-A7 H2
G8 H8-G8 H2′	A7 H8-A7 H3′	U9 H6-U9 H1′
U9 H6-U9 H2′	G8 H8-A7 H2′	U12 H6-G10 H1′
U9 H6-U9 H3′	G8 H8-G8 H3′	U12 H6-G10 H2′
G10 H8-G10 H2′	U9 H5-U9 H4′	U12 H6-G10 H8
G10 H8-G10 H3′	U9 H6-U9 H1′	U12 H6-U12 H1′
C11 H6-C11 H1′	G10 H8-G10 H1′	C13 H6-U12 H1′
C11 H6-C11 H2′	C11 H5-C11 H2′	C13 H6-U12 H6
C13 H6-C13 H2′	C11 H6-C11 H3′	
C13 H1′-A5 H2	U12 H6-U12 H1′	
U14 H6-C13 H2′	U12 H6-U12 H2′	
U14 H6-U14 H2′	C16 H1′-A2 H2	
U15 H6-U14 H2′		
U15 H1′-A3 H2		
U15 H6-U15 H2′		
C16 H6-U16 H2′		
all H5/H6		

Table 4: MC-SYM Script

			Sequence		
			5′-helical strand		
A	rG	1	reference		type−A
A	rA	2	connect	1	type−A
A	rA	3	connect	2	type−A
A	rG	4	connect	3	type−A
A	rA	5	connect	4	sample−A
			3′-helical strand		
A	rU	12	wc	5	sample−A
A	rC	13	wc	4	type−A
A	rU	14	wc	3	type−A
A	rU	15	wc	2	type−A
A	rC	16	wc	1	type−A
			Loop from 5′ end		
A	rC	6	connect	5	full−A′
A	rA	7	connect	6	full′
A	rG	8	connect	7	full′
A	rU	9	connect	8	full−B′
A	rG	10	CG−2−Hbond	6	full−B′
A	rC	11	connect	10	full−B′

Constraint

enter here the <7 Å, < 6 Å, < 4.5 Å NMR NOE constraints

Adjacency

0 3.5

Global

P P	3.0	N1 N1	2.0
C1′ C1′	3.0	N1 C1′	2.0
PSE PSE	3.0	N1 C4	2.0
P C1′	2.5	O2′ O4′	2.0
P PSE	2.5	O3′ C5′	2.0
C1′ PSE	2.5	P C3′	2.0
C4 C4	2.0		

α, β, γ, δ, and ϵ torsion angles (sample A). Loop nucleotides were less constrained. C6, with its C3′-*endo* ribose, and NOEs to A5 and A7, was allowed to sample the complete database of available N-type sugar backbone connectivities (full A′). A7 and G8 were permitted to sample the entire database for N- and S-type sugar backbone connectivities (full′). U9, G10, and C11 have sugar puckers that ap-

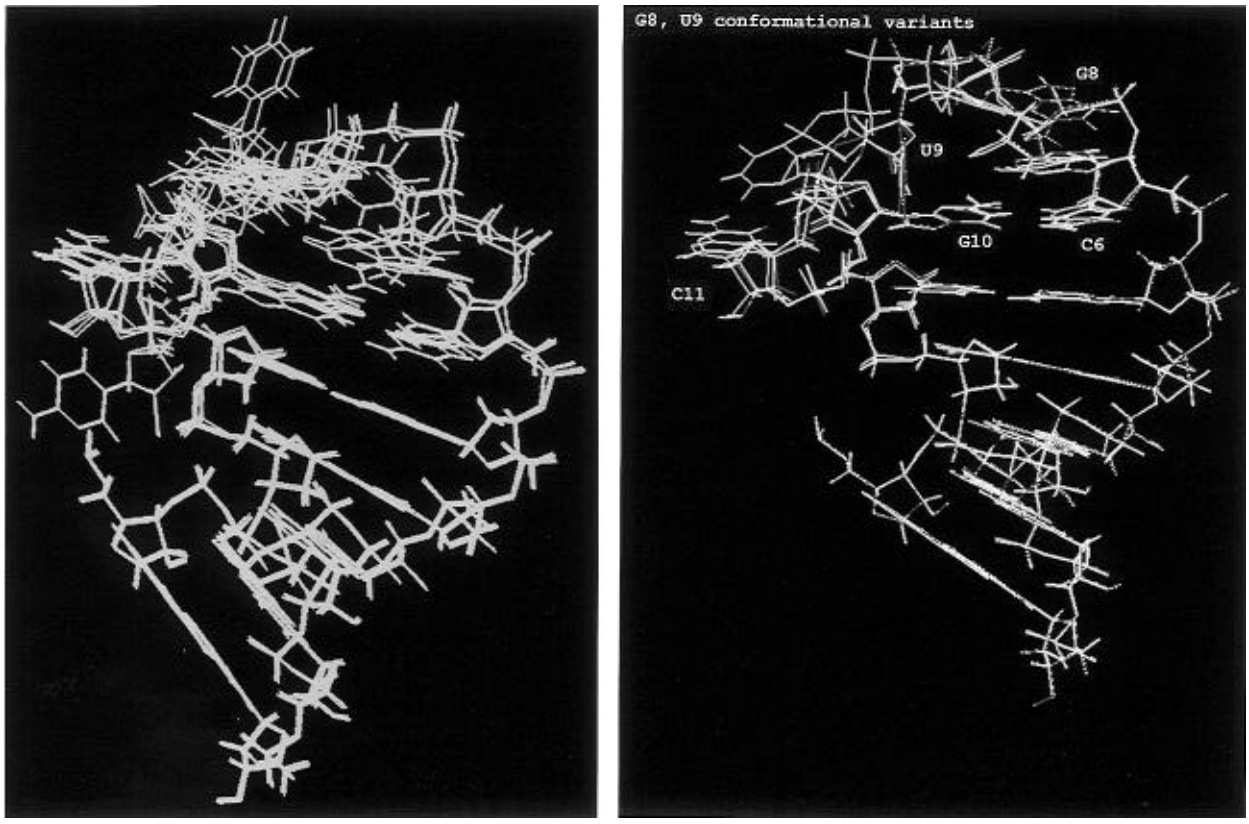

FIGURE 9: Structures of the IRE-G10 hairpin, after MC-SYM/energy minimization. (a, left) Superposition of eight structures in which C6·G10 has a Watson−Crick orientation but in which G8 is not stacked over A7. RMSD (all atom) of pairs varies from 1.3 to 2.8 Å. (b, right) Superposition of three structures in which C6·G10 is again in a Watson−Crick configuration but C6, A7, and G8 are stacked. The positions of U9 and, to a lesser extent, C11 are highly variable. RMSD varies from 0.14 to 1.8 Å.

proximate C2′-*endo*, and therefore these loop nucleotides were allowed the complete database of available S-type sugar backbone connectivities (full B′). NMR data (Table 3) were entered as a separate constraint section. Distances derived from NMR data are set much looser than used for other computational methods, to allow for the coarse sampling inherent in the MC-SYM construction. The pyrimidine H5/H6 distance, for example, is included in the <4.5 Å constraint, even though it is a fixed 2.45 Å. The adjacency section of the script contained the O3′−P bond distance and was varied in separate runs. An additional set of global constraints for minimum approach distances between atom types was used to aid in the reduction of solutions with atom overlap and to provide structures which would be suitable for energy minimization routines (Major et al., 1991).

The NMR data could not fully describe the tertiary C6/G10 interaction. In particular, NOEs to the C6 amino protons from the G10 imino proton were not observed in NOE experiments. Further experiments with [^{15}N]-cytosine may identify them, but the chemical shift of the G10 imino proton suggests that it may be hydrogen bonded to a carbonyl oxygen or perhaps only protected from exchange. Therefore, to model the loop structure, this interaction was described by "CG−2−Hbond", which samples all C·G pairings with two or more hydrogen bonds. For comparison, structures were also generated in separate MC-SYM runs with this interaction described by the looser "CG−pair" constraint, which allows for only one hydrogen bond.

In order to assess the influence of the constraints on the resulting structures, several parameters were systematically varied. Script variations included tightening or loosening the bounds of several NMR distances, altering the bounds of the adjacency (O3′−P distance), and changing loop nucleotide conformational database constraints. The NMR data are absolutely required to limit the allowed conformational space; in the extreme case, removing the NMR constraints produced 22 000 structures. Using the empirically derived script given in Table 4, with the adjacency 0−4, 126 structures were produced; with a smaller adjacency of 0−3.5, 32 structures were generated. The majority of those excluded structures were variants of the U9 position; since this nucleotide position is not fixed by NMR constraints, the reduced data set of 32 structures was selected for further analysis. The MC-SYM structures were converted to pdb files and subjected to 15 cycles of energy minimization using Insight/Discover, in the absence of constraints, to remove bad contacts. The structures were then examined for violations of the NMR constraints, and on this basis five were eliminated (they had two catenated bases produced by MC-SYM which were not separable in the minimization).

Structures could be loosely grouped into two subsets: in one family of 12 structures, the C6/G10 interaction was in a Watson−Crick orientation; in one subset of 15 structures, the C6/G10 was paired as a reversed Watson−Crick. In the loops with a Watson−Crick C6/G10 orientation, four loop conformations showed stacking of C6, A7, and G8 (Figure 9b), while in seven loops, only C6 and A7 were stacked and G8 was solvent exposed (Figure 9a). The G8 conformational heterogeneity arises through the absence of NOEs to anchor its position. The observation of a weak NOE from the G8H8 to the A7H8 at long mixing times suggests that some stacking of these bases may occur, but either the distance is greater

than 5 Å or their positions are dynamic, and therefore no constraint was included in the script. The position of U9 was highly variable in these loops, as expected from the lack of constraints. The C11 position was less variable, but the base was fully solvent accessible in all loops. In the family of structures with a reversed Watson−Crick geometry, only two structures showed stacking of C6, A7, and G8. In most of these structures, no base was stacked, and in some structures, the loop appeared crumpled and squashed.

In general, the loops with the reversed Watson−Crick interaction had higher energies (defined by the parameter set of Insight), although what this means in terms of the real structures is not clear. Eight loops with a Watson−Crick orientation were the lowest energy structures of all those calculated; the reversed Watson−Crick orientations had energies 5−10−fold higher. The two classes of structures also showed different NOE violations; in those loops with a reversed Watson−Crick orientation of C6/G10, the <7 Å constraint for G8H1′/A7H2 was violated by 2.5−2.7 Å. Of the structures with a Watson−Crick-type C6/G10 interaction, three structures in which G8 was not stacked on A7 had a similar violation. In the structures with a Watson−Crick geometry of C6/G10 with stacking of C6, A7, and G8, several had 0.24−0.3 Å violations of the A7H8/H2′ constraint.

DISCUSSION

The six nucleotides of the IRE hairpin loop, 5′CAGUGN, are conserved in mRNA 5′ UTR or 3′ UTR sequences. This loop is part of the binding site for the 98 kDa IRP. Single mutations in the loop sequence appear to result in a reduction in the IRP binding affinity (Jaffrey et al., 1993). These thermodynamic and NMR experiments show that although two nucleotides participate in a tertiary interaction across the loop, the remaining nucleotides have considerable conformational freedom.

Hairpin Stability. In predictions of RNA secondary structures based on thermodynamic parameters, loops are considered to be destabilizing; a penalty is incurred upon their formation that is larger as the loop size increases (Turner et al., 1988). In a recent investigation of the thermodynamic properties of six-nucleotide RNA loops, Serra et al. (1994) derived an algorithm for calculation of loop stability that included loop size ($\Delta G°_{37,i}$, where i = loop size), the sequence of the closing base pair, and a free energy increment for the interaction of the loop with the top of the stem ($\Delta G°_{37,MM}$). From these parameters, the free energy attributable to the loop itself can be derived.

This thermodynamic analysis can be applied to the six-nucleotide IRE loops for comparison with those loops described by Serra et al. (1994). In Table 1, the measured thermodynamic parameters for the three IRE hairpins and the IRE-R sequence are compared to the predicted values. It is clear that these hairpins are all more stable than the thermodynamic parameters predict, but the IRE-G10 and IRE-U11 hairpins are especially stable. The stem itself is calculated to have a free energy ($\Delta G°_{37}$) of −7.2 kcal/mol, based on sums of nearest neighbor interactions, and a six-nucleotide loop is given a +4.4 kcal/mol penalty (Turner et al., 1988); their sum gives the predicted hairpin stability of $\Delta G°_{37}$= −2.8 kcal/mol. Using the Serra and Turner (1995) calculation for the IRE loop sequence, $\Delta G°_{37L}$=+4.7 kcal/mol= 5.0 kcal/mol [$\Delta G°_{37,i}$(loop = 6)] + 0.4 kcal/mol [AU closing pair] − 0.7 [terminal C6C11 mismatch]. Using this value of the loop stability with the stem free energy gives the hairpin free energy $\Delta G°_{37}$ = −2.5 kcal/mol. However, experimentally, the stabilization free energy of IRE-G10 and IRE-U11 is near −5.4 kcal/mol, which is significantly more stable than predicted. The free energy of the IRE loop itself is thus calculated to be $\Delta G°_{meas}$(−5.4 kcal/mol) − $\Delta G°_{stem}$(−7.2 kcal/mol) = +1.8 kcal/mol. Both IRE-A10 and IRE-R loops are actually less stable by 1 kcal/mol [$\Delta G°_{37L}$= −4.5 − (−7.2) = 2.7 kcal/mol]. We note that the free energy of the IRE-G10 and -U11 loops is close to that of the UNCG unusually stable RNA tetraloops (Antao et al., 1991; Antao & Tinoco, 1992), which have been shown to have defined structures (Varani et al., 1991). These thermodynamic data therefore strongly suggest that the IRE loop has some tertiary interactions.

Loop Structure. MC-SYM was chosen to describe this RNA structure for several reasons. Perhaps the most important consideration is the observation that two of the six loop nucleotides have few NMR constraints on their position in the loop and probably have dynamic motion of the base and possibly of the sugar. Likewise, there are no NOEs observed from the imino proton assigned to G10 in the loop to offer a more detailed picture of its interaction. A portion of the duplex region does not adopt a stable double helix, making it difficult to resolve multiple sugar resonances and also to get accurate distances. This latter situation would be less problematic if this part of the stem is assumed to be typical A-form, as in fact the sugar puckers suggest it is. Finally, the conformations of nucleotides in this small RNA loop seemed likely to be represented in the existing experimental database of RNA conformations available to the MC-SYM algorithm, which would lead to the construction of an accurate structure.

Conformation of the loop is shown schematically in Figure 8 and modeled in Figure 9. NOEs from the C6H5 proton to both the A5H8 and A7H8 protons indicate that C6 is stacked between A5 and A7; this arrangement is present in all loop structures shown. In contrast, there are only very weak NOEs from G8H8 to A7H8 that are observed at long mixing times, making this stacking uncertain in terms of distance and dynamics. As a result of the loose constraints on its position, the conformation of G8 is not restricted in many of these structures. We note that the sugar pucker of G8 seems to be very sensitive to Mg^{2+}, and it is possible that, in other solution conditions, G8 may adopt a more defined geometry. The phosphodiester backbone must turn at U9 in order to close the loop; the downfield-shifted ^{31}P resonances and the C2′-*endo* sugar puckers of G8, U9, G10, and C11 are similar to previous descriptions of extended loop backbone conformations in situations where two nucleotides span the turn of the loop (Salemink et al., 1979; Varani & Tinoco, 1991). It is likely that U9 is swung out into solution to provide this backbone flexibility and span. Since uridine has the least propensity to stack of any RNA base, it can readily accommodate this exposed position.

The interactions of G10 with C6 somewhat restrict the loop structure. As shown in Figure 8, there are NOEs observed between the G10 aromatic H8 proton and the U12 ribose H1′ proton, indicating the intervening C11 is bulged out. The G10/C6 interaction could be a Watson−Crick or reversed Watson−Crick geometry; we favor a Watson−Crick

orientation, on the basis of the observation of weak NOEs that indicate some degree of A-form stacking of the three 5′ nucleotides. Loop structures with a reversed Watson−Crick show little or no stacking of these nucleotides, and although there are few violations of the NMR constraints, these structures are less convincing and difficult to reconcile with the thermodynamic data. In all structures, the intervening C11 is bulged out into solution, where it probably has a high degree of dynamic motion. This sixth loop nucleotide is not conserved in IRE hairpins, but on the basis of these structures, a G at this position would form a C·G pair; if loop structure or flexibility is important for recognition, then the prediction is that this variant seldom occurs.

The interaction between C6 and G10 could take one of several forms. The fast exchange of the G10 imino proton indicates that it is accessible to solvent water; its chemical shift suggests that it could be hydrogen bonded to a carbonyl oxygen. In the UUCG tetraloop, the G imino proton, which is found at 9.5 ppm (Varani et al., 1991), is apparently hydrogen bonded to the carbonyl oxygen of the first U (Allain et al., 1995); in the IRE loop, the C6 carbonyl oxygen is the analogous hydrogen bond acceptor. Alternatively, normal Watson−Crick pairing between C6 and G10 could occur, where the environment of the weakly bonded G10 imino proton determines its chemical shift. These experiments do not distinguish between these possibilities.

Thermodynamic data showed that this loop sequence is more stable than predicted. However, despite its relatively favorable free energy, several bases can adopt multiple conformations and thus presumably do not contribute significantly to loop stability. The C6/G10 interaction, and perhaps the enthalpy of stacking of C6 and A7, would seem to be mostly responsible for stabilizing this structure.

Structures of Other Six-Nucleotide RNA Loops. Several other six-nucleotide loop structures have been described. The hairpin loop of the HIV TAR element has a similar sequence, C1UGGGA6, and the resemblance between these loops was noted by Henderson et al. (1994) on the basis of their selection experiments with IRE/IRP binding and the structure mapping of the TAR element by Colvin and Garcia-Blanco (1992). In the structure described by Jaeger and Tinoco (1993), this loop was very flexible, although A6 was stacked on the closing base pair and G5 often stacked on A6. Michnicka et al. (1991) concluded, however, that the A6 was extruded from the loop structure. In a structure of this loop studied by Colvin et al. (1993) the loop-closing CG pair was swapped to GC; in that structure, also, the A6 is stacked on the stem. In none of these structures is there evidence for a C1/G5 interaction across the loop. There are apparently significant differences in the three structures, however, which may be due to solution conditions or, in the case of the Colvin et al. (1993) structure, may be due to the reversal of the loop-closing base pair.

Finally, we note that the eALAS 5′ UTR contains an IRE-G10 loop sequence in a stem closed with a U·A base pair (Dandekar et al., 1991). While the eALAS hairpin could compete with the ferritin IRE for IRP binding, its affinity to the IRP appeared to be significantly weaker than that of the ferritin IRE. By analogy to the HIV structures, the closing base pair may have altered the loop structure; this is an interesting question to pursue if the loop structure is a dominant recognition feature of the IRE·IRP association.

SUMMARY

We have shown that the six-nucleotide IRE hairpin is far more stable than predicted from thermodynamic parameters. The loop structure was described by NMR methods and modeled by MC-SYM/energy minimization and shows that the source of its stability is undoubtedly a hydrogen-bonding interaction between the conserved first (C) and fifth (G) loop nucleotides. Other features of the loop structure include stacking of the first two or three nucleotides on the 5′ side of the loop while the fourth and sixth nucleotides are unconstrained and apparently free to move.

ACKNOWLEDGMENT

We thank profusely Jeremy Williams, Michael Hodsdon, and Murad Nayal (Washington University) for their assistance in computational work. L.G.L. is also grateful for the advice of Jason Rife (Yale) and Yun-Xing Wang (Johns Hopkins).

REFERENCES

Allain, F. H. T., & Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
Antao, V. P., & Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
Antao, V. P., Lain, S. Y., & Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901−5905.
Aziz, N., & Munro, H. N. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8478−8482.
Barton, H. A., Eisenstein, R. S., Bomford, A., & Munro, H. N. (1990) *J. Biol. Chem. 265*, 7000−7008.
Basilion, J. P., Rouault, T. A., Massinople, C. M., Klausner, R. D., & Burgess, W. H. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 574−578.
Bax, A., & Davis, D. G. (1985a) *J. Magn. Reson. 65*, 207−213.
Bax, A., & Davis, D. G. (1985b) *J. Magn. Reson. 65*, 355−360.
Bax, A., & Potchapsky, S. (1992) *J. Magn. Reson. 99*, 638−643.
Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304−318.
Bettany, A. J. E., Eisenstein, R. S., & Munro, H. N. (1992) *J. Biol. Chem. 267*, 16531−16537.
Brush, C. K., Stone, M. P., & Harris, T. M. (1988) *Biochemistry 27*, 115−122.
Casey, J. L., Hentze, M. W., Koeller, D. M., Caughman, S. W., Roualt, T. A., Klausner, R. D., & Harford, J. B. (1988) *Science 240*, 924−928.
Casey, J. L., Koeller, D. M., Ramin, V., Klausner, R. D., & Harford, J. B. (1989) *EMBO J. 8*, 3693−3699.
Colvin, R. A., & Garcia-Blanco, M. A. (1992) *J. Virol. 66*, 930−935.
Colvin, R. A., White, S. W., Garcia-Blanco, M. A., & Hoffman, D. W. (1993) *Biochemistry 32*, 1105−1112.
Dandekar, T., Stripecke, R., Gray, N. K., Goossen, B., Constable, A., Johansson, H. E., & Hentze, M. W. (1991) *EMBO J. 10*, 1903−1909.
Frey, M. H., Leupin, W., Sorensen, O. W., Denny, W. A., Ernst, R. R., & Wüthrich, K. (1985) *Biopolymers 24*, 2371−2380.
Gautheret, D., Major, F., & Cedergren, R. (1993) *J. Mol. Biol. 229*, 1049−1064.
Gray, N. K., & Hentze, M. W. (1994) *EMBO J. 13*, 3882−3891.
Gray, N. K., Quick, S., Goossen, B., Constable, A., Hirling, H., Kuhn, L. C., & Hentze, M. W. (1993) *Eur. J. Biochem. 218*, 657−667.
Haile, D. J., Hentze, M. W., Rouault, T. A., Harford, J. B., & Klausner, R. D. (1989) *Mol. Cell. Biol. 9*, 5055−5061.
Harrell, C. M., McKenzie, A. R., Patino, M. M., Walden, W. E., & Theil, E. C. (1991) *Proc. Natl. Acad. Sci U.S.A. 88*, 4166−4170.
Hayatsu, H. (1976) *Prog. Nucleic Acid Res. 16*, 75−124.
Henderson, B. R., Menotti, E., Bonnard, C., & Kuhn, L. C. (1994) *J. Biol. Chem. 269*, 17481−17489.

Hentze, M. W., Caughman, S. W., Rouault, T. A., Barriocanal, J. G., Dancis, A., Harford, J. B., & Klausner, R. D. (1987) *Science 238*, 1570−1573.
Hentze, M. W., Caughman, S. W., Casey, J. L., Koeller, D. M., Rouault, T. A., Harford, J. B., & Klausner, R. D. (1988) *Gene 72*, 201−208.
Hirling, H., Henderson, B. R., & Kuhn, L. C. (1994) *EMBO J. 13*, 453−461.
Jaeger, J. A., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 12522−12530.
Jaffrey, S. R., Haile, D. J., Klausner, R. D., & Harford, J. B. (1993) *Nucleic Acids Res. 21*, 4627−4631.
Kellogg, G. W. (1992) *J. Magn. Reson. 98*, 176−182.
Kellogg, G. W., & Schweitzer, B. I. (1993) *J. Biomol. NMR 3*, 577−595.
Kikinis, Z., Eisenstein, R. A., Bettany, A. J. E., & Munro, H. N. (1995) *Nucleic Acids Res. 23*, 4190−4195.
Klausner, R. D., Rouault, T. A., & Harford, J. B. (1993) *Cell 72*, 19−28.
Kuhn, L. C., & Hentze, M. W. (1992) *J. Inorg. Biochem. 47*, 183−195.
Laing, L. G., & Draper, D. E. (1994) *J. Mol. Biol. 237*, 560−576.
Leclerc, F., Cedergren, R., & Ellington, A. E. (1994) *Nat. Struct. Biol. 1*, 293−300.
Legault, P., & Pardi, A. (1994) *J. Magn. Reson. B 103*, 82−86.
Leibold, E. A., & Munro, H. N. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2171−2175.
Leibold, E. A., Laudano, A., & Yu, Y. (1990) *Nucleic Acids Res. 18*, 1819−1824.
Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., & Cedergren, R. (1991) *Science 253*, 1255−1260.
Major, F., Gautheret, D., & Cedergren, R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9408−9412.
Michnicka, M. J., Harper, J. W., & King, G. C. (1993) *Biochemistry 32*, 395−400.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
Mullner, E. W., Neupert, B., & Kuhn, L. C. (1989) *Cell 58*, 373−382.
Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., & Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507−4513.
Philpott, C. C., Klausner, R. D., & Rouault, T. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 7321−7325.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
Rouault, T. A., Hentze, M. W., Caughman, S. W., Harford, J. B., & Klausner, R. D. (1988) *Science 241*, 1207−1210.
Salamink, P. J. M., Swarthof, T., & Hilbers, C. W. (1979) *Biochemistry 18*, 3477−3485.
Scaringe, S. A., Francklyn, C., & Usman, N. (1990) *Nucleic Acids Res. 18*, 5433−5435.
Serra, M. J., & Turner, D. H. (1995) in *Methods in Enymology*, (Johnson, M. L., & Ackers, G. K., Eds.) Vol. 259, pp 242−261, Academic Press, San Diego, CA.
Serra, M. J., Axenson, T. J., & Turner, D. H. (1994) *Biochemistry 33*, 14289−14296.
Sierzputowska-Gracz, H., McKenzie, R. A., & Theil, E. C. (1995) *Nucleic Acids Res. 23*, 146−153.
Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525−7526.
Sklenár, V., Miyashiro, H., Zon, G., Miles, Zon, H. T., Miles, T., & Bax, A. (1986) *FEBS Lett. 208*, 94−98.
Smallcombe, S. H. (1993) *J. Am. Chem. Soc. 115*, 4776−4785.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Stump, W. T., & Hall, K. B. (1993) *Nucleic Acids Res. 21*, 5480−5484.
Theil, E. C. (1994) *Biochem. J. 304*, 1−11.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biochem. Biophys. Chem. 17*, 167−192.
Usman, N., Egli, M., & Rich, A. (1992) *Nucleic Acids Res. 20*, 6695−6699.
Varani, G., & Tinoco, I., Jr. (1991) *Quart. Rev. Biophys. 24*, 479−532.
Varani, G., Cheong, C., & Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
Walden, W. E., Daniels-McQueen, S., Brown, P. H., Gaffield, L., Russell, D. A., Bielser, D., Bailey, L. C., & Thach, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9503−9507.
Wijmenga, S. S., Heus, H. A., Werten, B., Van der Marel, G. A., Van Boom, J. H., & Hilbers, C. W. (1994) *J. Magn. Reson. B 103*, 134−141.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

BI961310Q