Iron Regulatory Element and Internal Loop/Bulge Structure for Ferritin mRNA Studied by Cobalt(III) Hexammine Binding, Molecular Modeling, and NMR Spectroscopy†

Zofia Gdaniec,‡,§ Hanna Sierzputowska-Gracz,§,|| and Elizabeth C. Theil*,§

Departments of Biochemistry and Chemistry, North Carolina State University, Raleigh, North Carolina 27696-7622, and Institute for Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

Received August 11, 1997

ABSTRACT: The ferritin IRE, a highly conserved (96−99% in vertebrates) mRNA translation regulatory element in animal mRNA, was studied by molecular modeling (using MC-SYM and DOCKING) and by NMR spectroscopy. Cobalt(III) hexammine was used to model hydrated Mg^{2+}. IRE isoforms in other mRNAs regulate mRNA translation or stability; all IREs bind IRPs (iron regulatory proteins). A G·C base pair, conserved in ferritin IREs, spans an internal loop/bulge in the middle of an A-helix and, combined with a dynamic G·U base pair, formed a pocket suitable for Co(III) hexammine binding. On the basis of the effects of Co(III) hexammine on the ^{1}H NMR spectrum and results of automatic docking into the IRE model, the IRE bound Co(III) hexammine at the pocket in the major groove; Mg^{2+} may bind to the IRE at the same site on the basis of an analogy to Co(III) hexammine and on the Mg^{2+} inhibition of Cu(phen)$_2$ cleavage at the site. Distortion of the IRE helix by the internal loop/bulge near a conserved unpaired C required for IRP binding and adjacent to an IRP cross-linking site suggests a role for the pocket in ferritin IRE/IRP interactions.

RNA sequences in the noncoding region of mRNAs can regulate mRNA function. The predicted secondary structure is a hairpin distorted by a bulge, bulge loop, or internal loop. Specificity of the three-dimensional structure of RNA regulatory elements is recognized by proteins as in the Tat/TAR and Rev/RRE interactions of the HIV virus (*1*−*3*). Bulge loops and internal loops in RNA induce bends or distortions in helixes, creating specific three-dimensional structures and, often, metal binding sites (*4*). Little is known about the three-dimensional structure of natural regulatory elements in eukaryotic mRNAs.

The IRE (iron responsive element) family of isoelements is a particularly well characterized control element in normal animal mRNAs encoding proteins of iron metabolism. IREs are hairpins of 9 or 10 base pairs, interrupted by a bulge loop of 1−4 nucleotides with a conserved C residue and with a terminal hexaloop, CAGUGX (reviewed most recently in refs *5*−*8*). The metal complex Cu(phen)$_2$ binds at the internal bulge/loop (*9, 10*). All IREs recognize a family of RNA binding proteins, the IRPs (iron regulatory proteins); some IREs recognize other proteins as well, such as initiation factors (*11, 12, 14*). Single-copy IREs in the 5′-untranslated regions of mRNAs regulate ribosome binding, while pentuple-copy IREs in the 3′-untranslated regions are part of a rapid turnover element regulating mRNA stability; each type of IRE is highly conserved (96−99%) which contrasts with the lower sequence conservation (35−45%) between translation and rapid turnover IREs (*8*).

The ferritin IRE is the best characterized IRE in terms of structure and function. Assurance of the biological relevance of IRE studies with synthetic RNA, used here and in other types of experiments, has been uniquely provided by earlier investigations using natural ferritin mRNA [poly(A^+) RNA][1] to study IRE structure, the IRP binding site and IRE function in regulating protein synthesis (*9*−*11, 14*−*16*); ferritin poly(A^+) RNA showed function and/or chemical and enzymatic reactivity similar to those of the synthetic RNAs. The ferritin IRE is the most efficient of the translational regulatory IREs (*13*), possibly because of a conserved internal loop/bulge involving UGC/C rather than the bulge C of other IREs.

Previous NMR studies have focused on the role of the ferritin IRE terminal hexaloop (*17, 18*). In this study, a model of the complete IRE 30-mer is developed, assisted by NMR data from ^{15}N- and ^{13}C-labeled RNA and cobalt(III) hexammine/RNA complexes; the model is consistent with previous chemical and enzymatic studies. Co(III) hexammine significantly shifted proton NMR resonances of G7 and G27 in the internal loop/bulge region and docked in a pocket caused by distortion of the major groove in the middle of the IRE. The same region is also hypersensitive to cleavage by hydroxyl radical (*16*) and displays Mg-

† The work was supported in part by NIH Grant DK-20251.

* Corresponding author at Department of Biochemistry, North Carolina State University, Raleigh, NC 27695-7622. Phone: 919-515-5805. Fax: 919-515-5805. E-mail: Theil@bchserver.bch.ncsu.edu.

‡ Polish Academy of Sciences.

§ Department of Biochemistry, North Carolina State University.

|| Department of Chemistry, North Carolina State University.

[1] Poly(A^+) RNA from a natural cell rich in ferritin mRNA [the embryonic red cell in which ∼10% of the mRNA is ferritin mRNA (*11*)] was used with immunoprecipitation to examine control of ferritin synthesis (*11, 14, 15*) or with specific primers to examine IRE structure in the RNA after reaction with structure probes or IRP binding (*9, 16*).

S0006-2960(97)01981-8 CCC: $15.00
Published on Web 01/23/1998

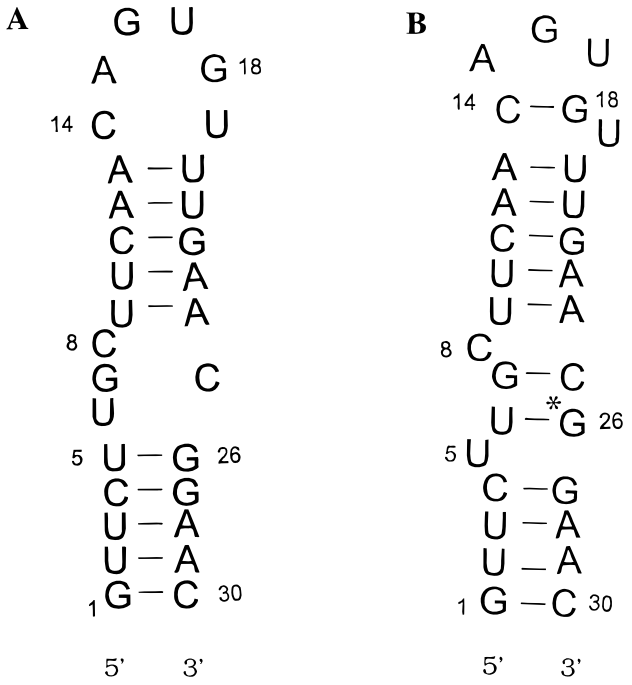

FIGURE 1: Secondary structure of the ferritin IRE. (A) The secondary structure predicted by m-fold (University of Wisconsin GCG-9.0); note the absence of the G•C base pairs across the hairpin loop and internal loop/bulge. (B) The secondary structure based on NMR spectroscopy (Figure 2). The asterisk indicates the dynamic, pH-sensitive G•U base pair (see Figure 2C).

sensitive changes in cleavage by $Cu(phen)_2$ (*9*), indicating solvent accessibility and suggesting that hydrated Mg^{2+} binds at the site. In the IRE model, G7/U6 in the internal loop/bulge and G18/U19 which cross-link to the IRP (*21*) are 22 Å apart, in contrast to only 18 Å in an IRE model without the interhelical pocket, which may relate to correct positioning in the IRP binding site.

MATERIALS AND METHODS

RNA Synthesis. The 30-mer representing the frog ferritin IRE (Figure 1) was synthesized using the double-stranded T7 polymerase site and the complement of the 30-mer as a template, as previously described (*17*); vertebrate ferritin IREs are highly conserved (*5*−*8*) (96−99%), but the frog ferritin IRE is the only one which has been studied in natural [poly(A^+)] mRNA as well as in synthetic mRNAs and RNA oligomers. The use of full-length double-stranded template increased the yield ~1.5−2-fold; reaction volumes were 24−90 mL. Cloned T7 polymerase was isolated as described by Studier et al. (*22*). RNA was purified by electrophoresis in urea/acrylamide gels as before (*17*), electroreluted, and concentrated by alcohol precipitation. To study the effect of pH on the detection of the G•U base pair (Figure 2C), commercially prepared (Cybersyn) RNA was used, but was purified by gel electrophoresis and dialyzed extensively against water before use.

RNA enriched in ^{13}C and ^{15}N was prepared using $^{13}C/^{15}N$ nucleotide triphosphates (NTPs) as described for the synthesis of RNA with natural abundance levels of the isotopes (*19*, *20*). The ^{13}C- and ^{15}N-enriched NTPs were prepared using crude rRNA from *Methylophilus methyltrophus* provided by the NIH Research Resource for Heavy Atoms at Los Alamos National Laboratory. The crude rRNA was digested with DNAse, extracted with phenol and chloroform/isoamyl alcohol (24:1), and precipitated with alcohol followed by digestion to nucleotides with nuclease P1 (*23*) and conversion to NTPs using nucleoside monophosphate kinase, guanylate kinase, pyruvate kinase, myokinase, phosphoenolpyruvate, and ATP as described by Nikonowicz et al. (*24*). After concentration, lyophilization, and alcohol precipitation, the crude NTPs were dissolved in cold 1 M triethylamine/borate buffer (TEAB) at pH 9.5 and desalted on an Affigel 601 (Biorad) column equilibrated in 1 M TEAB buffer at 5 °C (*25*); the NTPs were eluted with cold distilled water acidified to pH 4−5 with CO_2, lyophilized, dissolved in water, filtered through a washed Centricon 10 filter, and stored at pH 8.1 and −20 °C until use.

NMR Spectroscopy. RNA (0.5−1.0 mM in 10 mM sodium phosphate buffer and 0.1 mM EDTA at pH 6.8) was heated at 85 °C and slowly cooled in the NMR tube. Spectra were acquired on a Bruker DRX 500 MHz spectrometer. Spectra in H_2O were obtained either by the Watergate method (*26*) or by presaturation of the HDO signal for 2 s prior to applying an observation pulse or by using the jump−return water suppression and excitation maximum set to the imino resonances (*27*). Data for the two-dimensional (2D) NOESY experiment in 10% D_2O/90% H_2O were acquired at 12 °C using Watergate-water suppression [a 3−9−19 pulse sequence with the gradients for water suppression with excitation maximum set to the imino resonances (*26*)]. The spectrum was 2048 × 256 complex data points with a sweep width of 12 000 Hz, a mixing time of 250 ms, a recycle delay of 1.7 s, and 256 scans per slice. Spectra were processed with FELIX 95.0 software (Biosym/Molecular Simulations, Inc.) using an exponential weighing function or shifted sine-bell function to resolve overlapped imino protons.

NOESY, DQF-COSY, and TOCSY experiments were recorded in 99.996% D_2O on a 500 MHz GE Omega spectrometer or a Bruker 500 MHz spectrometer. Data sets with 2048 complex points in t_2 and 512 complex points in t_1 were acquired with 5000 Hz sweep widths in both dimensions and 128 scans per slice. NOESY spectra were acquired with mixing times of 120, 200, and 400 ms and a recycle delay of 2 s at 12 and 20 °C. The TOCSY spectrum was recorded with a 75 ms MLEV spin lock pulse and a recycle delay of 1.5 s. The DQF-COSY spectra were recorded with WALTZ decoupling of ^{31}P during acquisition and a recycle delay of 1.6 s. The diagonal and cross-peaks of DQF-COSY spectra were phased with antiphase absorption line shape in both directions. All spectra were processed with combinations of exponential and sine-skewed functions and zero-filled to 2K × 2K data points using XWINNMR Bruker or Felix 95.0 software.

Spectra with Co(III) hexammine and with varyious pHs were acquired on a Bruker DRX 500 MHz system. Imino proton spectra were obtained by the Watergate method (*26*). Typically, 2048 scans were collected. ^1H spectra in 10% D_2O/90% H_2O were collected at 12 °C in 16K point data sets consisting of 1024 scans each.

Spectra of double-labeled RNA were obtained on a Varian Unity Plus 600 MHz NMR spectrometer at the University of Chicago, Biological Sciences Division NMR Facility, used in consultation with Dr. Klaas Hallenga. The 2D ($^1H-^{15}N$) HSQC experiments were carried out using gradients during the INEPT and reverse INEPT magnetization transfer delays

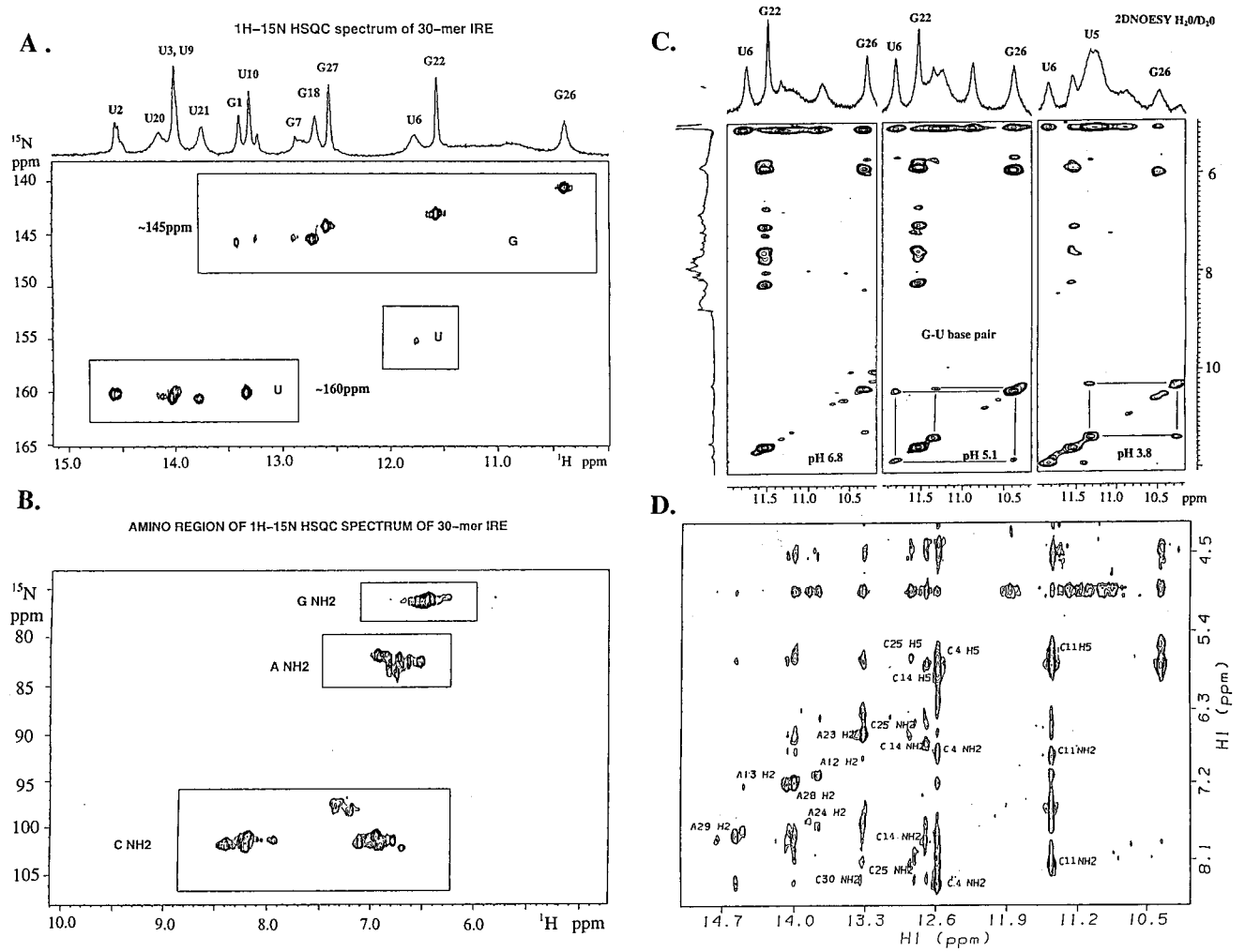

FIGURE 2: NMR data for the ferritin IRE: 2D (^{1}H−^{15}N) HSQC spectra of (A) the imino region, (B) the amino region for 2D NOESY in 90% H_2O/10% D_2O, (C) imino−imino region for three different values of pH showing enhanced detection of the G·U base pair at lower pH, and (D) the imino to aromatic and amino region. RNA (0.5−1 mM) was measured at 12 °C in 10 mM sodium phosphate and 0.1 mM EDTA at pH 6.8.

to defocus and refocus the solvent magnetization as described by Mori et al. (*28*). Experiments were optimized separately for imino (G N1H and U N3H) and amino nitrogens (C, G, and A). Data sets with 5760 complex points in t_2 and 360 complex points in t_1 were acquired. In the 2D (^{1}H−^{13}C) correlation experiment, constant time ^{13}C HMQC was used to enhance the resolution of the carbon dimension by refocusing the $^1J_{C-C}$ coupling with flip-back pulses for water elimination (*29*).

RNA samples with the native ferritin IRE sequence, containing the conserved sequence around the internal loop/bulge (Figure 1), formed opalescent gels during annealing at concentrations of >1 mM which irreversibly broadened the lines in the NMR spectrum even after diluting the sample, melting, and reannealing. Samples at concentrations below 1 mM used here and in previous studies (*17*) were clear under a variety of conditions and showed no evidence of intermolecular aggregation. Such properties, to date, have limited the types of multidimensional NMR experiments possible with the native ferritin IRE sequence.

Structure Modeling and Refinement. At the outset, the secondary structure of the RNA sequence was predicted using mfold version 9.0 from the University of Wisconsin GCG (Figure 1A). The program MC-SYM (macromolecular conformation by symbolic generation) (*30*) was used to generate a family of model RNA/IRE structures. Preliminary structures were generated using a nucleotide conformational database constructed from all available nucleic acid structures determined by X-ray crystallography and NMR spectroscopy. The INSIGHT II program (version 95.0 from Biosym/Molecular Simulations Inc.) was then used to visualize and refine the coordinates generated by MC-SYM. Initial structures from MC-SYM were refined using energy minimization with a Discover 95.0 module. Structures were examined for violations with a limited number of NMR constraints and bad contacts. The script for the model is available upon request.

Docking. The Co(III) hexammine ion was built within INSIGHT II using the Builder module and the published crystal structure (*31*). An RNA/Co(III) hexammine complex was constructed using the Docking module of the Biosym/MSI software package. Automatic docking of a ligand into the IRE molecule was performed employing an energy-driven Affinity method using the FIXEDDOCKING command. FIXEDDOCKING uses nonbond methods in DISCOVER and does not include solvation terms. The extensible systematic forcefield (ESFF) was used in all DISCOVER 3 calculations. In a DOCKING search, the maximum number

of structures was set to 10. Usually, less than 400 conjugate gradient minimization steps were needed to find the energy minimum for each acceptable structure. Hydrogen bonds were located with HBOND in MEASURE in the Viewer module of INSIGHT II.

RESULTS

Exchangeable Proton Spectra of the Ferritin IRE. In the IRE, the imino (10−15 ppm) and amino (6−9 ppm) regions of the ^{1}H NMR spectra recorded in H_2O provided information on hydrogen bonding between base pairs and mismatches. Most of the data were recorded at pH 6.8 in 10 mM sodium phosphate buffer at 12 °C after a survey of the imino proton spectra at different temperatures, pH values (6.8−3.2), and NaCl concentrations (0−150 mM) in order to find optimal conditions for making exchangeable proton assignments. Both 2D NOESY and 2D (^{1}H−^{15}N) correlated experiments were used to assign imino protons. Complete connectivities for the stem imino protons were obtained. Amino protons involved in base pairing were assigned from imino−amino correlations and ^{1}H−^{15}N HSQC spectrum optimized for the amino region.

The assignment of exchangeable proton resonances in the ferritin IRE was facilitated by 2D heteronuclear experiments on ^{15}N- and ^{13}C-labeled molecules. Resonance assignment began by identifying the directly bonded imino protons and nitrogens using the imino proton optimized 2D (^{1}H−^{15}N) HSQC spectrum shown in Figure 2. All but one uridine ^{15}N signal resonated in the region typical for A·U base pairs which was further confirmed by strong NOE interactions from uridine imino to adenosine H2 protons in the 2D NOESY spectrum.

Amino proton resonances for the IRE were assigned next using the amino proton optimized 2D (^{1}H−^{15}N) HSQC and 2D NOESY spectra. In a G·C base pair, strong intra-base pair NOEs generally arise between the G imino protons and the C amino protons. Figure 2B shows a plot of the region containing the cross-peaks for the cytidine amino groups. Separate resonances were observed for the two protons in all but one cytidine amino group. A broad cross-peak was observed for guanosine amino proton resonances in a separate region of this spectrum, due to chemical exchange-induced line broadening caused by the rotation of the amino group around the C−N bond (*19*). Adenine amino groups were not well resolved in the (^{1}H−^{15}N) HSQC spectrum because of an intermediate exchange regime.

Five cytidine amino groups showed characteristic bonded/nonbonded amino proton patterns indicative of involvement in Watson−Crick hydrogen bonds (Figure 2B,D). Three are in the upper and lower stems of the IRE (Figure 1). The only other two G·C base pairs possible from the secondary structure (Figure 1A) are across the hairpin hexaloop (G18·C14) and across the bulge loop region (G7·C25) below unpaired C8 and above U5/U6. The residue with the imino proton at 12.86 ppm had a correlation to C amino protons at 7.89 and 6.71 ppm, while the imino proton at 12.69 ppm showed NOEs to amino protons at 8.11 and 6.63 ppm, consistent with the restricted rotation of a C residue in a G·C base pair (*32*), and thus are assigned to Watson−Crick G·C base pairs. One of the imino resonances, 12.69 ppm, is assigned to a G·C pair across the hairpin hexaloop (G18·C14) by comparison of the NMR spectra of the wild type IRE sequence and a base substitution, G18A (*17*). Note that the assignment of the imino proton of G18 in the hairpin loop of the IRE 30-mer differs from that of 16-mer fragment of the IRE (*18*) in terms of chemical shift values as well as in terms of stability, judged by the broadness of the imino signal; the difference is likely due to the influence of the internal loop/bulge and the lower stem on the IRE structure. That the structure of one region of the IRE can be sensitive to structure in other parts of the IRE is illustrated by the marked changes observed throughout the IRE in the G18A substitution (*17*) and by changing the concentration of magnesium [Mg^{2+}] (*9*). Finally, then, the fifth G·C base pair in the ferritin IRE, with an imino resonance at 12.86 ppm, is G7 paired with C25 in the internal loop/bulge region (Figure 1B). The two G·C base pairs, G7·C25 and C14·G18, each of which spans an IRE loop were not predicted by the mfold program (Figure 1A).

G26, U5, and U6 are in the internal loop/bulge region of the IRE (Figure 1). Imino proton resonances at 10.37 and 11.87 ppm could not be unambiguously assigned to either G26 or U5 or U6 from the proton spectrum alone. However, in the ^{13}C- and ^{15}N-enriched sample, the U and G imino nitrogens have distinctive chemical shifts and are readily identified by the nitrogen chemical shift. A G·U base pair can be identified by the ^{15}N chemical shifts (140.47 ppm for G and 155.3 ppm for U) in Figure 2A; note that the signal at 155.3 ppm has a very low intensity and could be observed only at the low contour plot level. The presence of the G·U base pair deduced from the ^{15}N chemical shifts is supported by the observation of a cross-peak characteristic of a G·U base pair in a 2D NOESY spectrum recorded in H_2O/D_2O at 12 °C, varying the value of the pH (Figure 2C). The G·U cross-peak intensity was very sensitive to pH (Figure 2C), becoming stronger as the pH decreased from 6.8 to 5.1. Two cross-peaks characteristic of G·U base pairs were observed in the 2D NOESY spectrum (Figure 2C), indicating the occurrence of two conformers, one with a G26·U5 pair and one with a G26·U6 pair. The distribution of conformers varied with pH (Figure 2C) with the G26·U6 pair predominating at higher values of pH. The properties of the G·U base pair shown in Figure 2 suggest that the IRE in the region of the bulge/loop region of the G·U base pair is dynamic, a conclusion supported by enhanced solvent exchange (*17*) susceptibility to radical cleavage [Fe-EDTA/hydrogen peroxide and $Cu(phen)_2$/dioxygen] (*9, 16*).

Model of the Ferritin IRE. MC-SYM version 1.3 was used to generate a folded starting structure of the ferritin IRE using the A type helix for the stem nucleotides G1−C4, G27−C30, U9−A13, and U20−A24, an assumption supported by the NMR data shown in Figure 2. Watson−Crick base-paired helical nucleotides were assigned a single conformation, that of a nucleotide in an idealized A-helix. Multiple conformations were allowed for the internal loop/bulge region and hexaloop nucleotides. Satisfactory distance constraints for modeling the A-helical part of the IRE were obtained from 2D NOESY experiments in H_2O (Figure 2D) and D_2O. Sugar puckers were based on data from DQF/COSY experiments in D_2O.

The loop regions of the IRE had fewer constraints. On the basis of the NMR data, C14·G18 and G7·C25 were defined as Watson−Crick base pairs; MC-SYM sampled 10

different conformations of Watson−Crick base pairs. To facilitate study of the internal loop/bulge, a hexaloop structure, CAGUGU, was generated with a script that, in addition to the C14•G18 pairs, forced G16 to stack over A15, in analogy to the hexaloop described for the IRE 16-mer subdomain (upper stem/hairpin hexaloop) studied by Liang and Hall (*18*). In the internal loop/bulge, C8, G7, U6, U5, and C25, a Watson−Crick base pair was included for G7•C25 in the script, on the basis of the NMR. G26•U6 was included as a base pair in the script, on the basis of the NMR data in Figure 2. Because the position of C8 could not be fixed by NMR constraints, different terms permitting a large degree of sampling were first used for this nucleotide. Structures with C8 bulged out and stacked into a helix were generated and than tested. Residue C8 is thought to be critical for recognition by the IRP since it cannot be mutated without severe loss of activity (*5*−*7*). Elimination of the bulge U6 also diminished IRP recognition (Y. Ke and E. C. Theil, unpublished observations).

The script was constructed in two stages in order to reduce the number of solutions generated by the MC-SYM script when using a large degree of sampling. First, only a 16-mer including the hexaloop (but excluding the internal loop/bulge and lower stem) was generated. Because of the limited input data, only a small number of nucleotide conformations were sampled to decrease the number of allowed solutions. Eight different hexaloop structures were generated which differed mainly in the position of U19; this position is not conserved in IREs. Cytidine was used in the 16-mer model (*18*) and bulged out, as does U19 in all the hexaloop models generated in this study. U17 was also looped out. The positions selected for A15, G16, and U17 are unlikely to affect greatly any other part of the IRE because the AGU sequence and position 19 are sites in the IRE hexaloop that project into the solution on the basis of the accessibility to both large and small structure probes in solution in both ferritin and transferrin receptor IREs (*9*, *10*, *16*, *38*, *42*).

Next, nucleotides in the hexaloop were fixed and the internal loop was modeled. The input file consisted of explicit base pairing and stacking terms for generating an A type helix (supported by the NMR data) for the stem nucleotides. For the other residues, information on stacking, non-Watson−Crick base pairs, and other conformational features served to restrict the conformational assignments to those nucleotides in the database that possess the same conformational features. In those cases where no conformational information was available, different sample conformations from the database were tested, which were compatible with known geometries of RNA structures determined by X-ray crystallography and NMR spectroscopy. The MC-SYM structures were converted into pdb format and subjected to 100 cycles of energy minimization with steepest descent and 100 cycles of minimization with conjugate gradients.

The MC-SYM script with the hexaloop fixed as described, generated 47 structures which varied in the internal loop/bulge region. The structures were grouped into five families consisting of 25, 8, 4, 5, and 5 structures. Superposition, after energy minimization, was performed for five structures representing each of the MC-SYM families (Figure 3A) and for 15 structures from the largest MC-SYM family (Figure 3B). Only the nucleotides from U3 to U20 and from A23 to A28, which bracket the internal loop/bulge, were used to create the best fit. All the conformations within one family are very similar, being mainly variants of position C8 and sugar−phosphate backbone conformers of G26 (Figure 3B) that likely reflect the dynamic features of the G26 and the U6•G26 base pair observed in NMR spectra (Figure 2) and the fact that U5 readily stacks in the models (Figure 3).

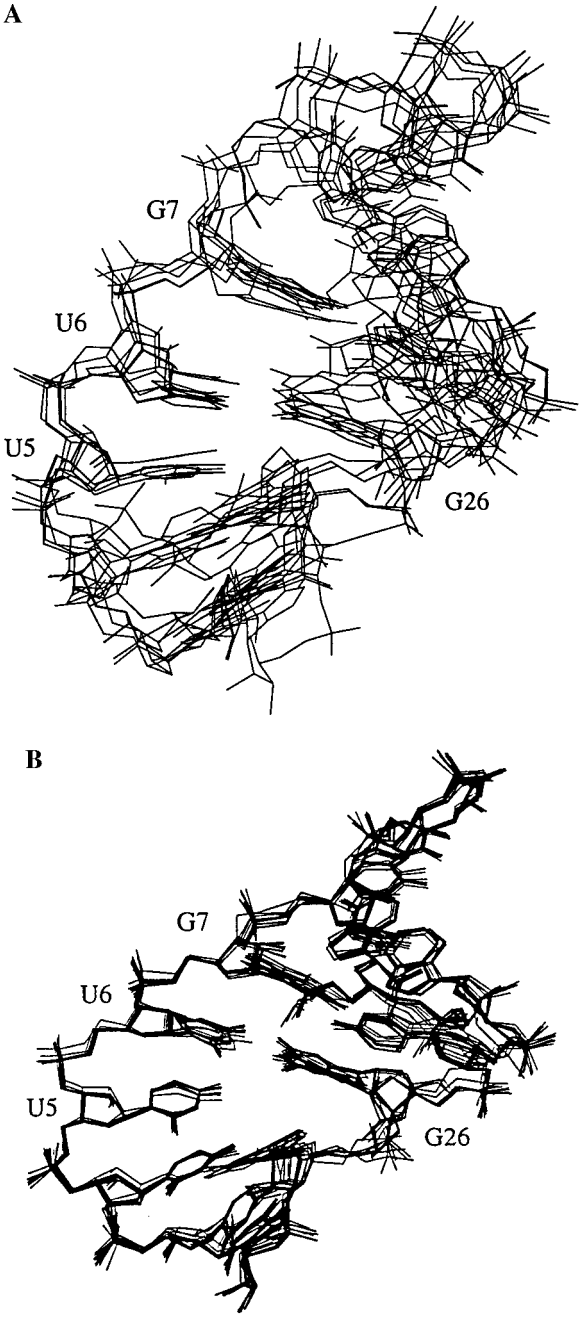

FIGURE 3: View of the internal loop/bulge region of the IRE structure. Superposition was carried out with nucleotides U3−U20 and U23−U28, which bracket the internal bulge/loop region. The view is into the minor groove: (A) five structures, one each representing the five MC-SYM families; and (B) 15 of the 25 structures in the largest MC-SYM family.

The model (Figure 4) shows the distortion of the helix imposed by the structure of the internal loop/bulge region, C8, G7, U6, and U5 and C25, G26, G27, and G7−C25. The distance between the G18 and G7 is 22 Å in the ferritin IRE model, and the major groove appears to be widened at the junction of the upper and lower helixes. In contrast, in an

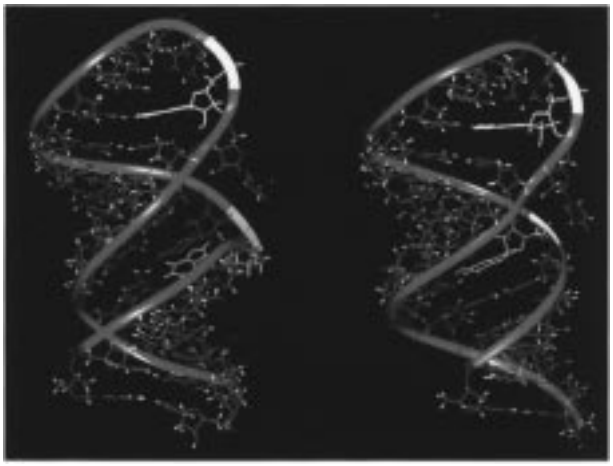

FIGURE 4: IRE models, obtained using MC-SYM, assisted by constraints obtained from NMR spectra of the IRE (Figure 2). G7 and G18, which form G•C base pairs across loops (see Figure 2), are shown in white and are at sites cross-linked to IRP-1 (*21*). C8, required for IRP binding, is shown in orange. U5 is stacked into the helix, while C8 appears to rotate freely (see Figure 3). The internal loop/bulge region, which is conserved in all ferritin IREs, changed the distance between the two IRP-1 contact sites from 18 to 22 Å, which may relate to IRP binding (*42, 43*): (left) the ferritin IRE and (right) a ferritin IRE model without the internal loop/bulge (C8 and U5 deleted).

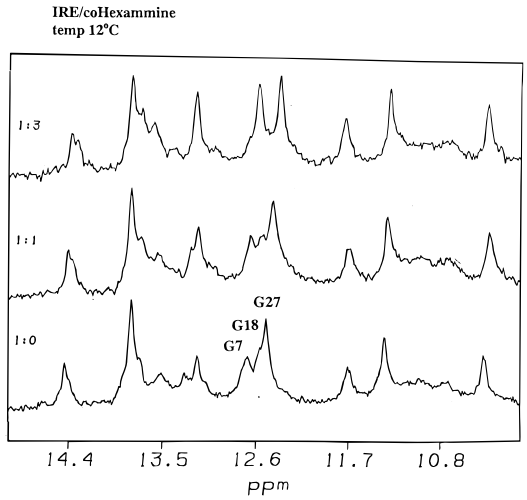

FIGURE 5: Effect of Co(III) hexammine on the ^{1}H NMR spectrum of the ferritin IRE. RNA (0.5 mM) was measured at 12 °C in 10 mM sodium phosphate and 0.1 mM EDTA at pH 6.8. Co(III) hexammine, present during melting and annealing, increased the T_m 4, 8, or 10 °C for 1:1, 2:1, and 3:1 Co:RNA, respectively: (top) 3:1 Co:RNA, (middle) 1:1 Co:RNA, and (bottom) 0:1 Co:RNA.

A-helix with the same sequence but without the bulges (deletion of C8 and U5), the distance between G18 into G7 is only 18 Å (Figure 4), and the major groove is narrow and undistorted. G18 and G7 are protein (IRP) contact sites, where intersite distances may be important for binding.

Co(III) Hexammine Effects on the IRE. To gain more insight about the loop regions of the IRE and to further test the model, one-dimensional NMR spectroscopy was used to monitor the Co(III) hexammine effects on the IRE. Co(III) hexammine can provide information about Mg^{2+} sites in RNA (*33*). Figure 5 shows the results of IRE titration with Co(III) hexammine up to a 1:3 ratio; the concentration of Co(III) hexammine was 0.5−1.5 mM. One set of imino protons of IRE and amino protons of Co(III) hexammine (3.65 ppm) observed during the titration experiment indicted that Co(III) hexammine was in fast exchange on the NMR time scale. Addition of Co(III) hexammine to the IRE had the largest observable effects on the chemical shift of two imino resonances: G7 and G27 in the internal loop/bulge region. G27/G26 is also a major site for binding and cleavage by $Cu(phen)_2$ (*9, 10*). Such an observation suggests that the metal ion bound specifically in the pocket formed by bases located in the internal loop/bulge region (Figure 4) near the dynamic G•U base pair. The T_m of the IRE with Co(III) hexammine was increased 4, 8, and 10 °C for 1:1, 2:1, and 3:1 metal:RNA, respectively (Y. Ke and E. C. Theil, unpublished results).

In the ferritin IRE, N7 and O6 atoms of guanine and the carbonyl O4 group of uracil residues are potential acceptor sites for hydrogen bonding with Co(III) hexammine ion. To determine how the proposed model structure of the ferritin IRE fit the NMR results of Co(III) hexammine titration, the intermolecular association between IRE and Co(III) hexammine was examined. A model of the RNA/metal ligand complex was built by automatic docking of a flexible ligand into the model structure of RNA. The representative structure (Figure 4) of the IRE was fixed in its conformation during the automatic docking process. Such an approach is consistent with recent NMR studies on the structure of the P5b stem from a group I intron ribozyme which bound the Co(III) hexammine ion and where the structures of RNA with and without Co(III) hexammine bound were found to be identical within the precision of the NMR data (*33*). Superposition of 10 structures of the Co(III) hexammine complex generated by the DOCKING program using an Affinity option is presented in Figure 6.

The metal ion binds to the dynamic internal loop/bulge region which forms a suitable binding pocket for Co(III) hexammine and contains four nucleotides conserved in all ferritin IREs. All structures of the generated RNA/metal complexes can be grouped into two classes. In one class, a family of hydrogen bonds, located using the HBOND option in InsightII, are found in a complicated interaction with G27, G26, U5, U6, and C25; in the second class, additional hydrogen bonds are found to G7. Positioning C8, in or outside the bulge loop, had no effect on the model. All the hydrogen bonds in the complexes are to base nitrogen and oxygen; no interactions with the backbone were observed in the model. The Co(III) hexammine:IRE ratio (1:1) increases the T_m by 4 °C (Y. Ke and E. C. Theil, unpublished observations). The effect of Co(III) hexammine on T_m, combined with the effect of Co(III) hexammine on the NMR spectrum and the model, suggests that the Co(III) hexammine interaction at the dynamic G•U base pair selects and stabilizes one of the conformations around G26 that are possible in the model (Figure 3).

DISCUSSION

A model of the natural IRE in ferritin mRNA, generated by a combination of molecular modeling (MC-SYM) and NMR studies (Figures 2 and 3), is a stack of two A-helixes of 5 base pairs each, distorted by an internal loop/bulge region with a terminal hairpin loop. Earlier NMR studies of the IRE (*17, 18*) considered only the terminal hexaloop and obtained evidence for the G•C base pair (*17, 18*) also

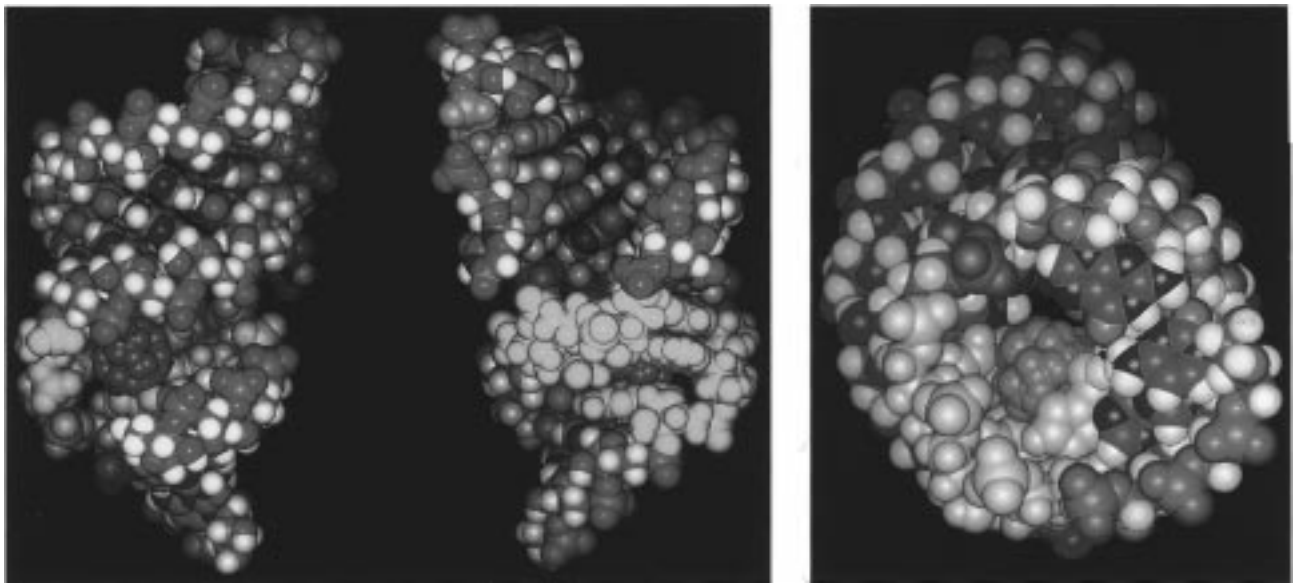

FIGURE 6: Structure of the IRE/Co(III) hexammine complex. Docking was used to examine the interaction of Co(III) hexammine (blue) with the IRE. Note the specific interactions with G7, G27, G26, U5, U6, and C25 (yellow) in the pocket of the internal loop/bulge region. ^{1}H resonances of G7 and G27 on the outer edge of the pocket were the most affected by Co(III) hexammine binding, based on the NMR spectra (Figure 5). G18 in the terminal hexaloop is shown in orange: (A, left structure in left panel) view into the major groove and (right, structure in left panel) view into the minor groove and (B) view from the bottom (5′- and 3′-termini).

observed here (Figure 2). Solution studies of the ferritin IRE using RNAses, dimethyl sulfate, transition metals complexes that are nucleases, and SELEX of variants of the hairpin loop with an IRP-1 ligand (*9, 10, 16, 34*−*38*) are all consistent with the model. Both the internal loop/bulge and the hairpin loop are spanned by a G·C base pair (G7·C25 and C14·G18, respectively) (Figure 1B). The G·C base pairs observed in the NMR spectra (Figure 2) are between bases conserved in all ferritin IREs, but were not predicted (Figure 1A); the G·C pair in the terminal hexaloop had also been indicated in earlier chemical and mutagenesis studies (*9, 10, 16*−*18, 34*−*38*), but before this study there was no experimental evidence for the G·C pair in the internal loop/bulge. On the basis of the imino resonances (12.69 and 12.86 ppm), the relatively high accessibility to D_2O, and relatively low temperature stability (*17*), the two G·C base pairs are in similar environments. Moreover, each base pair is adjacent to a dynamic region of the IRE (U6, U5, C8, or U19) and protein contact sites (*21*), suggesting that a combination of structural specificity and flexibility is important for the RNA/protein interaction.

Residue C8, which does not participate in a Watson−Crick base pair and is conserved in all IREs (*5, 6, 8*), could not be fixed by NMR constraints. Free rotation of the amino group of cytosine residues is limited only to the unpaired form of the base and is represented by the single cytidine amino group in the 2D (^{1}H−^{15}N) NMR data (Figure 2B). Thus, the flexibility of C8, and consequent free rotation of the amino group, may be related to a requirement of water-mediated protein complexation (*32*).

Co(III) hexammine binds to the ferritin IRE where the major groove is distorted by the internal loop/bulge of the ferritin IRE, based on the effect on the 1D NMR spectrum (Figure 5). Modeling the Co(III) hexammine/RNA interaction (Figure 6) provided a more detailed view of the internal loop/bulge region and the Co(III) hexammine site than is possible from the NMR data alone. The interaction of the IRE with Co(III) hexammine appears to be analogous to the binding of Co(III) hexammine in the major groove of the P5b stem of the group 1 intron of Tetrahymena (*33*). In the ferritin IRE, the Co(III) hexammine site is located in the major groove in a pocket formed by a G·U base pair and adjacent G and U residues; in the model, hydrogen bonds between Co(III) hexammine and U5, U6, G7, G26, and G27 are predicted. $Cu(phen)_2$ appears to bind in or near the Co(III) hexammine site of the ferritin IRE on the basis of cleavage of G7, G26, and G27 in both natural ferritin mRNA [poly(A^+)] and in vitro transcripts (*9, 35, 36*). $Cu(phen)_2$ recognizes distortions in RNA helixes associated with bulge loops (*39, 40*).

Hydrated Mg^{2+} and Co(III) hexammine can bind at similar sites in RNA, often involving G residues, as illustrated by the group I intron of Tetrahymena (*33, 41*). By analogy, the Co(III) hexammine site in the internal loop/bulge of the ferritin IRE could also be a Mg^{2+} binding site. The fact that Mg^{2+} decreased the accessibility of $Cu(phen)_2$ to G26 and G27 in the internal loop/bulge region (*9*) suggests competition between Mg^{2+} and $Cu(phen)_2$ for the same IRE site and strengthens the idea that the internal loop/bulge region of the ferritin IRE is a Mg^{2+} binding site. Since the IRE forms a fairly specific structure without Mg^{2+} or Co(III) hexammine (Figures 2 and 4) at low salt concentrations, finding conditions where the effect of metal or salt on IRE function can be studied independently of effects on protein is difficult, except by mutation (*38*; reviewed in refs *5*−*8*). The relatively stable structure of the IRE without metal, except for the internal loop/bulge, suggests that the metal will be involved in fine-tuning interactions in an initial RNA/protein complex, especially in the region of the internal loop/bulge, or may facilitate protein dissociation.

Hydrogen bonds between Co(III) hexammine and the IRE molecule involving five of the six NH^{3+} groups and U5, U6, G26, and G27 are possible. Such a hydrogen bond network could stabilize the dynamic features of base pairing exempli-

fied in Figure 2 and could account for the increase in the T_m when Co(III) hexammine binds (Y. Ke and E. C. Theil, unpublished observations). The interhelical region of IREs appears to be different in other IREs, based on the predicted structure which has only the C8 bulge. However, the IREs in erythroid aminolevulinate and *m*-aconitase mRNAs have additional bulged residues near the C bulge which may produce that same dynamic behavior of the interhelical region observed with the ferritin IRE (Figure 2C). The transferrin receptor IREs, which have a bulge C and no other predicted bulge regions, nevertheless have a broad melting transition, multiple cleavage sites [$Cu(phen)_2$] (Ke and E. C. Theil, unpublished observations), and multiple Pb^{2+} hydrolysis sites (*42*) in the IRE interhelical region, which also is the behavior of an RNA with conformational heterogeneity in the interhelical region. Taken together, the NMR data, the IRE model, and the IRE chemical reactivity suggest that a fairly commodious or dynamic metal binding pocket in the middle of the IRE helix is important for optimal ligand/protein interactions.

NOTE ADDED IN PROOF

A paper describing the structure of a consensus B(bulge)-IRE appeared recently (*43*). Many of the features are shared except in the region of the IL/B (internal loop/bulge); regulation by the IL/B-IREs in vivo is more efficient than that by B-IREs.

REFERENCES

1. Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J., and Feigon, J. (1994) *Biochemistry 33*, 5357−5366.
2. Puglisi, J. D., Chen, L., Blanchard, S., and Frankel, A. D. (1995) *Science 270*, 1200−1203.
3. Brodsky, A. S., and Williamson, J. R. (1997) *J. Mol. Biol. 267*, 624−639.
4. Luebke, K. J., and Tinoco, I., Jr. (1996) *Biochemistry 35*, 11677−11684.
5. Roault, T. A., and Klausner, R. D. (1996) *J. Biol. Inorg. Chem. 1*, 494−499.
6. Hentze, M. W., and Kuhn, L. C. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8175−8182.
7. Goessling, L. S., Rup, D., and Thach, R. E. (1996) *Prog. Nucleic Acids Res. Mol. Biol. 55*, 121−134.
8. Theil, E. C. (1997) in *Metal Ions in Biological Systems* (Sigel, H., and Sigel, A., Eds.) (in press).
9. Wang, Y.-H., Sczekan, S. R., and Theil, E. C. (1990) *Nucleic Acids Res. 18*, 4463−4468.
10. Wang, Y.-H., Lin, P.-N., Sczekan, S. R., McKenzie, A. R., and Theil, E. C. (1991) *Biol. Met. 4*, 56−61.
11. Shull, G. E., and Theil, E. C. (1982) *J. Biol. Chem. 257*, 14187−14191.
12. Dix, D. J., Lin, P.-N., Kimata, Y., and Theil, E. C. (1992) *Biochemistry 31*, 2818−2822.
13. Bhasker, C. R., Burgiel, G., Neupert, B., Emery-Goodman, A., Kuhn, L. C., and May, B. K. (1993) *J. Biol. Chem. 268*, 12699−12705.
14. Shull, G. E., and Theil, E. C. (1983) *J. Biol. Chem. 258*, 7921−7923.
15. Dickey, L. F., Wang, Y.-H., Shull, G. E., Wortman, I. A., III, and Theil, E. C. (1988) *J. Biol. Chem. 263*, 3071−3074.
16. Harrell, C. M., McKenzie, A. R., Patino, M. M., Walden, W. E., and Theil, E. C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4166−4170.
17. Sierzputowska-Gracz, H., McKenzie, A. R., and Theil, E. C. (1995) *Nucleic Acids Res. 23*, 146−153.
18. Liang, L. G., and Hall, K. B. (1996) *Biochemistry 35*, 13586−13596.
19. Varani, G., Aboul-ela, F., and Allain, H.-T. (1996) *Prog. Nucl. Magn. Reson. Spectrosc. 29*, 51−127.
20. Pardi, A. (1995) *Methods Enzymol. 261*, 350−386.
21. Basilion, P. J., Rouault, T. A., Massinople, M. C., Klausner, R. D., and Burgess, H. W. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 574−578.
22. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and DuBendary, J. W. (1990) *Methods Enzymol. 185*, 60−88.
23. Hayne, S. L., and Whitesides, G. M. (1990) *Appl. Biochem. Biotechnol. 23*, 205−222.
24. Nikonowicz, E. P., Sirr, A., Legault, P., Jackson, F. M., Baer, L. M., and Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507−4513.
25. Batay, R. T., Foruda, M., Kujawinski, E., Puglisi, J. D., and Williams, J. (1992) *Nucleic Acids Res. 20*, 4515−4523.
26. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.
27. Sklenar, V., and Bax, A. (1987) *J. Magn. Reson. 74*, 469−479.
28. Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. M. (1995) *J. Magn. Reson., Ser. B 108*, 94−98.
29. Hallenga, K., and Lippens, G. (1995) *J. Biomol. NMR 5*, 59−66.
30. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. (1991) *Science 253*, 1255−1260.
31. Clegg, W. (1985) *Acta Crystallogr. C 41*, 1164−1166.
32. Williams, L. D., Williams, N. G., and Shaw, B. R. (1990) *J. Am. Chem. Soc. 112*, 829−832.
33. Kieft, J. J., and Tinoco, I., Jr. (1997) *Structure 5*, 713−721.
34. Bettany, A. J. E., Eisenstein, R. S., and Munro, H. M. (1992) *J. Biol. Chem. 267*, 16531−16537.
35. Dix, D. J., Lin, P.-N., McKenzie, A. R., Walden, W. E., and Theil, E. C. (1993) *J. Mol. Biol. 231*, 230−240.
36. Henderson, B. R., Menotti, E., Bonnard, C., and Kuhn, L. C. (1994) *J. Biol. Chem. 269*, 17481−17489.
37. Butt, J., Kim, H.-Y., Basilion, J. P., Cohen, S., Iwai, K., Philpott, C. C., Altschul, S., Klausner, R. D., and Rouault, T. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 4345−4349.
38. Thorp, H. H., McKenzie, R. A., Lin, P.-N., Walden, W. E., and Theil, E. C. (1996) *Inorg. Chem. 35*, 2773−2779.
39. Sigman, D. S. (1990) *Biochemistry 29*, 9097−9105.
40. Hermann, T. H., and Heumann, H. (1995) *RNA 1*, 1009−1017.
41. Cate, J. H., Gooding, A. R., Powell, E., Zhou, K., Golden, B. L., Kuindrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1686.
42. Schlegel, J., Gegout, V., Schlager, B., Hentze, M. W., Westhof, E., Ehresmann, C., Ehresmann, B., and Romby, P. (1997) *RNA 3*, 1159−1172.
43. Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A., and Pardi, A. (1997) *J. Mol. Biol. 274*, 72−83.BI9719814

BI9719814