Interaction of RNA Hairpins with the Human U1A N-Terminal RNA Binding Domain†

K. B. Hall

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

Received February 28, 1994; Revised Manuscript Received May 24, 1994®

ABSTRACT: The isolated 102 amino acid N-terminal RNA binding domain (RBD) of the human U1A protein specifically interacts with a short RNA hairpin containing the U1 snRNA stem/loop II sequence. This recognition is nucleotide-specific, for substitutions of critical nucleotides in the RNA loop decrease binding affinity up to 10^6-fold, as measured by nitrocellulose filter binding experiments. The magnitude of the loss of binding free energy with single-nucleotide substitution in the conserved GCA sequence suggests that the interaction between the RBD and RNA occurs through a number of interdependent specific contacts in the complex. ^{13}C and ^{15}N NMR experiments, using isotopically-labeled RNA together with unlabeled protein, show that the chemical shifts of many protons from the bound RNA are substantially different from those of the free RNA, especially in the loop region of the hairpin. All these data suggest that there is a conformational change in the RNA upon formation of the RBD–RNA complex.

Many proteins that bind to RNA contain a distinctive motif, variously called the RNA recognition motif (RRM), RNA binding domain (RBD), or ribonucleoprotein motif (RNP) (Bandziulis et al., 1989). This motif is small, approximately 90 amino acids, and is distinguished by the presence of 2 distinctive sequences, RNP1 and RNP2, an octamer and hexamer sequence, respectively, that are conserved both in sequence and in position within the RBD (Dreyfuss et al., 1988; Bandziulis et al., 1989; Mattaj, 1989; Query et al., 1989; Kenan et al., 1991). This RBD can be found in a protein as a single domain, such as in SC35, U1 70K, or hnRNP C, or in tandem arrangements, such as in U1A, ASF, U2B″, and hnRNP A1 which have two RBDs, U2AF which has three, and poly(A) binding protein I with four [see Keene and Query (1991) and Dreyfuss et al. (1993) for reviews of RBD proteins]. For many of these proteins, the RBD has been identified as the site of interaction with the RNA substrate.

The structure of two RBDs is known: the human U1A N-terminal domain from X-ray crystallography (Nagai et al., 1990) and NMR (Hoffman et al., 1992), and the hnRNP C protein from NMR (Wittekind et al., 1992). Both contain the $\beta\alpha\beta$–$\beta\alpha\beta$ fold predicted for RBD sequences (Ghetti et al., 1989, 1990) with a β sheet formed from four antiparallel β strands. The adjacent strands $\beta3$ and $\beta1$ contain the RNP1 and RNP2 sequences, respectively. Two α helices are positioned on one side of the β sheet. In NMR experiments with hnRNP C, interaction with rU_8 has been shown to perturb the chemical shifts of amino acids in the β sheet and at the C- and N-termini, suggesting that these regions are the sites of contact with the RNA (Gorlach et al., 1992). For the U1A RBD, the interacting surface of the protein has not been clearly defined, although several sequence-swap experiments with the closely related protein U2B″ have identified the $\beta2$ strand and the loop between $\beta3$ and $\beta2$ as critical for RNA binding specificity (Scherly et al., 1989; Bentley & Keene, 1991).

The U1A protein is part of the U1 snRNP, which is one of the snRNPs that participate in pre-mRNA splicing. The protein binds to stem/loop II of the U1 snRNA (Scherly et al., 1989), where its function in the snRNP is unknown; it may be a scaffolding protein, to restrict the conformational variability of the snRNA, analogous to many of the ribosomal proteins. However, recent work (Boelens et al., 1993; Gunderson et al., 1994) has shown that U1A also binds to its own mRNA 3′ untranslated region (UTR), where it can inhibit polyadenylation and regulate its translation. The structure of the 3′ UTR consists of three loops, two of which are sites of U1A binding (van Gelder et al., 1993).

The N-terminal 102 amino acids of the human U1A protein, here designated 102A RBD, form an autonomous RDB that is capable of recognizing a short RNA hairpin that contains the sequence of the U1 snRNA stem/loop II (Scherly et al., 1989; Nagai et al., 1990; Jessen et al., 1991; Hall & Stump, 1992). Since the structure of the RDB itself is known, and its RNA target has been identified, it is a particularly attractive candidate for detailed investigation into its RNA binding properties. Many mutations in the RBD have been made, and their effect on its association with RNA has been measured (Scherly et al., 1989; Nagai et al., 1990; Jessen et al., 1991; Boelens et al., 1991). Here we take the opposite approach, which is to study the RNA in the complex, both through mutations that affect its association with the 102A RBD and by using NMR to look directly at the RNA resonances in the complex. Significant changes are observed in the chemical shift of many RNA resonances upon formation of the 102A-RNA complex, both in the bases and in the ribose sugars, which indicate that the environment of many nucleosides is altered when the protein is bound. The most substantial changes in chemical shift are localized to the RNA loop, which is the site of the phylogenetically conserved GCA sequence.

MATERIALS AND METHODS

Materials. [α-^{32}P]- and [γ-^{32}P]NTPs were purchased from NEN Research Products and $^{15}NH_4Cl$ and [U-^{13}C]glucose from Cambridge Isotope Laboratories. NMR 520-1B 3 mm microtubes were from Wilmad Glass. Purification of 102A and 95A RBDs was as described (Hall & Stump, 1992). T7 RNA polymerase was purified from an overproducing strain (Grodberg & Dunn, 1988), as was SP6 RNA polymerase (Jorgensen et al., 1991).

† This work is supported by the Lucille P. Markey Charitable Trust (90-47) and the NIH. K.B.H. is a Lucille P. Markey Scholar.

® Abstract published in *Advance ACS Abstracts,* August 1, 1994.

0006-2960/94/0433-10076$04.50/0

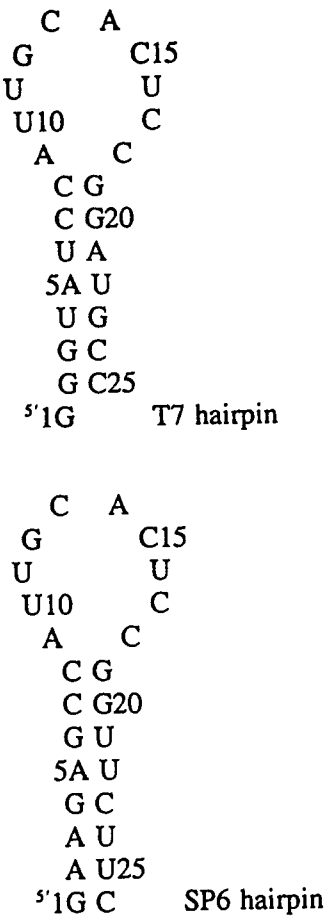

FIGURE 1: Hairpin sequences of the T7 and SP6 RNA transcripts used for these experiments, with the numbering scheme indicated.

RNA Synthesis. RNA molecules were synthesized enzymatically from short DNA oligonucleotide substrates by T7 RNA polymerase (Milligan et al., 1987) or by SP6 RNA polymerase (Stump & Hall, 1993). Figure 1 gives the sequences of the wild-type transcripts and the numbering system used to indicate substitutions. Molecules were also chemically synthesized using phosphoramidites from Milligen Biosearch or Glen Research (2'-*O*-methyladenosine). Chemically synthesized RNAs were deprotected as described (Hall & McLaughlin, 1992), except that evaporation in the presence of pyridine and toluene was eliminated. RNA was purified from 20% polyacrylamide/8 M urea gels by soaking in 3 M sodium acetate overnight for ^{32}P-labeled samples and by electroelution (Schleicher & Schuell) for unlabeled samples and NMR samples. RNAs were labeled in enzymatic syntheses with [α-^{32}P]CTP and/or [α-^{32}P]UTP for use in nitrocellulose filter binding assays or with [^{13}C]ATP or [^{15}N]-NTPs for NMR experiments (Nikonowitz et al., 1992). Chemically synthesized RNAs were labeled at the 5' end with polynucleotide kinase and [γ-^{32}P]ATP for use in binding assays. For binding experiments, the concentration of RNA was determined from the specific activity of the incorporated radiolabeled nucleotide. For melting experiments and NMR experiments, the RNA concentration was determined spectrophotometrically, using the appropriate extinction coefficient (Cantor & Tinoco, 1965; Fasman, 1975). RNAs used for thermal melting analysis were dialyzed against MilliQ water and then lyophilized. RNAs used in NMR experiments were dialyzed against buffer, and exchanged with D_2O by successive speed-vac evaporations. RNA–protein complexes for NMR were prepared by mixing at micromolar concentrations and dialyzing against 100 mM NaCl, 10 mM sodium phosphate, pH 6, and 1 mM $MgCl_2$, and concentrating at 4 °C with a 3000 MWCO Microcon filtration unit using the buffer prepared in D_2O.

Each RNA hairpin used for binding assays was also used in thermal melting experiments to ensure that it formed a monomer in solution. The absorbance vs temperature data for the hairpins were measured at 260 nm in a Gilford 250 spectrophotometer interfaced to a PC. The salt concentration varied: 250 mM NaCl/10 mM sodium cacodylate, pH 6 or 7, $\pm MgCl_2$ or 100 mM NaCl with or without $MgCl_2$, to match the conditions of binding and NMR experiments. For a monomolecular transition, there should be no concentration dependence to the melting temperature, and deviations from this observation would indicate that the RNAs were forming dimers. The concentrations of the RNAs measured varied from 10^{-3} to 10^{-7} M, using cuvettes with path lengths from 1 to 0.01 cm (NGS Precision Cells, Inc.)

Binding Assays. Nitrocellulose filter binding experiments were done at least twice for every RNA or solution condition, as described previously (Hall & Stump, 1992). Schleicher & Schuell 0.45 μm filters were presoaked at least 30 min in the binding buffer at the appropriate temperature. Bound radiolabeled RNA was quantified using a Betagen Betascope Blot Reader (Model 603), and the fraction bound was normalized to the total RNA present. The RNA bound to the filter in the absence of protein was designated as the background, and this value was subtracted from each data point. The data were fit to a Langmuir isotherm using nonlinear regression Kaleidegraph software. Retention efficiency varied with the salt concentration and temperature, and values typically ranged from 30 to 50%. Such low retention efficiencies seem to be typical of RNA–protein systems (Carey & Uhlenbeck, 1983; Romaniuk, 1985).

NMR Experiments. A Varian Unity-500 spectrometer was used for these experiments, with a Nalorac ^{1}H/^{13}C/^{15}N triple resonance microprobe. For observation of exchangeable protons, a 1331 pulse was used for water suppression in 1D experiments, and for heteronuclear multiple quantum coherence (HMQC) experiments (Mueller, 1979), a JR pulse was used for observation (Plateau & Gueron, 1982). ^{15}N/^{1}H HMQC experiments in water used $^1J_{NH}$ = 92 Hz, and a sweep width of 10 000 Hz (^{1}H) and 2000 Hz (^{15}N), with WALTZ decoupling during acquisition. For nonexchangeable protons, ^{13}C/^{1}H HMQC experiments used 3000 Hz (^{1}H) and 16 000 Hz (^{13}C), unless the ^{13}C dimension was folded, with GARP-1 ^{13}C decoupling during acquisition, and $^1J_{CH}$ = 160–180 Hz. 2D total correlation spectroscopy (TOCSY)–^{13}C-HMQC experiments (Marion et al., 1989) used MLEV17 for the TOCSY spinlock, with $^1J_{CH}$ = 160 Hz. All spectra were processed with the hypercomplex method (States et al., 1982), using Varian VNMR software. Proton chemical shifts were referenced to water at 4.8 ppm, ^{13}C to external TMS, and ^{15}N to NH_4Cl.

Complexes prepared for NMR contained 10% excess protein, to ensure that all RNA was bound. Under conditions of the NMR experiments (100 mM NaCl, 10 mM sodium phosphate, pH 6, and 1 mM $MgCl_2$), all the RNA is bound, given the subnanomolar K_D (Hall & Stump, 1992).

RESULTS

Previous experiments with the U1A 102A N-terminal RBD have shown that it will bind specifically to short RNA hairpins that contain the sequence of the U1 snRNA stem/loop II (Scherly et al., 1989; Lutz-Freyermuth et al., 1990; Jessen et al., 1991; Hall & Stump, 1992). In experiments here, short RNA hairpins, in which specific nucleotide substitutions have

been introduced, are used to measure the binding affinity to the 102A RBD, using nitrocellulose filter binding. In addition, isotopically labeled RNAs are used in NMR experiments to look directly at the RNA and compare the RNA structure free and bound to 102A.

RNA Hairpin Stability. The absorbance vs temperature profiles of all RNAs were measured both to determine if they were in fact a monomeric hairpin and to observe any changes in the melting profile as a function of the substitutions. With one exception, all RNAs showed an upper melting temperature (T_M) that was independent of RNA concentration, indicating that these are monomeric species. The lower melting transition was dependent on the RNA sequence, and was generally quite broad. The exception was the wild-type T7 RNA, which at millimolar concentrations showed a significant change in the melting temperature, indicating that some dimerization or aggregation was occurring. Since the filter binding assays were done in conditions where the RNA concentration is 10^{-11} or 10^{-12} M, the RNAs are therefore certain to be monomers (hairpins), and thus the results obtained cannot be ascribed to RNA dimers which are not likely to be substrates for the protein. Melting profiles of the SP6 wild-type hairpin ± $MgCl_2$ and of the C15G and wild-type T7 RNAs are shown in Figure 2. The addition of $MgCl_2$ at this low NaCl concentration generally results in an increase in the melting temperature of the RNA hairpin, particularly noticeable in the cooperative transition centered near 67 °C. As shown in the SP6 stem RNA profile, the lower base line of the transition is not well-defined.

Most single-nucleotide substitutions in the hairpin loop did not alter the melting temperature of the stem, although the A14G hairpin had a T_M = 62 °C. The shape of the lower transition was sensitive to the loop sequence, however. The C15G substitution shows the largest change in this region of the melting curve. The substitution of C15G apparently alters the stability of the loop, as shown by the poorly defined transition near 45 °C (in 1 mM $MgCl_2$, 250 mM NaCl, and 10 mM sodium phosphate, pH 6) compared to the normal RNA sequence in Figure 2b. Other loop substitutions, G12A, C13U, and A14G, also show changes in this lower base line which are less pronounced. These data suggest that the purine/purine and pyrimidine/pyrimidine substitutions have not greatly perturbed the stability of the loop, while the C15G substitution does destabilize it.

Alterations in the 10-nucleotide hairpin loop, where it was reduced it to a 4-nucleotide GCAC loop or put in a single-stranded RNA, were reflected by thermal melting experiments that showed melting profiles significantly different from those of the normal RNA hairpin. The melting curve of the GCAC loop in 100 mM NaCl/1 mM $MgCl_2$ showed a single cooperative transition centered around 93 °C with a flat lower base line. The high melting temperature is consistent with an extended stem of 10 base pairs, as predicted for this sequence. In contrast, the U-loop RNA transition showed a steeply sloping lower base line, although the melting temperature of the stem remains near 69 °C in 100 mM NaCl/1 mM $MgCl_2$ (data not shown), showing that this loop has little structure. Melting of the single-stranded "stemless" RNA showed significant hyperchromicity due to stacking of the bases, but no sharp transition or concentration dependence, as predicted for melting of a single-stranded RNA.

These thermal melting data showed first, that the RNAs adopted the predicted conformation, a critical feature for interpretation of binding affinities. Second, these experiments showed that the hairpin is very stable. Its high melting

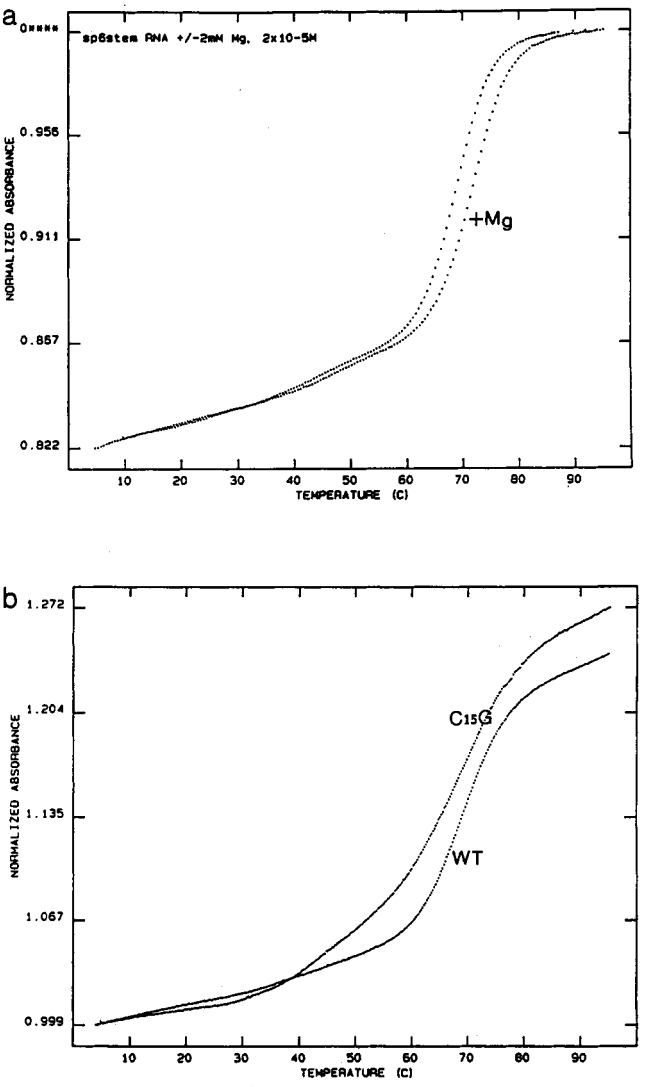

FIGURE 2: Normalized absorbance at 260 nm vs temperature for several RNA hairpins. Absorbances are normalized using the value at 90 °C. (a) "SP6 RNA" hairpin at 2×10^{-5} M, in 250 mM NaCl/10 mM sodium phosphate, pH 7, ±2 mM $MgCl_2$. RNA concentrations over a 200× concentration range showed an identical T_M. (b) A comparison of the melting profiles of two "T7 RNA" hairpins, with the wild-type loop sequence and the C15G substitution at 3×10^{-6} M, in 250 mM NaCl, 10 mM sodium phosphate, pH 7, and 1 mM $MgCl_2$.

temperature makes it unlikely that the duplex is disrupted in association with the RBD.

Role of RNA Loop Nucleotides in Complex Formation. RNA hairpins containing specific nucleotide substitutions were used in nitrocellulose filter binding experiments to measure the equilibrium constant of complex formation. The binding affinity of the 102A RBD for the wild-type RNA hairpin is very tight. In the standard assay conditions for these experiments, 250 mM NaCl, 10 mM sodium cacodylate, pH 6, and 1 mM $MgCl_2$, the dissociation constant (K_D) is typically $(4 \pm 3) \times 10^{-10}$ M for the interaction between 102A and the SP6 RNA hairpin (this value varies somewhat with different RNA and protein preparations, as well as having an error intrinsic to the assay). Several RNA variants bound the 102A RBD with comparable affinity (Table 1), including the chemically synthesized RNA with an SP6 stem sequence, the T7 stem RNA hairpin, a hairpin with a 5 base pair stem, and a chemically synthesized RNA containing 2′-*O*-methyl-A14. This modified nucleotide is normally present in human U1 snRNA but, at least in these assays, does not confer additional affinity to the association.

Table 1: Binding Affinities of RNA Substrates to 102A[a]

RNA hairpin	K_D (M)	K_A (M^{-1})	$\Delta G°$ (kcal/mol)	$\Delta\Delta G°$ [b]
SP6 stem (5′GAAGAGCC/3′CUUCUUGG)	$(4 \pm 3) \times 10^{-10}$	$(2.5 \pm 1.5) \times 10^{9}$	-12.8 ± 0.5	
T7 stem (5′GGGUAUCC/3′CCGUAGG)	$(4 \pm 3) \times 10^{-10}$	$(2.5 \pm 1.5) \times 10^{9}$	-12.8 ± 0.5	0
5 bp SP6 stem (5′GAAGC/3′CUUCG)	$(4 \pm 3) \times 10^{-10}$	$(2.5 \pm 1.5) \times 10^{9}$	-12.8 ± 0.5	0
mA14[c] (SP6 stem)	$(4 \pm 3) \times 10^{-10}$	$(2.5 \pm 1.5) \times 10^{9}$	-12.8 ± 0.5	0
C15G (T7 stem)	$(4 \pm 3) \times 10^{-9}$	$(2.5 \pm 1.5) \times 10^{8}$	-11.4 ± 0.4	1.4
dC RNA[d] (SP6 stem)	$(4 \pm 3) \times 10^{-9}$	$(2.5 \pm 1.5) \times 10^{8}$	-11.4 ± 0.4	1.4
GCAC tetraloop (T7 stem)	$(5 \pm 3) \times 10^{-7}$	$(2 \pm 1) \times 10^{6}$	-8.6 ± 0.3	4.2

[a] Binding measured in 250 mM NaCl, 10 mM sodium cacodylate, and 1 mM $MgCl_2$, pH 6, at room temperature. [b] $\Delta\Delta G° = \Delta G°$(mutant) – $\Delta G°$(wild type) for complex formation. [c] 2′-*O*-Methyladenosine. This RNA is chemically synthesized. [d] All rC are replaced with dC. This RNA is chemically synthesized.

Table 2: Binding Affinities of RNA Substrates to 102A[a]

RNA	K_D (M)	K_A (M^{-1})	$\Delta G°$ (kcal/mol)	$\Delta\Delta G°$ [b]
G12A (T7 stem)	$(2 \pm 1) \times 10^{-7}$	$(5 \pm 2) \times 10^{6}$	-9.1 ± 0.3	7.6
C13U (T7 stem)	$(7 \pm 3) \times 10^{-12}$	$(1 \pm 1) \times 10^{11}$	-15.1 ± 0.4	1.6
A14G (T7 stem)	$(3 \pm 1) \times 10^{-7}$	$(3 \pm 1) \times 10^{6}$	-8.8 ± 0.3	7.9
A9C (SP6 stem)	$(3 \pm 1) \times 10^{-9}$	$(3 \pm 1) \times 10^{8}$	-11.5 ± 0.4	5.2
G→I RNA (SP6 stem)	$(7 \pm 3) \times 10^{-9}$	$(1 \pm 1) \times 10^{8}$	-11.1 ± 0.3	5.6
U loop (SP6 stem)	$(3 \pm 1) \times 10^{-9}$	$(3 \pm 1) \times 10^{8}$	-11.5 ± 0.3	5.2
U10G hexaloop (T7 stem)	$(9 \pm 1) \times 10^{-8}$	$(1 \pm 1) \times 10^{7}$	-9.6 ± 0.3	7.1
stemless (5′GAACAGA-AUUGCACUCCGAACAU3′)	$(3 \pm 2) \times 10^{-6}$	$(3 \pm 2) \times 10^{5}$	-7.5 ± 0.3	9.2

[a] Binding measured in 150 mM NaCl, 10 mM sodium cacodylate, and 1 mM $MgCl_2$, pH 6, 25 °C. [b] $\Delta\Delta G°$ is calculated by subtracting the free energy ($\Delta G°$) of the 102A–RNA complex with SP6 wild-type RNA from that of the complex formed with the RNA variants. Since the affinity of the complex with wild-type RNA cannot be measured at this low salt concentration, the $\Delta G°$ of that complex is extrapolated from higher salt data and relative to dissociation constants of other RNA–protein complexes, to give $\Delta G° = -16.7$ kcal/mol ($K_D = 4 \times 10^{-13}$ M). The error in $\Delta\Delta G°$ should be about 20%, given this extrapolation.

Other RNAs bound to 102A with somewhat reduced affinity, at this salt concentration. RNA in which all seven rC's were substituted with dC binds with a 10X reduced affinity [$(K_D = (4 \pm 2) \times 10^{-9}$ M], suggesting that not all these ribose hydroxyls are directly involved in binding or that their loss can be readily compensated. The C15G RNA substitution has the same 10X reduction in affinity, but this loss of affinity is unlikely to result from the same energetic compensations as for the dC RNA. With this dC chimeric RNA, the balance of entropy/enthalpy compensation in complex formation is likely to reflect the absence of seven hydroxyls in the chimera which would normally have some water associated with them. The similar binding affinities for the dC and C15G RNA substrates are expected to result from different energetic contributions, but from these limited data, nothing can be said about the specific enthalpic or entropic driving forces that describe these associations.

Some of these RNA variants bound very weakly to the protein in our standard assay conditions of 250 mM NaCl/1 mM $MgCl_2$, and, therefore, to measure the dissociation constant using reasonable amounts of protein, the salt concentration was lowered to 150 mM NaCl/1 mM $MgCl_2$. These data are given in Table 2. Binding of wild-type RNA is too tight to be assayed in these conditions with this method. Several of these substitutions resulted in dramatic changes in the affinity (dissociation constant) of the complex. In particular, the dissociation constants of the 102A–RNA complexes formed with the G12A [$K_D = (2 \pm 1) \times 10^{-7}$ M] and A14G [$K_D = (3 \pm 1) \times 10^{-7}$ M] RNAs indicate that these substitutions substantially affect recognition. In contrast, the C13U substitution was far less disruptive to the complex. G12, C13, and A14 are phylogenetically conserved in U1 snRNAs, and so the effect of substitutions in this region might be anticipated to be severe. However, these results show that substitution at each nucleotide in this sequence does not perturb complex formation to the same extent, suggesting that some substitutions are tolerated more readily than others. A less severe loss of affinity is observed for the A9C substitution, at the base of the stem, which suggests that this region of the loop is also important for complex formation. These substitutions do affect the structure of the RNA loop, as observed in the lower transition of the thermal melting data. However, the substitution with the largest effect on this lower transition, C15G, is not the most disruptive for protein binding, showing that the structures adopted by these mutant loops may not necessarily be related to their ability to bind to the RBD.

One possible effect of these nucleotide substitutions on complex association may be due to steric clashes that block formation of specific contacts. The purine–purine and pyrimidine–pyrimidine substitutions are designed to minimize these effects, but the geometric arrangement of potential hydrogen bonding moieties on the bases is, of course, not preserved. However, the substitution of inosine for guanosine at all positions in the RNA is designed to only remove a potential hydrogen bond donor, since I-C base pairs do not disrupt the geometry of the helix. The absent C2 amino group would be found in the minor groove of the RNA duplex, and there are five G(I) nucleotides in the stem, while the phylogenetically conserved GCA sequence contains the only guanosine nucleotide in the RNA loop. This I-RNA binds to 102A with a nanomolar binding constant (Table 2) in the low-salt buffer, a 100X higher affinity than observed for the single RNA substitutions G12A and A14G. This loss of affinity for I-RNA cannot be ascribed to steric clashes, but rather indicates that these amino groups contribute to protein binding, and while it is tempting to assign this loss of binding free energy to the loss of specific contacts at G12, it is likely that there is a loss of some minor groove interactions as well.

Loop Structure and Affinity. The conserved sequence GCAC was maintained in three RNA loop variants. In one RNA, the 10-nucleotide loop was reduced to a GCAC 4-nucleotide loop by 3 nucleotide substitutions: 5′AGU-GCACACU. This RNA molecule has a melting temperature near 93 °C, a consequence of producing a 10 base pair stem.

This RNA is bound weakly by 102A RBD [$K_D = (5 \pm 3) \times 10^{-7}$ M in the high-salt buffer], even though it retains the phylogenetically conserved GCA sequence. The single U10G substitution results in an RNA with a hexamer loop, which is also a poor substrate for 102A binding [measured in the low-salt buffer (Table 2)]. Perhaps the RBD cannot make specific contacts with the RNA sequence in these hairpins, or the loss of other contacts with loop nucleotides or backbone leads to loss of affinity.

If closing the loop blocks protein access to bases, then making a flexible loop should facilitate access. Putting the GCAC sequence in the context of a U-loop (5′UUUGCACUUU) should provide a relatively unstructured environment. This RNA binds 30× tighter than the U10G RNA, but still far weaker than the wild-type RNA. Thus, while the conserved GCAC sequence is required for binding specificity, it is not sufficient to provide all the binding free energy of the association in the context of an unstructured loop. The nonconserved nucleotides in the loop may serve both to provide a structural context for the GCA sequence, and thus lower the entropic cost of association, and also to contribute to nonspecific interactions with the protein.

Contribution of the Stem to Complex Formation. The effect of the RNA stem on binding affinity has been measured by altering its sequence and length, and by disrupting it. In the experiments here, RNAs with the stem sequences 5′GGGUAUCC/3′CCGUAGG with 7 base pairs and a 5′ overhang (from T7 RNA polymerase transcriptions) and 5′GAAGAGCC/3′CUUCUUGG with 8 base pairs (from SP6 RNA polymerase transcriptions) bound equally well to the 102A RBD, as did the shorter 5 base pair 5′GAAGC/3′CUUCG hairpin (Table 1). Perhaps significantly, the shorter stem does not include a GU base pair, nor any unpaired bases (such as the U–U pair in the normal U1 snRNA stem), yet it binds as well as the longer stems. These results suggest that the regular arrangements of phosphates provided by the stem are enough to provide the necessary contacts between the RNA and RBD. Removal of 3 base pairs at the end of the 8 base pair stem does not reduce the binding affinity of the RNA–RBD complex; this result could be due to lack of direct contact with the RBD, but could also be explained through enthalpy/entropy compensation in the complex that occurs as a result of the loss of these base pairs.

When the stem is absent, and the RNA loop sequence is in the context of a 23-nucleotide single-stranded RNA, the RNA is essentially not a substrate for 102A binding. The single-stranded RNA binds with a micromolar dissociation constant (Table 2), which in these assays is the level of nonspecific binding. This result indicates that the RBD is unable to make the sequence-specific contacts with the loop sequences when that loop is not presented at the end of a stem. The constraint of the stem could be to present a particular loop geometry to the protein or to provide electrostatic contacts through phosphate backbone/protein interactions which position the loop for specific contacts.

Comparisons of the Free Energy of Complex Formation. Tables 1 and 2 list the free energy of complex formation ($\Delta G°$) for each RNA substrate, as well as the $\Delta\Delta G°$. The difference in the free energy of association, $\Delta\Delta G° = \Delta G°(\text{mutant}) - \Delta G°(\text{wild type})$, is a measure of how much binding free energy is lost due to the mutation(s) in the RNA. If the interaction free energy were simply the sum of each independent interaction, then taking the sum of $\Delta\Delta G°$ calculated for each single RNA mutation should equal the total free energy of the wild-type RNA interaction. However, the magnitude of a single substitution, such as G12A or A14G, indicates that this description of independent interactions is inappropriate for this system. Rather, these results suggest that the association is best described by interdependent interactions that are coupled in their contribution to formation of the complex.

The interpretation of these free energy differences is not obvious, however. For example, the observation that the G12A or A14G substitution costs the complex 7 kcal/mol of free energy can be interpreted several ways. The first is simply in terms of new structural constraints on the complex, where the substitution of these nucleotides interferes with formation of specific hydrogen bonds either through steric considerations or through loss of a critical protein–RNA contact. A second interpretation is that these nucleotide substitutions change the water or ion distribution around the RNA, and thus the entropic or enthalpic contribution to the free energy is altered, leading to a less favorable association. Especially for RNA loop structures, the effect of counterions on their structure and stability (and flexibility) is likely to contribute to their interactions with proteins. In fact, previous experiments (Hall & Stump, 1992) have shown that there is a Mg^{2+} dependence to this association, which could reflect stabilization of an unfavorable RNA loop structure. In the absence of more detailed data, it is not possible to unambiguously assign the source (thermodynamic and/or structural) of these free energy changes (and the different dissociation constants). One prediction from these data is that the pattern of entropy/enthalpy compensation in the association would be different for different RNA substrates.

Temperature Dependence of Complex Formation. Measurement of the dissociation constant (K_D) of the RNA–protein complex as a function of temperature not only can indicate the stability of the association but also can provide thermodynamic parameters of the association to identify the energetic driving forces. The van't Hoff enthalpy of association, calculated from the slope of the line in a plot of $\ln(K_A)$ vs $1/T$ ($K_A = 1/K_D$), can provide information about the driving forces for formation of the complex. The van't Hoff plot for the SP6 wild-type hairpin RNA-102A interaction in 250 mM NaCl, 10 mM sodium cacodylate, pH 6, and 1 mM $MgCl_2$ is shown in Figure 3. It is obvious that this is not a linear relation: the association constant has a maximum near 10 °C, with decreasing values at both higher and lower temperatures. The free energy of association, $\Delta G°_{obs} = -RT \ln(K_A)$, is nearly invariant with temperature, however, and indicates that the entropy and enthalpy terms are compensating with temperature. Similar behavior has been observed for some specific DNA–protein complexes (Ha et al., 1989) such as *lac* repressor and *Eco*RI endonuclease, as well as the nonspecific interaction between *Escherichia coli* SSB and oligoadenylate (T. Lohman, personal communication).

Shortening the RBD Affects Binding Affinity. The positions of the 1–5 N-terminal amino acids and the 91–95 C-terminal amino acids are disordered in the crystal structure of the 95A RBD (Nagai et al., 1990). The 102A RBD used for these experiments has seven additional amino acids at its C-terminus compared to the RBD used for crystallography, and thus this region might also be anticipated to be disordered. However, the sequence from position 95–102, Ala-Lys-Met-Lys-Gly-Thr-Phe-Val, contains two positively charged amino acids as well as a phenylalanine, all of which might potentially interact with the RNA. A comparison of the binding affinity of these two RBDs, 95A and 102A, shows that the 95A affinity is reduced by about 30X from that of 102A in 250 mM NaCl, 10 mM sodium cacodylate, pH 6, and 1 mM $MgCl_2$ (data not

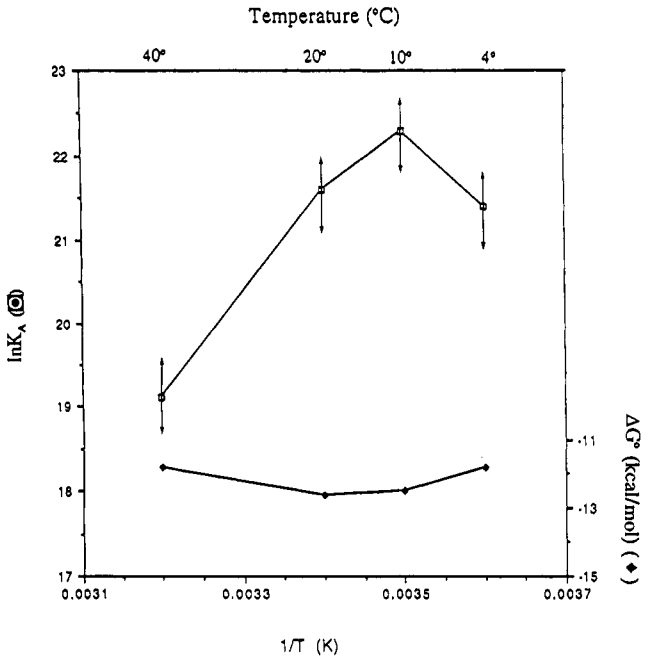

FIGURE 3: van't Hoff plot (ln K_A vs $1/T$) of the binding of the 102A RBD and the 26-nucleotide SP6 stem RNA, in 250 mM NaCl, 10 mM sodium cacodylate, pH 6, and 1 mM $MgCl_2$.

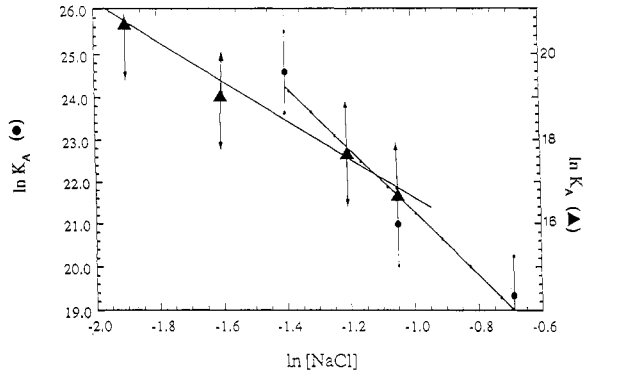

FIGURE 4: Comparison of the [NaCl] dependence of the association constants (K_A) for 102A (●) and 95A (▲) complexes. The slope of the line, $\delta \ln(K_A)/\delta \ln$ [NaCl], is proportional to the number of ions released upon complex formation. Experiments were done at room temperature, in 10 mM sodium cacodylate (pH 6)/1 mM $MgCl_2$.

shown), showing that this deleted end contributes to affinity.

The NaCl dependence of the RNA–102A interaction shows that about seven ions are released upon complex formation (Hall & Stump, 1992), which is part of the electrostatic contribution to the association. A similar analysis of the 95A–RNA interaction as a function of NaCl concentration shows that only five ions are released in this complex (see Figure 4). The slope of the line in the plot of ln K_A vs ln[NaCl] is proportional to the number of ions released: $\delta \ln(K_A)/\delta \ln [M] = -m'\psi$, where [M] is the salt concentration, m' is the number of ions released, and ψ is the fractional counterion bound per phosphate. Since the value of ψ for this RNA is undefined, we use the slope alone as an indication of the ions released. For 102A, the slope is calculated using the three values at the highest salt concentrations where the association constants are well-defined (those at lower salt concentrations are not reliable since the RNA is bound too tightly). Thus, for the 102A–RNA complex, $\delta \ln(K_A)/\delta \ln [M] = 7.4 \pm 1.6$; for 95A, the slope is -5.4 ± 0.9; both were calculated by least-squares fit of the data. Note that these experiments do not address the individual contributions of anions and cations, but measure the sum of the two as net ion release (Record et al., 1976, 1978). The truncation reduces the number of ions released, which may be related to the lower affinity of the 95A–RNA complex.

This C-terminal sequence of the RBD has been previously shown to be important in binding to RNA, since mutation of Lys98 to Gln decreased binding although Lys96Gln did not alter affinity (Nagai et al., 1990). Truncation of the domain to amino acid 91 resulted in no detectable binding by immunoprecipitation (Scherly et al., 1989) while truncation at A96 reduced the affinity (Nagai et al., 1990; Jesson et al., 1991). The difference in the NaCl dependence for formation of the two complexes in the data shown here supports the idea that the two lysines in this "arm" of the protein contribute to electrostatic interactions in the complex.

NMR Experiments with RNA and the RNA–102A Complex. The RNA hairpin in the complex can be observed directly using the hydrogen-bonded imino protons from the base pairs and by using ^{13}C and ^{15}N isotope-labeled RNA in NMR experiments that look only at these labeled nuclei and their covalently attached protons. From a comparison of the RNA bound and free, it is possible to look for differences in the RNA spectra that occur upon formation of the complex, and so potentially identify those nucleotides that are associated with the RBD as well as to look for any evidence of conformational changes that occur as a result of binding of the protein.

(A) Imino Protons. The 1D imino proton spectra of the free and bound SP6 stem RNA are shown in Figure 5. These protons are observed in experiments in H_2O, where, although they are in exchange with the bulk water, the exchange rate is slowed due to hydrogen bonding and thus those protons that are in the stem of the hairpin can be readily observed. In addition, if any imino protons from the loop are protected from exchange either by hydrogen bonding within the loop structure or by a loop conformation that excludes water, then these protons could be observed as well. There are three U's and one G in the loop that could potentially contribute an imino proton to the spectrum.

As shown in Figure 5, the protons arising from the stem region are observed in experiments with both free and bound RNA. The proton resonances are assigned using 1D NOE experiments, using selective irradiation of each proton. The assignments are given in Table 3. The G1C26 imino proton resonance (peak g at 12.3 ppm) is a doublet, due to inhomogeneity at the 3′ end of the RNA transcript (a result of *in vitro* transcription and incomplete separation during purification). In the free RNA, one extra resonance is observed at 13.8 ppm (peak c), as is a very weak resonance near 11.8 ppm. No NOEs are observed to or from peak c, which is assigned to an imino proton from an AU base pair on the basis of its chemical shift. This extra AU pair appears to come from the loop, suggesting that there is some distinctive structure there, under these conditions of 1 mM free Mg^{2+}. (In the absence of Mg^{2+}, this resonance is not observed). There are three uridines and two adenosines in the loop, but it is not clear which two are paired. The weak resonance at 11.8 ppm in this spectrum has a greater intensity in conditions of higher salt and higher $MgCl_2$ concentration. Other NMR data indicate that the loop structure is very sensitive to the $MgCl_2$ concentration, such that at higher concentrations, the loop structure becomes more defined (data not shown). It therefore seems likely that the weak 11.8 ppm resonance also comes from an imino proton in the loop. It is worth noting that at higher $MgCl_2$ concentrations, the binding affinity of the protein becomes weaker (Hall & Stump, 1992), suggesting that a

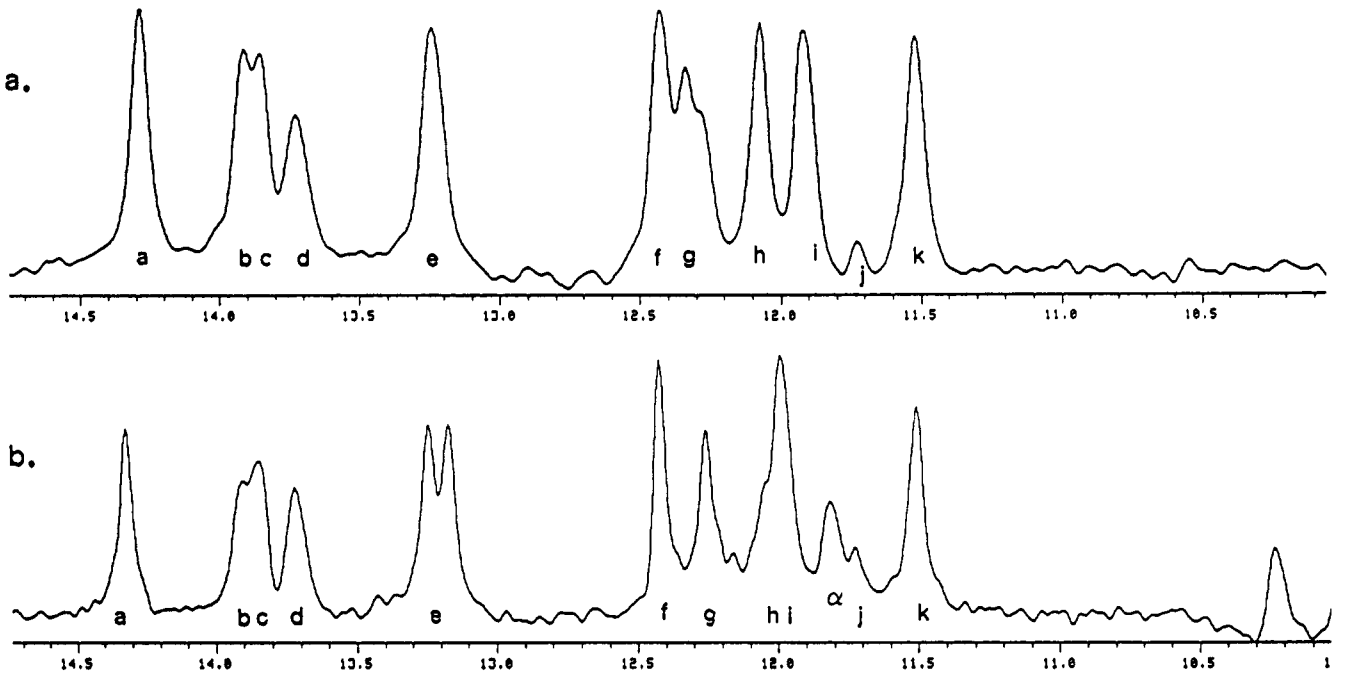

FIGURE 5: Imino proton NMR spectra of the SP6 RNA free (a) and bound to 102A (b) at 15 °C. Uniform [^{15}N]guanosine RNA concentration is 0.7 mM in (a); [RNA]:[102A] ratio = 1:1, at 0.7 mM in (b). Both samples were dialyzed against 100 mM NaCl, 10 mM sodium phosphate, pH 6, and 1 mM $MgCl_2$. Assignments shown were determined from 1D selective NOE experiments.

Table 3: Imino Proton Assignments[a]

resonance	chemical shift (ppm) free	chemical shift (ppm) bound (102A)	assignment
a	14.3	14.35	A5U22
b	13.9	13.9	A3U24
c	13.8	13.9	loop AU[b]
d	13.7	13.7	A2U25
e	13.2	13.2	G7C20
		13.15	
f	12.4	12.4	C4G23
g	12.3	12.25	G1C26
h	12.1	12.05	C8G19
i	11.9	12.0	G6U21
j	11.8	11.7	G12[b]
α		11.8	loop U
k	11.5	11.5	G6U21

[a] Peak i corresponds to the imino proton from U21, while peak k is from G6, which together form a GU pair. Peak α is present only in the complex. Peak g is a doublet due to inhomogeneity at the 3′ end of the RNA. Peak j has a very low intensity, and in the complex is only observable in HMQCJR spectra. Chemical shifts are for RNA and complex at 15 °C, in 100 mM NaCl, 10 mM sodium phosphate, pH 6, and 1 mM $MgCl_2$. [b] Tentative identification.

highly structured loop is not a good substrate for the RBD.

In the spectrum of the bound RNA, shown in Figure 5b, most of the same imino proton resonances are found as for the free RNA, in particular, those coming from the stem of the RNA. The chemical shifts of these proton resonances are similar but not identical in the two species, suggesting that some change in the helix environment or geometry has occurred. 1D NOE experiments show the same connectivities between imino protons as for the free RNA, although the NOEs are more difficult to observe. There are, however, some obvious changes in the spectrum of bound RNA relative to that of the free RNA. First, there is a new resonance (α) near 11.9 ppm in the spectrum of the bound RNA (at higher temperatures, this resonance splits into two). Peak α shows no clear NOEs to any other proton, but is observable from 4 to 30 °C, indicating that its environment is stable. Second, several resonances from adjacent base pairs in the stem have either shifted (G6/U21, k/i) or split (C7G20, peak e), suggesting that there may be an interaction with the protein at this site in the duplex. Peak *c*, assigned to a loop AU pair, is still present in the spectrum of the complex, although its chemical shift is changed. Finally, all resonances are broadened, which is probably a reflection of the increased molecular mass of the complex (now 19 kDa).

Heteronuclear ^{15}N–^{1}H NMR, using uniformly ^{15}N-labeled RNA, provides another way to observe these imino protons. The HMQCJR experiment, shown in Figure 6, compares the imino protons from [^{15}N]guanosine nucleotides in free and bound RNA (cross-peak labels correspond to Table 5 and Figure 5). The weak resonance observed near 11.8 ppm (peak j) in the imino proton spectrum of RNA gives a very weak ^{15}N-^{1}H correlation in the region of the ^{15}N spectrum corresponding to the guanosine N1 resonances (not visible at the contour level shown). The low intensity of this cross-peak is no doubt due to the exchange properties of this proton, which is in more rapid exchange than the other, hydrogen-bonded imino protons. These HMQCJR experiments do identify peak j as a resonance from a G imino proton, which is tentatively assigned to loop G12, since there is no evidence that the conformational heterogeneity in the RNA stem would account for this extra resonance. The G1C26 doublet is obvious here; again, this results from heterogeneity at the 3′ end of the RNA transcript. In addition, peak e, from C7G20, appears to be a doublet here, suggesting some heterogeneity at this end of the RNA stem.

The same HMQCJR experiment, looking at the 102A–[^{15}N]guanosine RNA complex, still shows a weak (G) imino proton resonance at 11.8 ppm, again not visible at this contour level, but which is more apparent at low temperature. Peak g, from G1C26, is again broad due to the two environments of G1. Peak e (C7G20) is now clearly a doublet, while peak h (C8G19) also shows some suggestion of being composed of two proton frequencies. The heterogeneity in these latter two resonances suggests more conformational heterogeneity or dynamics in the complex. Note that these two CG pairs are at the base of the RNA loop. The striking result is that the

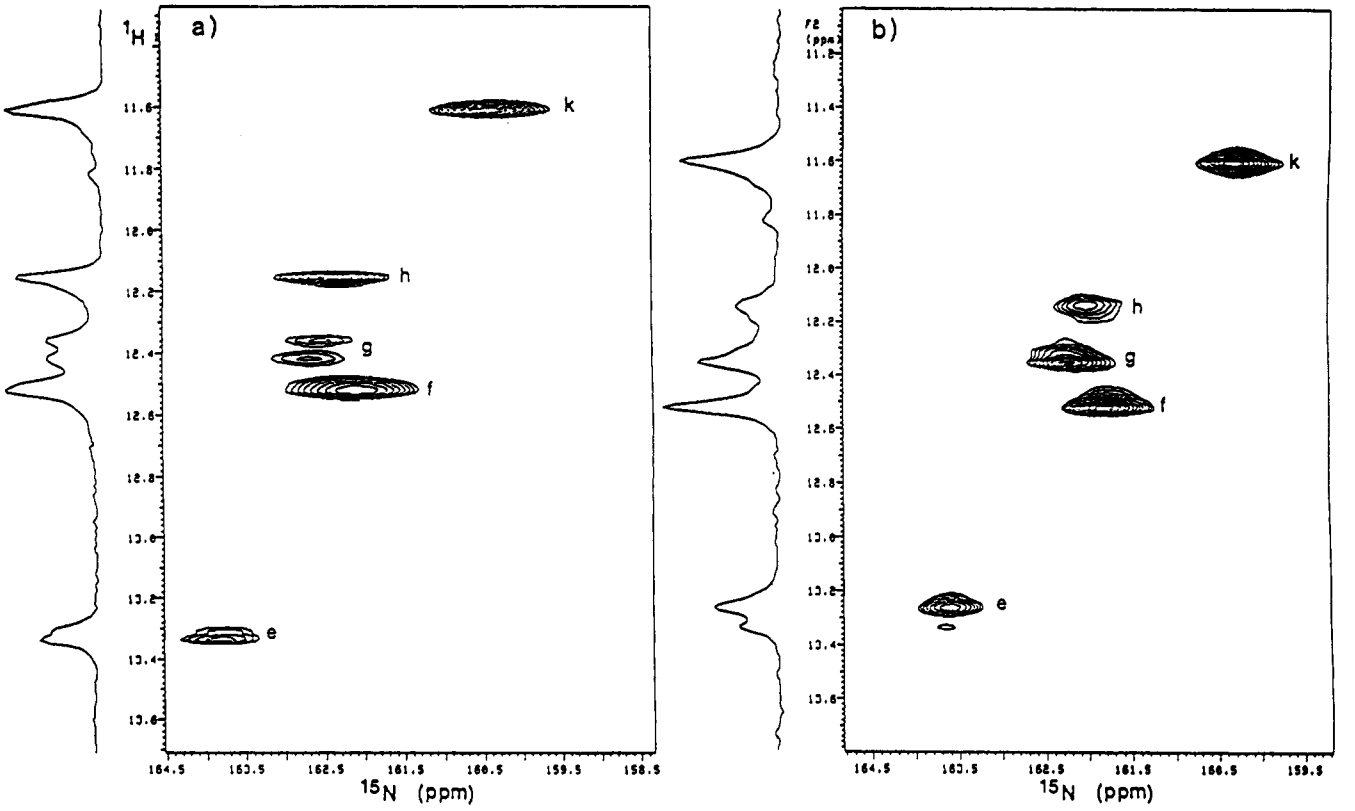

FIGURE 6: 2D ^{1}H/^{15}N HMQCJR spectra of uniform [^{15}N]guanosine RNA (a) free and (b) bound to 102A. Spectra were acquired at 15 °C. The resonance at 11.8 ppm is assigned to a G imino proton, but its exchange is very rapid so its intensity is weak. Peak α (Figure 5b) is identified as a UNH, based on its absence in these spectra.

new resonance(s) observed around 11.8–11.9 ppm in the 1D imino proton spectrum of the complex must come from uridine imino proton(s), since they do not appear in this HMQC experiment. These new resonances indicate that the imino protons in the loop are being protected from exchange in the complex, either through close association with the protein or from a conformational change of the loop that restricts their access to solvent.

(B) [^{13}C]Adenosine as a Probe of RNA Conformation. The direct ^{1}H/^{13}C correlations are used to observe the protons from the ribose rings and the bases. In these experiments, another RNA sample containing uniformly ^{13}C-labeled adenosine was used to observe the spectra of free and bound RNA.

The aromatic region of the ^{1}H/^{13}C HMQC spectra of free and bound [^{13}C]adenosine RNA is shown in Figure 7. In the free RNA, four AH2 resonances are clustered together from 7.0 to 7.3 ppm (^{1}H), and one is near 7.9 ppm. The five AH8 resonances are found near 7.8 ppm. Note that there is unfortunately little chemical shift dispersion in the ^{13}C frequencies at either the C2 or the C8 positions. However, in the spectrum of bound RNA, it is clear that one AH2 resonance has shifted downfield to 8.18 ppm. The proton chemical shifts of four AH8 protons are also shifted downfield relative to their position in free RNA; only one, near 7.6 ppm, appears unchanged. The large change in chemical shift of the aromatic proton from one adenosine base (from near 7.1–8.18 ppm) suggests that its environment has greatly changed, either through proximity to the RBD or from a conformational change in the RNA loop. There is also a change in the stem environment, which may be related to the changes observed in the imino proton spectra.

The ribose region of the ^{1}H/^{13}C HMQC spectra of free and bound RNA is shown in Figure 8. In the spectrum of the free RNA, the H1′/C1′ cross-peaks are all found near 94 ppm (^{13}C); H2′/C2′ near 76 ppm; H3′/C3′ near 73 ppm; H4′/C4′ near 83 ppm; and H5′,5″/C5′ near 65 ppm (Nikonowitz & Pardi, 1993). Although there are five adenosine nucleotides in the molecule, which presumably are in different environments, there is virtually no chemical shift dispersion in either the proton or the carbon dimension. However, when the complex is formed, this situation changes dramatically, as shown in Figure 8B. Now, all five ribose H1′ resonances are visible, and new cross-peaks appear at different proton and carbon frequencies, suggesting that there has been a dramatic change in the environment of these riboses. Several cross-peaks still contain more than one resonance, which accounts for some of the different cross-peak intensities (as do different relaxation rates).

A ribose ring in a canonical RNA A-form duplex adopts a C3′-*endo* conformation, and as a result, the coupling constant between the H1′ proton and the neighboring H2′ proton is on the order of 1–3 Hz, which is too small to be observed in COSY-type experiments. Such a lack of connectivity can be confusing for assignment purposes but useful for diagnosis of sugar pucker conformations that differ from the normal A-form. In ^{1}H TOCSY experiments of the free RNA, H1′/H2′ cross-peaks do appear, which indicate that there are regions of the RNA, no doubt in the loop, in which the ribose is not constrained to the standard C3′-*endo* conformation. The number of these cross-peaks increases as the temperature increases, indicating that the loop conformation is flexible and the sugar puckers are changing (data not shown).

In TOCSY experiments with the free RNA at 15 °C and either a 40 or a 70 ms isotropic mixing time, only four cross-peaks are observed between the H1′ and H2′ protons, which have been assigned to pyrimidine nucleotides in the RNA loop (using TOCSY–HMQC experiments with uniformly ^{13}C-

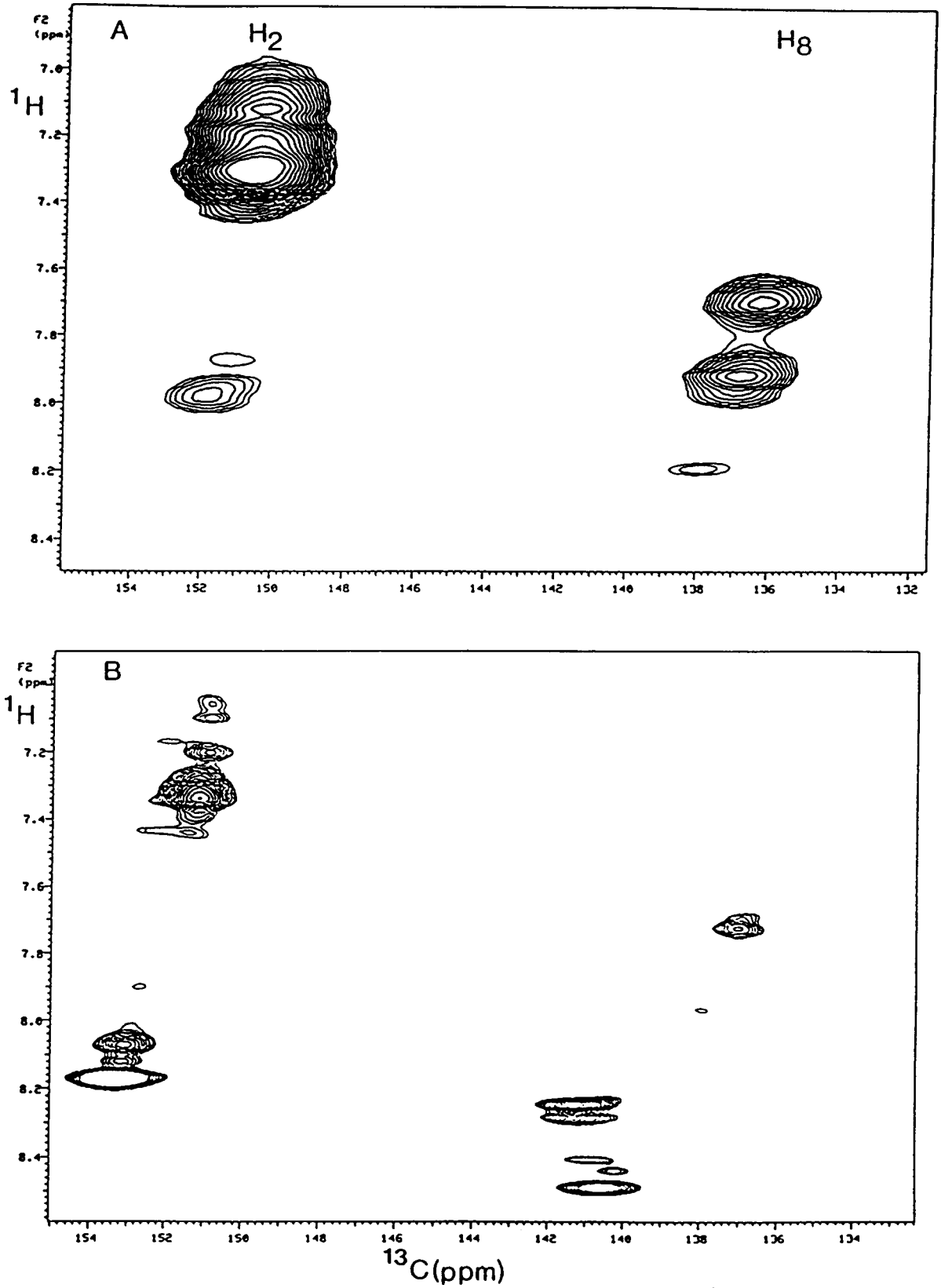

FIGURE 7: 2D $^{1}H/^{13}C$ HMQC spectra of the aromatic region of [^{13}C]adenosine RNA free (A) and bound (b) to 102A at 15 °C. The five AH8 cross-peaks are clearly resolved in the spectrum of the complex; one AH2 resonance has shifted about 0.8 ppm downfield in the complex.

labeled pyrimidines). In 2D TOCSY–HMQC experiments with [^{13}C]adenosine RNA, only the isotope-labeled ribose sugars are observed, since HMQC acts as a filter for the proton TOCSY experiment; no H1′–H2′ cross-peaks are observed from adenosine riboses. Thus, even the two adenosine residues in the loop, under these conditions, have constrained ribose conformations. However, in the TOCSY–HMQC spectrum of the bound [^{13}C]adenosine RNA, again at 15 °C with 40 ms mix, two H1′ to H2′ connectivities are clearly observed. At 70 ms mixing, connectivities are observed also to H3′ protons (Figure 9), indicating that these two riboses are no longer in a C3′-*endo* conformation. Notably, the proton chemical shifts

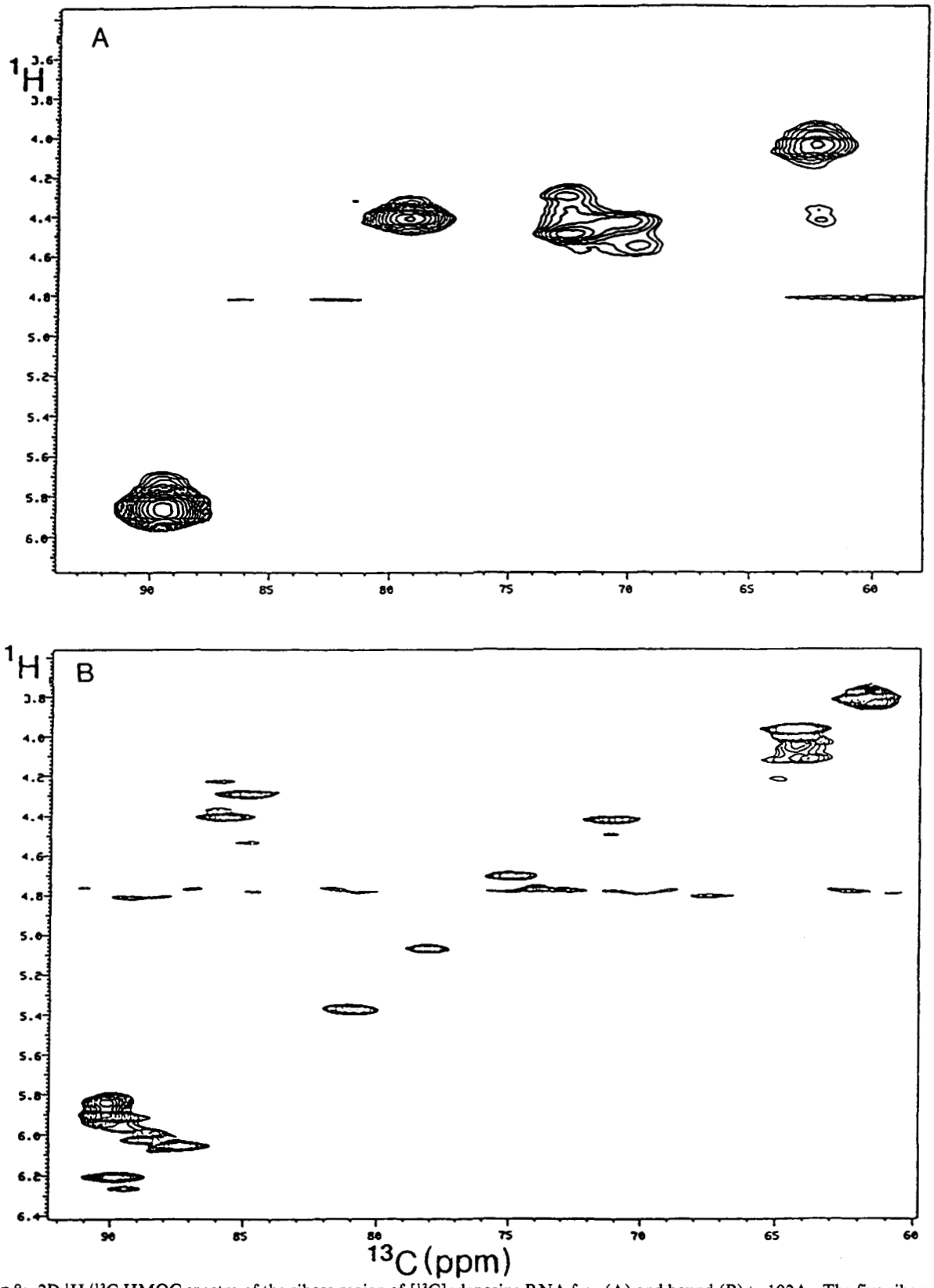

FIGURE 8: 2D ^{1}H/^{13}C HMQC spectra of the ribose region of [^{13}C]adenosine RNA free (A) and bound (B) to 102A. The five ribose C1′/H1′ cross-peaks are found together in the free RNA, but are clearly separated in the bound complex.

of one ribose ring are very unusual—the TOCSY experiment shows that one ribose has an H1′ proton at 6.2 ppm, which is coupled to an H2′ at 5.36 ppm, an H3′ at 5.04 ppm, and an H4′ at 4.4 ppm; the ^{13}C chemical shifts for this ribose are also unusual. The other ribose which shows measurable H1′-H2′ coupling has more normal chemical shifts, with an H2′ resonance at 4.70 ppm and an H3′ at 4.42 ppm. The different puckers of a ribose ring would alter its coupling constants, but not the chemical shift of its protons, and so a different explanation for the large chemical shift changes of this adenosine ribose must be found; by analogy to the experiments of Heus and Pardi (1991) with the GCAA tetraloop, this downfield shift could be due to a ring current effect from an adjacent aromatic ring in the plane of the ribose.

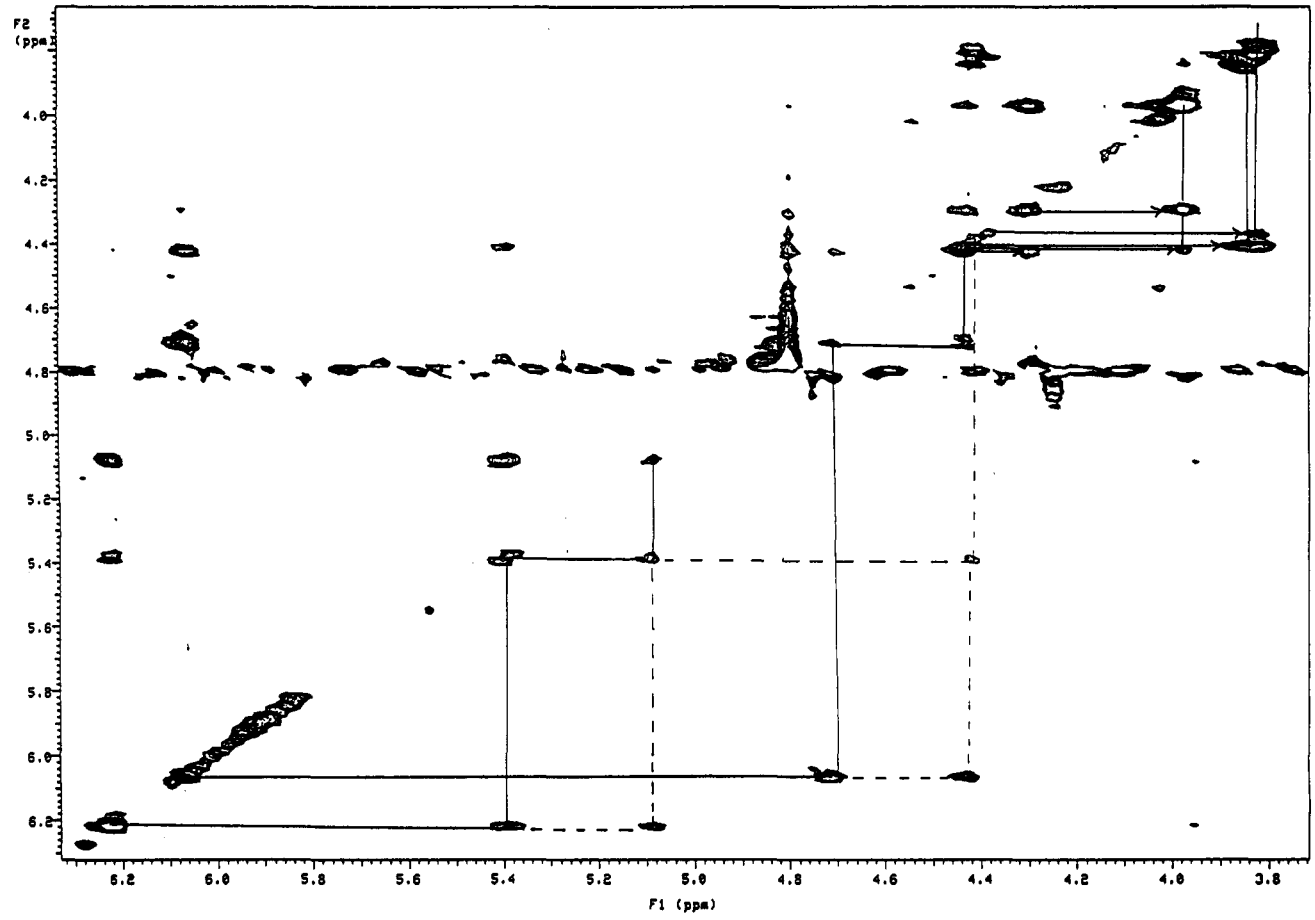

FIGURE 9: 2D (^{1}H/^{1}H) TOCSY–HMQC spectrum of the ribose region of the bound [^{13}C]adenosine RNA. τ_m = 70 ms, 15 °C. The two ribose sugars observed in this spectrum are not found in the same experiment with free RNA, and indicate a conformational change of two of the adenosine nucleotides. Connectivities within the two sugars are shown.

DISCUSSION

The 102A RBD may be the most well-behaved example of this motif available for detailed biochemical characterization, since unlike many isolated RBDs, it is soluble and correctly folded upon overproduction in *E. coli*, it does not aggregate at millimolar concentrations, it exists as a monomer in solution, and it binds a unique RNA substrate with high affinity. For these reasons, it represents an ideal system in which to study RBD–RNA association by biochemical and physical methods. However, the behavior of other RBDs that makes them difficult to work with may also reflect their functional properties, especially in those cases where several RBDs are linked together in a single protein. When an RBD is expected to interact through protein–protein contacts, and also bind RNA (such as U2B″, Scherly et al., 1990b), the question of which protein surface participates in these various functions becomes particularly important, especially since conformational changes in the RBD upon interactions with other molecules would be expected to alter the available binding surfaces. In this sense, the 102A interaction with its RNA hairpin may represent the simplest class of RBD–RNA interactions.

In order to understand the structural basis of sequence-specific recognition by an RBD, it is necessary to identify those regions of the protein that define the form and sequence of its RNA target. This class of RNA binding motif is capable of interacting with single-stranded RNAs as well as RNA hairpins, making it a remarkably versatile domain in spite of its small size. We use the 102A–RNA interaction as a model system to map the regions of the RNA and protein that are involved in the sequence-specific recognition of the RNA by this RBD. Because of its solubility, stability, and clearly defined RNA substrate, the U1A 102A domain is an excellent model system for these types of biochemical and biophysical experiments. Its RNA hairpin substrate is a more sophisticated RNA than the single-stranded RNAs recognized by other RBDs, such as hnRNP C (Swanson & Dreyfuss, 1988), hnRNP A1 (Dreyfuss et al., 1993), or poly(A) binding protein (Sachs et al., 1987), and so the sites of interaction between RNA and 102A RBD may not be identical to those of simpler systems.

While the data are clearly not complete, we propose the following interpretation of the results presented here. The formation of the RNA–protein complex requires that the RNA be presented to the RBD in a specific form, which for this RBD is an RNA hairpin. It is necessary that the phylogenetically conserved GCA sequence in the loop is available to the protein, as sequestering it in a small loop or putting it in a single strand does not allow the proper contacts to be formed. The contribution to the binding free energy is not a simple sum of individual interactions, as shown by the difference in the binding free energy of 102A to various RNA mutants. Formation of the complex apparently occurs through a series of interactions that are interdependent; failure to make one contact can influence subsequent contacts. These binding data and NMR results suggest that there is a conformational change in the RNA upon association with the protein. There are many systems in which this induced-fit association, first hypothesized by Koshland (1958) for enzyme–substrate complexes, is an accurate description of the recognition process: Frankel and Kim discuss the example of the leucine zipper of GCN4 (1991), but induced fit can also describe the arginine–TAR association (Puglisi et al., 1992), λ cro repressor

(Clarke et al., 1991) interaction with DNA (see Spolar & Record, 1994), and DNA-induced dimerization of *E. coli* Rep helicase (Wong et al., 1992). However, in this RBD–RNA complex, it should not be assumed that the conformation of the bound RNA is a static, rigid one; the NMR data could be interpreted to indicate that the RNA hairpin is dynamically associated with the RBD.

The temperature dependence of the association is nonlinear, and so the van't Hoff plot (ln K_A vs $1/T$) cannot be simply interpreted to give the enthalpy ($\Delta H°$) of association. Rather, one interpretation of a nonlinear van't Hoff plot, observed for several DNA–protein associations, involves either the burial of hydrophobic residues during association (Ha et al., 1989) to account for a real heat capacity change (ΔC_p) or a conformational change in either of the two molecules to provide the appropriate surface for the association, accompanied by an unfavorable entropic contribution (Spolar & Record, 1994). While these data cannot exclude the possibility that burying hydrophobic residues is the source of the nonlinearity, other data support the interpretation that there is a conformational change of the RNA upon association with 102A. What is certainly true is that there is entropy/enthalpy compensation in this association, where the driving force (entropic or enthalpic) changes with temperature. Such compensation must be occurring also for association of the RBD with other RNA variants, but these data are not sufficient to fully describe how these thermodynamic parameters change with RNA sequence.

In this context of the thermodynamics of complex formation, it is important to mention previous data from the hnRNP A1 RBD interaction with dT_8, in which phenylalanines were covalently cross-linked to the DNA (Merrill et al., 1988). These data have suggested that these hydrophobic amino acids intercalate between the nucleic acid bases and thus would contribute to the favorable enthalpy of association (through aromatic ring stacking). The NMR data of the RNA–102A complex do not detect intercalation of any amino acid into the stem of the RNA, since such an event would cause significant changes in the imino proton chemical shifts as well as disrupt the sequential imino NOEs between the bases. Thus, we can rule out intercalation of phenylalanine or tyrosine side chains into the stem of the RNA hairpin. However, the unusual proton chemical shifts of one adenosine nucleotide (base and ribose) in the complex suggest that a ring current effect is responsible, and the critical question of the identity of that aromatic ring is unanswered. The downfield shift of the ribose proton resonances indicates that the sugar is not stacked over this ring, however, but is located in its plane (since the effect is to deshield the protons). This ring could be an aromatic amino acid, but it could also be a nucleic acid base.

Removing the last seven amino acids from 102A to create 95A results in a loss of 2 kcal/mol of binding free energy, relative to the 102A–RNA interaction, or approximately 30× decrease in affinity. These data agree with previous experiments in which the protein was truncated at Ser91 (Scherly et al., 1989) or Lys96 (Jessen et al., 1991), or where Lys98 was replaced with Gln (Nagai et al., 1990). Lys96 and Lys98 were hypothesized to contribute to binding affinity through electrostatic interactions (Jessen et al., 1991), and the experiments here show that, in fact, truncation at Ala95 results in a loss of two ions released upon complex formation, relative to the 102A–RNA interaction. Deletion of this end of the RBD does not significantly destabilize the protein, based on calculations of the standard free energy of denaturation (in guanidine) for the 95A and 102A RBD (data not shown). We speculate, therefore, that this end of the RBD is flexible and acts like an arm to make nonspecific contacts with the negative charges on the RNA phosphate backbone. Our hypothesis is that it interacts with the RNA stem, perhaps to fold around the helix analogous to the arm of λ repressor (Sauer et al., 1990). However, an important caveat is that in the context of the intact U1A protein, what in 102A is the C-terminal end could adopt a completely different conformation in relation to the complex.

Many mutations in the RBD have been shown to contribute to the affinity of the U1A–RNA interaction (Nagai et al., 1990; Jessen et al., 1991; Scherly et al., 1989; Boelens et al., 1991). Several substitutions in the RNP2 have been shown to eliminate specific binding (Jesson et al., 1991), as did one in RNP1, R52Q (Nagai et al., 1990), while another, Q54F, increased the dissociation constant 100X in those assays. Experiments described here show that mutations in RNA can also have severe effects on the affinity of this association. In order to fully describe the interactions between RBD and RNA, these mutation results must be correlated to identify those sites that are in direct specific contact, those nonspecific sites that contribute to overall affinity, and those sites that confer conformational stability and flexibility to the RNA, RBD, or both. Thus, the surface of the 102A RBD that is in direct contact with the RNA hairpin seems still undefined, and more experiments are necessary to identify the RNA–protein contacts.

Rather than observe the RNA, the RBD was observed in NMR experiments with hnRNP C and rU_8, its preferred substrate (Gorlach et al., 1992). These data showed that complex formation resulted in changes in the chemical shifts of many amino acids, localized in the β sheet and the C- and N-terminal regions (Gorlach et al., 1992). As these authors suggested, the N- and C-terminal regions could adopt an ordered structure upon association with the RNA, and our results with 95A and 102A suggest that the same phenomenon could apply here. In the interaction of hnRNP C with rU_8, the amino acids of the α helices apparently did not participate in binding to the RNA, since their chemical shifts were unchanged in the free and bound states. Those results, showing that the surface of the β sheet is part of the binding site for RNA, agree with the previous hnRNP A1–dT_8 cross-linking data (Merrill et al., 1988). We suggest, however, that this mode of RBD–RNA interaction may not be conserved in cases where the RNA substrate itself contains significant structure. The bases in a single strand of rU or dT present a relatively nonpolar surface for interaction with aromatic amino acids on the surface of the protein, as indeed the hnRNP C and hnRNP A1 data show. For the case of 102A–RNA interactions, however, there is no simply single-stranded region of the RNA that could sit on the β sheet; a single-stranded RNA substrate does not bind to 102A with reasonable affinity. The RNA stem, which is required for binding, presents a negatively charged array of phosphates to the RBD, and since the surface of the 102A β sheet is nonpolar, it seems unlikely that the RNA stem can be positioned there. Since the RNP1 octamer (β3) and RNP2 hexamer (β1), which make up a large part of the β sheet, are not unique to 102A, these regions are unlikely to participate in sequence-specific interactions with the RNA, and so the RNA loop (at least the conserved GCA sequence) is probably not directly associated with this part of the RBD. As other investigators have shown, the β2 strand at the edge of the β sheet (Scherly et al., 1990) and the loop connecting β2 and β3 (Bentley & Keene, 1991) are critical for sequence-specific association with RNA stem/loop II. Thus, a model of 102A–RNA interaction should tentatively position the RNA loop in proximity to this region of the protein, and must also provide for electrostatic interactions with the RNA phosphate

backbone. One model of this interaction (Jesson et al., 1991) uses the crystal structure of 95A and a static RNA structure; however, our data indicate that a more complicated picture is necessary, since this interaction is not a lock and key association, but probably involves conformational changes of the RNA. In addition, there is a critical sequence-dependence to the association that indicates that specific nucleotides are in direct contact with the protein.

In summary, the data presented here show that substitutions in the RNA sequence lead to dramatic loss of binding affinity by the 102A RBD; NMR data suggest that a conformational change in the loop of the RNA occurs upon formation of the complex. These data together suggest that the complex is formed through a concerted rearrangement of the RNA, and probably the protein as well, to form the optimal and base-specific contacts leading to this high-affinity and very specific association. The formation of the complex can thus be best described as an induced fit between the two molecules, where individual nucleotides make substantial contributions to the binding free energy not only through their contact with the protein but also through their facilitation of subsequent contacts. Our model of the interactions between RNA and 102A involves nonspecific contributions from the RNA stem and sequence-specific contributions from the loop.

ACKNOWLEDGMENT

I thank W. Tom Stump for superb technical assistance (and protein and [^{13}C]- and [^{15}N]NTPs) and Jim Kranz for loopless and U10G binding data. Thanks to Jim Kranz, Jirong Lu, and Tim Lohman for helpful comments and to Art Pardi for enlightening discussions.

REFERENCES

Bandziulis, R. J., Swanson, M. S., & Dreyfuss, G. (1989) *Genes Dev. 3*, 431–437.

Bentley, R. C., & Keene, J. D. (1991) *Mol. Cell. Biol. 11*, 1829–1839.

Boelens, W., Scherly, D., Jansen, E. J. R., Kolen, K., Mattaj, I. W., & van Venrooij, W. J. (1991) *Nucleic Acids Res. 19*, 4611–4618.

Boelens, W., Jansen, E. J. R., van Venrooij, W. J., Stripecke, R., Mattaj, I. W., & Gunderson, S. I. (1993) *Cell 72*, 881–892.

Cantor, C. R., & Tinoco, I., Jr. (1965) *J. Mol. Biol. 13*, 65–77.

Carey, J., & Uhlenbeck, O. C. (1983) *Biochemistry 22*, 2610–2615.

Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C., & Pabo, C. O. (1991) *Science 254*, 267–270.

Dreyfuss, G., Swanson, M. S., & Pinol-Roma, S. (1988) *Trends Biochem. Sci. 13*, 86–91.

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., & Burd, C. (1993) *Annu. Rev. Biochem. 62*, 289–321.

Fasman, G. D., Ed. (1975) *Handbook of Biochemistry and Molecular Biology*: *Nucleic Acids*, Vol. I, 3rd ed., CRC Press, Cleveland, OH.

Frankel, A. D., & Kim, P. S. (1991) *Cell 65*, 717–719.

Ghetti, A., Padovani, C., Di Cesare, G., & Morandi, C. (1989) *FEBS Lett. 257*, 373–376.

Ghetti, A., Bolognesi, M., Cobianchi, F., & Morandi, C. (1990) *FEBS Lett. 277*, 272–276.

Gorlach, M., Wittekind, M., Beckman, R., Mueller, L., & Dreyfuss, G. (1992) *EMBO J. 11*, 3289–3295.

Grodberg, J., & Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245–1253.

Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., & Mattaj, I. W. (1994) *Cell 76*, 531–542.

Ha, J. H., Spolar, R. S., & Record, M. T. (1989) *J. Mol. Biol. 209*, 801–816.

Hall, K. B., & McLaughlin, L. W. (1992) *Nucleic Acids Res. 20*, 1883–1889.

Hall, K. B., & Stump, W. T. (1992) *Nucleic Acids Res. 20*, 4283–4290.

Heus, H. A., & Pardi, A. (1991) *Science 253*, 191–194.

Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495–2499.

James, T. L., James, J. L., & Lapidot, A. (1981) *J. Am. Chem. Soc. 103*, 6748–6750.

Jessen, T. H., Outbridge, C., Teo, C. H., Prichard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447–3456.

Jorgensen, E. D., Durbin, R. K., Risman, S. S., & McAllister, W. T. (1991) *J. Biol. Chem. 266*, 645–651.

Keene, J. D., & Query, C. C. (1991) *Prog. Nucleic Acids Res. 41*, 179–202.

Kenan, D. J., Query, C. C., & Keene, J. D. (1991) *Trends Biochem. Sci. 16*, 214–220.

Koshland, D. E. (1958) *Proc. Natl. Acad. Sci. U.S.A. 44*, 98–102.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A., & Clore, G. M. (1989) *Biochemistry 28*, 6150–6156.

Mattaj, I. W. (1989) *Cell 57*, 1–3.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., & Williams, K. R. (1988) *J. Biol. Chem. 263*, 3307–3313.

Michnicka, M. J., Harper, J. W., & King, G. C. (1993) *Biochemistry 32*, 395–400.

Milligan, J. F., Groebe, D. R., Witherall, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.

Mueller, L. (1979) *J. Am. Chem. Soc. 101*, 4481–4484.

Nagai, K., Outbridge, C., Jessen, T. H., Li, J., & Evans, P. R. (1990) *Nature 348*, 515–520.

Nikonowitz, E. P., & Pardi, A. (1993) *J. Mol. Biol. 232*, 1141–1156.

Nikonowitz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., & Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507–4513.

Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76–80.

Query, C. C., Bentley, R., & Keene, J. (1989) *Cell 57*, 89–101.

Record, M. T., Lohman, T. M., & de Haseth, P. L. (1976) *J. Mol. Biol. 107*, 145–158.

Record, M. T., Anderson, C. F., & Lohman, T. M. (1978) *Q. Rev. Biophys. 11*, 103–178.

Romaniuk, P. J. (1985) *Nucleic Acids Res. 13*, 5369–5387.

Sachs, A. B., Davis, R. W., & Kornberg, R. D. (1987) *Mol. Cell. Biol. 7*, 3268–3276.

Sauer, R. T., Jordan, S. R., & Pabo, C. O. (1990) *Adv. Protein Chem. 40*, 1–61.

Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., & Mattaj, I. W. (1989) *EMBO J. 8*, 4163–4170.

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., & Mattaj, I. W. (1990a) *Nature 345*, 506–512.

Scherly, D., Dathan, N. A., Boelens, W., van Venrooij, W. J., & Mattaj, I. W. (1990b) *EMBO J. 9*, 3675–3681.

Spolar, R. S., & Record, M. T. (1994) *Science 263*, 777–784.

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.

Swanson, M. S., & Dreyfuss, G. (1988) *Mol. Cell. Biol. 8*, 2237–2241.

van Gelder, C. W., Gunderson, S. I., Jansen, E., Boelens, W., Polycarpou-Schwartz, M., Mattaj, I. W., & van Venrooij, W. J. (1993) *EMBO J. 12*, 5191–5200.

Wittekind, M., Gorlach, M., Freidrichs, M., Dreyfuss, G., & Mueller, L. (1992) *Biochemistry 31*, 6254–6265.