RNA (1995), *1*:55–63. Cambridge University Press. Printed in the USA.

Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain

W. TOM STUMP and KATHLEEN B. HALL

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

ABSTRACT

The N-terminal RNA binding domain (RBD) of the human U1A snRNP protein binds tightly and specifically to an RNA hairpin that contains a 10-nucleotide loop. The protein is one of a class of RNA binding proteins that adopts a $\beta\alpha\beta\beta\alpha\beta$ global fold, which in turn forms a four-stranded antiparallel β-sheet. This sheet forms the primary binding surface for the RNA, as shown by the crosslinking results described here, and in more detail by a recently described co-crystal of this RBD with an RNA hairpin (Oubridge C, et al., 1994, *Nature 372*:432–438). The RNA hairpin sequence used in the crosslinking experiments, containing 5-iodo-uridine, is a variant of the normal U1 snRNA sequence which is able to form a crosslink with the protein, in contrast to the wild-type sequence, which does not. This single uridine substitution in the 10-nucleotide loop is the site of crosslinking to one tyrosine (Tyr 13) in the β1 strand of the U1A N-terminal RBD. This same uridine is also crosslinked to a mutant Tyr 13 Phe RBD, at this Phe 13 substitution.

Keywords: 5-iodo-uridine RNA; protein mutagenesis; site-specific crosslinking; U1A RBD

INTRODUCTION

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 identified by the presence of two 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; Birney et al., 1993; see Keene & Query, 1991, and Dreyfuss et al., 1993, for reviews of RBD proteins).

The tertiary structures of three RBDs are known: the human U1A N-terminal domain (Nagai et al., 1990; Hoffman et al., 1991), the hnRNP C protein (Wittekind et al., 1992), and RBD(2) of *Drosophila sxl* (Lee et al., 1994); and also the secondary structure of two others: from hnRNP A1 (Garrett et al., 1994) and the human U1A C-terminal RBD (Lu & K.B. Hall, in prep.). All contain the $\beta\alpha\beta\beta\alpha\beta$ fold predicted for RBD sequences (Ghetti et al., 1989) with a β-sheet formed from four antiparallel β strands. The adjacent strands β3 and β1 contain the RNP1 and RNP2 sequences, respectively. Two α-helices are positioned on one side of the β-sheet.

Because many of these RBDs discriminate among RNA substrates, understanding the molecular basis of their sequence-specific recognition process could allow predictions of RNA targets, as well as manipulations of target specificity. Several experiments have suggested sites on RBDs that participate in interaction with RNA. In experiments with hnRNP A1, a DNA substrate, $d(T)_8$, was photocrosslinked to the protein and the sites of covalent addition were mapped (Merrill et al., 1988). Two phenylalanines, one in the RNP-consensus octamer (RNP1) and one in the hexamer (RNP2), were the primary sites of addition. These data have led to a model in which the surface of the β-sheet is the site of RNA binding. These data agree with recent results of Gorlach et al. (1992, 1994) with the human hnRNP C RBD and rU_8 and with results of Howe et al. (1994) in NMR studies of the U1A N-terminal RBD. For the U1A N-terminal RBD, the interacting surface of the protein was also identified biochemically by several sequence-swap experiments with the closely related protein U2B″, where the β2 strand and the loop between β3 (RNP1) and β2 were shown to be critical for RNA binding specificity (Scherly et al., 1990; Bentley & Keene, 1991). Most recently, a co-crystal struc-

Reprint requests to: Kathleen B. Hall, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, USA.

ture of this RBD with a 21-nucleotide RNA has been solved (Oubridge et al., 1994). Those results show an intricate hydrogen bonding network formed between the protein and the RNA loop as the RNA sits on the β-sheet. Previous biochemical data, including binding of RNA and protein mutants, thermodynamics, and work in solution using NMR, can now be compared with this co-crystal structure (Scherly et al., 1990; Boelens et al., 1991; Jessen et al., 1991; Hall & Stump, 1992; Hall, 1994; Howe et al., 1994). In the experiments described here, another biochemical method is used to probe this interaction. Photochemical crosslinking, using photoreactive 5-I-uridine RNA, is used to map the RNA physically onto the protein. Results of sequence analysis of the crosslinked species indicate that there is one site on the RNA that is crosslinked to one amino acid on the protein. These results from solution are compared to what is predicted from the co-crystal.

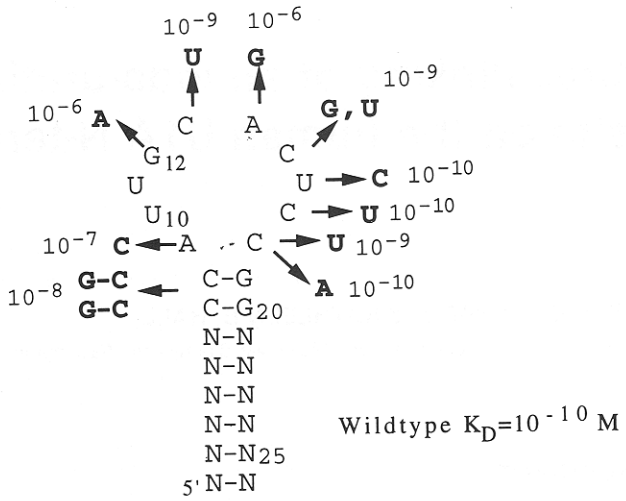

FIGURE 1. Sequences of the RNA hairpins used for the binding and crosslinking experiments, with the binding affinities to 102A (dissociation constants, K_D) shown. Binding was measured in 200 mM NaCl, 1 mM $MgCl_2$, 10 mM sodium cacodylate, 20 μg/mL BSA, 0.5 μg/μL tRNA, pH 6, 22 °C.

RESULTS

Association of RNAs and proteins

The N-terminal RBD of the human U1A protein binds to stem/loop II of the U1 snRNA (Scherly et al., 1989; Lutz-Freyermuth et al., 1990) and as an autonomous domain can bind tightly and specifically to a short RNA hairpin (Jessen et al., 1991; Hall & Stump, 1992; Hall, 1994; Oubridge et al., 1994). In addition, U1A also binds to its own mRNA 3′ untranslated region (UTR), where it can inhibit polyadenylation and regulate its translation (Boelens et al., 1993; van Gelder et al., 1993); this interaction is again mediated by the N-terminal RNA binding domain.

The first 102 amino acids of the U1A N-terminal RBD (102A) binds the wild-type sequence of a 25-nucleotide RNA hairpin tightly and specifically (Jessen et al., 1991; Hall & Stump, 1992; Hall, 1994). To determine potential sites of interaction between the RNA and 102A in crosslinking experiments, the RNA loop nucleotides were systematically substituted with uridine. Because formation of a stable complex is dependent on the RNA sequence, the binding constants of these substituted RNAs were determined to assess the effect of the mutation.

As shown in Figure 1, the substitution of selected nucleotides with uridine suggested that only the 5′ seven nucleotides of the loop were critical for contact with the protein. Substitution of U16C and C17U did not alter the affinity. When C18 was replaced with uridine, the binding constant was reduced 10×; in the C18A substitution, however, the RNA is bound with wild-type affinity. These results indicate that the C18U substitution, which allows formation of the A9U18 base pair, impedes the ability of 102A to bind the RNA. Nucleotide A9 is one of the most significant bases in formation of the complex, judging by the effect of mutation on binding (Hall, 1994), and so the effect of the AU base pair is presumably to obstruct access to this base. Nucleotides G12 and A14 were not replaced in these uridine-scanning experiments because previous work showed them to be essential for complex formation (Hall, 1994).

These substitution data indicated that nucleotides 9–15 were critical for formation of the complex, and also that 102A uses A9 as a contact but not C18, suggesting that the RBD is selectively interacting at the base of the RNA loop. Previous mutation data had shown that association is not affected by substitution of base pairs in the RNA stem (Hall, 1994), but those experiments had not altered the closing C8G19 base pair, which is also present in the U1 snRNA sequence. As Figure 1 shows, substituting this CG pair with GC results in a loss of 2 kcal/mol of binding free energy, suggesting that there are specific contacts between the RNA and protein at this position, or the conformation of the loop is altered by these new closing nucleotides and so affects the ability of the protein to associate with the RNA. However, a third, less interesting, possibility is that the base pairing arrangement of the stem and loop is altered, resulting in a six-nucleotide loop that is not recognized by the protein.

Photochemical crosslinking

The wild-type RNA sequence and the uridine-scanning variants were enzymatically synthesized with 5-I-uridine, to produce RNAs that were uniformly substituted with 5-I-U. As shown in Figure 2, introduction of the iodo group did not disrupt complex formation and in fact increased the binding affinity of C13U for

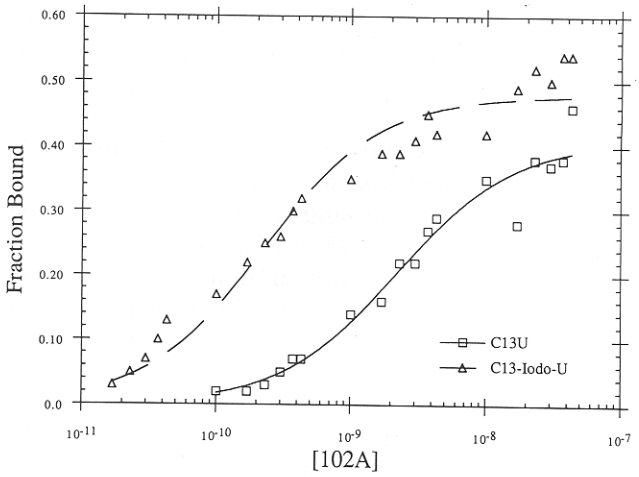

FIGURE 2. Binding isotherm of the C13U RNA and its 5-I-U variant bound to the 102A protein. Data points are shown; smooth line is the fit to a Langmuir isotherm, assuming a two-state model and a 1:1 stoichiometry. Binding assays include 200 mM NaCl, 10 mM sodium cacodylate, pH 6, 1 mM $MgCl_2$, 20 μg/mL BSA, 0.5 μg/μL tRNA, 22 °C.

102A. A similar result was observed for the R17 hairpin/coat protein interaction, using 5-Br-U RNA (Gott et al., 1991) and 5-I-U RNA (Willis et al., 1993).

These 5-I-U RNAs, containing α-^{32}P-ATP, were then used in photochemical crosslinking experiments with 102A. An example of the reaction products is shown in Figure 3, where it is clear that only one species is observed in the SDS-PAGE gel, which migrates slower than the input RNA or the free protein. Surprisingly, there are no significant intramolecular crosslinks in the RNA alone, although some degradation is seen, especially at longer irradiation times. Table 1 gives the percentage of input RNA crosslinked to the protein. Notably, the wild-type loop sequence did not yield sufficient amount of crosslinked complex for subsequent analysis. Only one RNA sequence, C13U, produced significant amounts of crosslinked product (typically 70% of input RNA).

This C13U RNA was also synthesized with 5-Br-uridine, and the crosslinking efficiencies of the two photoreactive RNA molecules were compared, using the Spectroline 312-nm light source. In these experiments with the 102A RBD, the 5-I-U RNA was crosslinked at 56% efficiency after 20 min of irradiation and at 58% after 90 min. The 5-Br-U RNA showed an efficiency of 2.4% after 20 min, and only 13% after 90 min (data not shown). Because the amount of RNA degradation was greater in the iodo-U-RNA samples than in the Br-U-RNAs, destruction of the RNA was not the cause of the poor Br-U-RNA crosslinking efficiency. A similarly high yield for iodinated uridine compared to the brominated species was reported for the R17/coat protein complex (Willis et al., 1993).

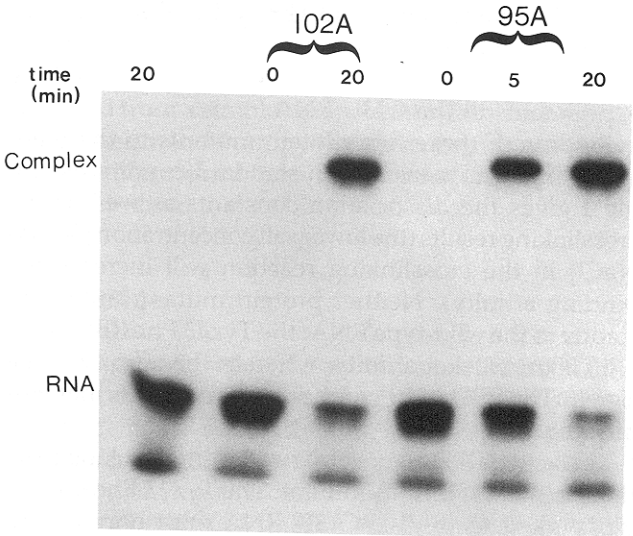

FIGURE 3. Products of a photochemical crosslinking reaction using 5-I-U C13U RNA and 102A or 95A RBD; minutes of irradiation at 4 °C are indicated. Free RNA and complex mobilities on the 15% SDS/PAGE gel are shown; free protein is found about midway between (not shown). There is some decomposition of the RNA observed, even in the absence of irradiation.

Crosslinked sites

The C13U RNA:102A complex was used for subsequent determination of the specific crosslinking sites on the RNA and the protein. To determine the site(s) on the RNA, two nucleases were used to digest the crosslinked RNA, to look for the disappearance of specific fragments in urea/polyacrylamide gels. To determine the site on the protein, the complex was electroblotted onto Immobilon, and directly sequenced.

TABLE 1. Binding affinity and crosslinking efficiency.

Protein	RNA	K_D[a] (M)	Crosslink efficiency[b] (%)
102A	WT	$2 \pm 1 \times 10^{-10}$	3
	C13U	$3 \pm 2 \times 10^{-9}$	67
	C15U	$3 \pm 2 \times 10^{-9}$	3
	C17U	$2 \pm 1 \times 10^{-10}$	3
	C18U	$2 \pm 1 \times 10^{-9}$	10
95A	WT	$1 \pm 1 \times 10^{-9}$	3
	C13U	$3 \pm 2 \times 10^{-8}$	66
Y13T	WT	$3 \pm 2 \times 10^{-5}$	
	C13U		1
Y13F	WT	$3 \pm 2 \times 10^{-7}$	
	C13U		69

[a] Dissociation constant measured by nitrocellulose filter binding, using RNA containing uridine.

[b] Crosslinking efficiency measured after 20 min of irradiation, using 5-I-U RNA. Values given are for one particular experiment, but are typical. Incubation conditions are described in the Materials and methods.

Although it was likely that the C13U substitution itself was the site of crosslinking on the RNA, it was also possible that this substitution altered the RNA loop structure and so allowed a new protein contact or changed the proximity at one of the other 5-I-U sites. Therefore, the RNA site was mapped by nuclease digestion of the RNA, similar to the method used for the R17 interaction (Gott et al., 1991). Because the RNA degraded with time, these manipulations had to be done soon after isolation of the complex.

The results of the nuclease digestions are shown in Figure 4, where the crosslinked RNA is compared to the free I-U RNA. Because only the four adenosines are α-^{32}P-labeled, only three fragments are observed in the T1 digestion: 9 nt (U4–G12), 7 nt (U13–G19), and 1 nt (G20); and three fragments in the RNase A digestion: 4 nt (G1–U4 and G19–U22, which runs as two bands due to the 5'pppG), 2 nt (G12, U13), and 1 nt (C8). In each digestion, it is clear that only one α-^{32}P-RNA product is missing from the crosslinked RNA; in the case of the T1 digestion, it is that fragment from U13 to G19, and in the RNase A digestion, it is the dimer G12U13. Thus, these results show that the C13U is the site that crosslinks to the protein.

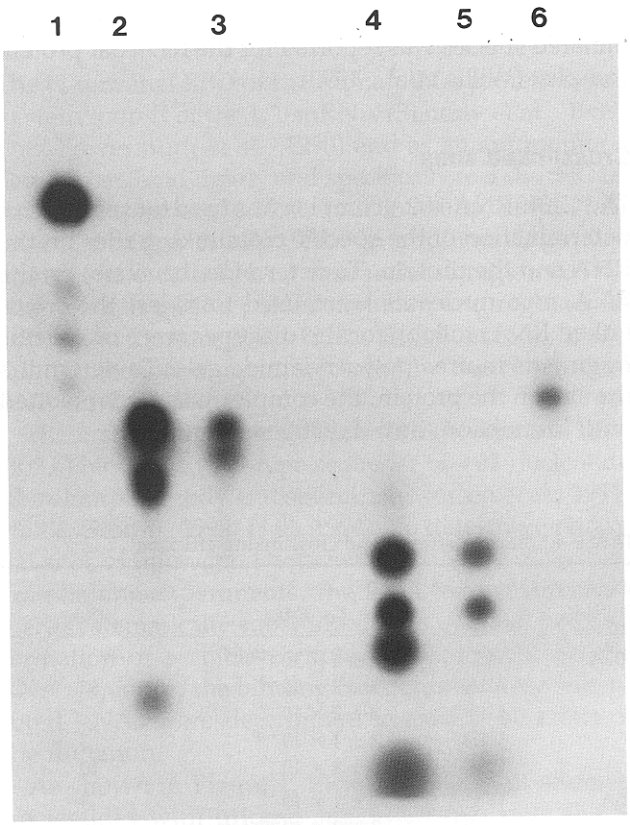

FIGURE 4. Nuclease digestion of the α-^{32}P-ATP-labeled C13U RNA, free and crosslinked to 102A, resolved on a 20% polyacrylamide/8 M urea gel. Lane 1, free RNA, no crosslinking or nuclease digestion. Lane 2, free RNA + nuclease T1. Three products (9 nt U4–G12, 7 nt U13–G19, and 1 nt G20) are observed. Lane 3, crosslinked RNA + nuclease T1. The 7-nt product is absent; the 1-nt product is not visible in this exposure. Lane 4, free RNA + RNase A. Four products (4 nt G19–U22, 4 nt pppG1–U4, 2 nt G12, U13, and 1 nt C8) are observed. Lane 5, crosslinked RNA + RNase A: the 2-nt product is absent. Lane 6, intact complex (observed band is a degradation product that is observed occasionally). The digestion pattern of the irradiated RNA in the absence of protein was nearly identical to the free RNA (not shown).

To locate the crosslinking site on the protein, the intact complex with the 25-nt RNA covalently bound was resolved from free protein on an SDS-PAGE gel and electroblotted onto an Immobilon membrane. A portion of the blot was stained to visualize the band, and a portion of the band was used in direct amino acid sequencing. Sequencing showed that of the first 20 N-terminal amino acids, one was completely absent, whereas the others appeared clearly (data not shown). The missing amino acid, tyrosine 13, is therefore identified as the site of crosslinking with the protein. No further sequencing of the protein was done.

Protein mutations and RNA binding

To assess the contribution of Tyr 13 to complex formation, two mutations were made: Tyr 13 Phe and Tyr 13 Thr were constructed by PCR mutagenesis, in the background of the 95A RBD. The 95A construct lacks seven amino acids at the carboxyl-terminal of the RBD, and binds to the RNA with a reduced affinity (approximately 30× weaker; Table 1) but retains the specificity for the RNA substrate (Hall, 1994; data not shown). These two mutants were determined to have the proper secondary structure, based on comparisons of their CD spectra with that of 95A (data not shown). In addition, their stability was measured by guanidine denaturation: the Tyr 13 Phe protein is destabilized by 0.3 kcal/mol and the Tyr 13 Thr by 1 kcal/mol relative to the wild-type 95A sequence (data not shown). These two constructs were used in binding and crosslinking experiments to the C13U RNA.

Binding of these two protein mutants to the wild-type RNA was measured in standard conditions. Table 1 gives the dissociation constants as well as the crosslinking results (the lower salt concentration [50 mM NaCl] in the crosslinking reaction will increase the binding affinity). Neither protein mutant binds efficiently to the wild-type RNA: the Tyr 13 Phe RBD binds with 300× weaker affinity, whereas the affinity of the Tyr 13 Thr RBD is 10^4× weaker than the normal 95A binding.

These two RBDs were also used in crosslinking reactions to the C13U RNA mutant. The Tyr 13 Phe protein crosslinked to the I-U-C13U RNA with normal efficiency, whereas the Tyr 13 Thr protein did not crosslink to C13U. Although it is possible that the binding constant is too weak to allow efficient crosslinking of the Tyr 13 Thr to RNA, it is also likely that threonine is not a reactive site for the photochemical addition. Because other protein mutants with similar affinities for

the RNA did crosslink under identical conditions (data not shown), the most likely explanation for the Tyr 13 Thr result is that this amino acid is not a target of the photoactivated I-U.

Crosslinking results with the Tyr 13 Phe mutant were interesting because it was not obvious that Phe 13 would be the site of covalent addition in this mutant, for there is no report of I-U or Br-U addition at phenylalanine. The Tyr 13 Phe crosslink could arise through addition to Phe 13, although alternatively, the introduction of Phe 13 could have altered the binding surface of the complex, to provide a new site for RNA association and crosslinking. Therefore, the first 20 amino acids of the Phe 13 RBD crosslinked to the C13U RNA were sequenced, as previously described, to determine if the site of addition was again at position 13. Results from that analysis showed that all these amino acids, with the exception of Phe 13, were clearly observed; Phe 13 was absent. Thus, covalent addition of the C13U RNA to the Tyr 13 Phe RBD also occurs at position 13 on the β1 strand.

The reduced binding affinity measured for both protein mutations could be interpreted to mean that there are specific contacts formed between the RNA and Tyr 13 that are lost with these substitutions, or that other amino acids are prevented from interacting with the RNA due to a rearrangement of the binding surface, or that there is an entropic or enthalpic penalty for altering the ions and waters bound to the protein or in the protein:RNA interface. The co-crystal structure (Oubridge et al., 1994) shows no direct contacts between Tyr 13 and RNA. Because only stacking interactions connect this amino acid and nucleotide base, Phe 13 should be able to substitute for Tyr 13, and the crosslinking result would suggest that it does, but energetically the Phe 13 RBD is not equivalent to wild type. The energetic cost of this substitution cannot be assessed from a simple comparison with the co-crystal; the enthalpy and entropy that describe this association must be determined from the temperature-dependence of the binding.

DISCUSSION

It now seems clear that the binding site of RNA on RNA binding domains utilizes the surface of the β-sheet of the protein. These crosslinking results agree with previous results with other RBD:RNA complexes, including photochemical crosslinking of $d(T)_8$ to the RNA binding domains of hnRNP A1 (Merrill et al., 1988), and NMR experiments with the hnRNP C1 RBD and $r(U)_8$ (Gorlach et al., 1992) and with the U1A N-terminal 101A RBD and an RNA hairpin (Howe et al., 1994). The Tyr 13 in β1 (RNP2) that is crosslinked in the 102A:RNA complex corresponds to the Phe 16 site found in the hnRNP A1:dT_8 crosslinking.

Because there is now a structure of a co-crystal of the 102A:RNA complex, these crosslinking results and binding data can be compared to those contacts that are described in the crystal structure. There is always the possibility that the solution structure and the crystal structure of a complex will show some differences. In an association like this one, where the charges on the RNA backbone are only partially shielded by the protein, the crystal-packing forces may distort the RNA in order to form a stable structure. In turn, there is also the concern that the crosslinking sites of an RNA and a protein may not describe the normal binding mode of the complex, but instead report on a transient species trapped by the photochemistry. Even worse, in an example such as this one, where a mutant RNA was used to map the crosslinking site, the result may describe a state that normally is not highly populated. In the case of the 102A:RNA complex, we can use these two complementary methods in a direct comparison, to ask if they describe the same interaction. Because the solution conditions include $MgCl_2$, where the co-crystal does not, there is the chance that the structure of the RNA loop, which is sensitive to the Mg^{2+} concentration (data not shown), will influence the contacts it makes with the protein surface. A comparison of the two methods is therefore useful.

Crosslink and co-crystal

In the 102A:RNA complex, Tyr 13 and U13 must be within a few angstroms for crosslinking to occur. As shown in the co-crystal (Oubridge et al., 1994), Tyr 13 is stacked on the 5′ side of C13 (using our RNA numbering system); the presumption we make is that it is also found in this orientation in the complex with the C13U RNA. There are no hydrogen bonds between Tyr 13 and C13 in the crystal, so the C13U substitution would not lose any specific interactions there, and the stacking should not be disrupted. This hypothesis is supported by the binding data; at 4 °C, where the crosslinking reactions are done, the binding constant of the 102A:C13U complex is identical to that with the wild-type RNA sequence, at least in 200 mM NaCl, 1 mM $MgCl_2$, pH 6 (data not shown). The temperature dependence of this association suggests that, at 4 °C, the complex is able to form efficiently with both wild-type and mutant RNA. However, we note that at higher temperatures, the disparity in the binding constants becomes greater, and there is a higher entropic cost for association of the RBD with the C13U RNA (data not shown). Thus, although the structure in this region that is found in the co-crystal would predict that this crosslink could be formed, its formation is dependent on a stable association that keeps these two aromatic groups in the appropriate geometric orientation. If the crosslinking were done at 37 °C, for example, we predict that the efficiency would be much lower, due

to motion of the RNA and protein side chains. However, the crosslinking data do show that the Tyr 13 and U13 proximity and geometry that are observed in the structure of the co-crystal are preserved in solution, at least at low temperatures.

The C13U RNA is the only variant that produces a crosslink to the RBD. The uridines normally found in the RNA sequence, U10, U11, and U16, are apparently not in the right orientation with an appropriate amino acid nearby to produce a crosslink. This observation is suggested by the co-crystal results, although not all amino acids surrounding these bases are shown (Oubridge et al., 1994); a more detailed description of their environments will provide a useful comparison. However, the last three bases at the 3′ side of the loop (U16, C17, and C18) are not in contact with the RBD because the binding data for mutations at these sites shows that they can be readily substituted without interfering with the binding affinity (Fig. 1). In the structure of the co-crystal, these three bases point away from the protein surface.

Role of Tyr 13 in RNA binding

For these biochemical experiments, we have substituted Tyr 13 with Phe 13 and Thr 13 in order to assess the contribution of Tyr 13 to RNA binding. These protein mutation experiments can be difficult to interpret because amino acid substitutions can destabilize the protein or perturb its structure, and therefore impair its ability to bind RNA. However, Tyr 13 Phe and Tyr 13 Thr in 95A do not greatly destabilize the RBD: in particular, the small loss of stabilization free energy (0.3 kcal/mol) for the Tyr 13 Phe mutant suggests that it has not suffered a large perturbation in its structure. Tyr 13 was shown in the 95A crystal structure to form a hydrogen bond with its hydroxyl to Gln 54 (Jessen et al., 1991); however, the loss of this interaction in the Phe 13 RBD has minimal energetic consequences for the protein stability.

Neither the Tyr 13 Phe RBD nor the Tyr 13 Thr RBD, despite their stability and structural integrity, binds RNA as tightly as does the wild-type RBD (Table 1). The 300× loss of affinity ($\Delta\Delta G^0 = 2.9 \pm 0.3$ kcal/mol) observed with the Phe 13 RBD may suggest that several contacts in the complex have been lost, but could also report on a loss of favorable enthalpy or an increase of unfavorable entropy, for example, if there has been a change in the number or positions of water molecules associated with complex formation. In the co-crystal, the hydrogen bond between the hydroxyl of Tyr 13 and the side chain of Gln 54 is maintained in the complex, where it is presumed to constrain the conformation of Gln 54. In the Phe 13 RBD interaction with RNA, this hydrogen bond to Gln 54 is absent, of course; its absence may, as suggested (Oubridge et al., 1994), allow Gln 54 to form other interactions. However, this Phe 13 substitution does not abolish RNA binding, contrary to prediction (Oubridge et al., 1994), although it reduces the affinity. The observation of a crosslink from U13 to Phe 13 is conclusive evidence that these residues are in close proximity, most likely in the same relative orientation as in the Tyr 13 RBD, given the identical crosslinking efficiency. Phe 13 is thus able to maintain the stacking interaction with the RNA in the absence of the anchoring hydrogen bond found in Tyr 13.

Role of the COOH-terminus in RNA binding

One detail of these mutant RBDs needs to be discussed. These constructs are in the shorter form of the protein, terminating at amino acid 95. The 12 carboxyl-terminal amino acids of the RBD (from 90 to 102) have been implicated in association with RNA. Truncation of these residues either reduces the affinity of the protein for the RNA, such as with this 95A construct where binding affinity was 30× weaker than for the 102A (Table 1), or in the truncated Ala 96 RBD and the Lys 98 Gln mutant of 102A (Nagai et al., 1990), or abolishes it when the RBD was terminated at Ser 91 (Scherly et al., 1989). Assuming that the substitutions or truncations did not destabilize the protein structure (the 95A stability is lower than the 102A by about 1 kcal/mol; data not shown), these results suggest that these amino acids participate in interactions with the RNA. In the co-crystal, there are some contacts between amino acids 91 and 92 and the RNA (Oubridge et al., 1994), and NMR data show that these amino acids are sensitive to RNA binding (Howe et al., 1994). It is clear that the seven amino acids from 96 to 102 are not necessary for RNA binding, however, nor are they required for specific crosslinking because the 102A and 95A RBDs crosslink with equal efficiency to the C13U RNA (Table 1). This end of the RBD is implicated in RNA binding of hnRNP C (Gorlach et al., 1994), and earlier experiments with U1A suggested that sequences in this region, and beyond, were important for RNA binding (Scherly et al., 1991). These sequences clearly deserve more careful biochemical study.

Specificity, affinity, and dynamics

The nonpolar residues in the RNP1 consensus (β3) and the RNP2 consensus (β1) are conserved among RBD proteins as, for example, there are usually aromatic amino acids at positions analogous to Tyr 13 and Phe 56 (Birney et al., 1993). In the co-crystal structure, neither of these residues forms a direct contact with the RNA, supporting the idea that these conserved amino acids are not involved in binding specificity. We hypothesize that one role of these nonspecific interactions may be to provide the geometric orientation and the initial

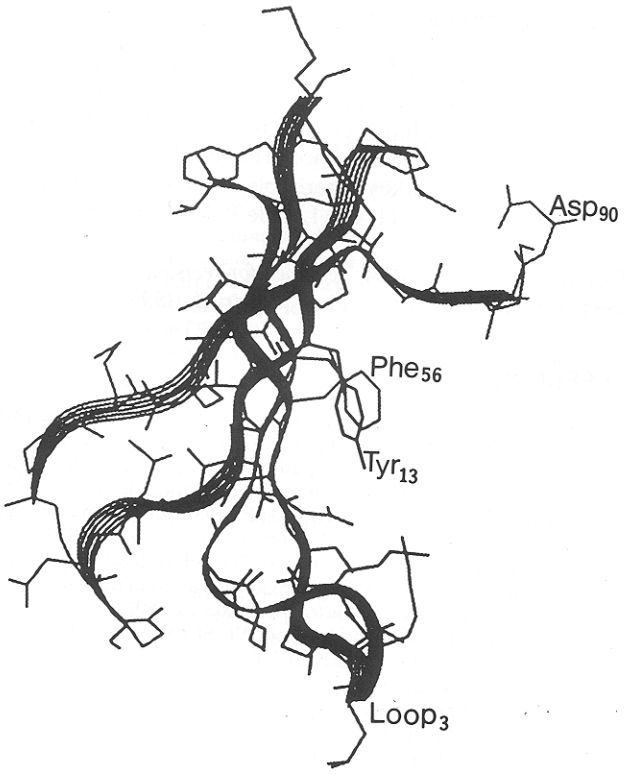

FIGURE 5. Reconstruction of the tertiary structure of the 95A protein, showing the surface of the β-sheet with the amino acid side chains visible (courtesy Dr. C. van Gelder). Tyr 13 is labeled, which is the site of crosslinking to the RNA hairpin.

binding free energy for the RNA to allow other specific interactions between RNA and protein to form. As the structure of the RBD shows, the concavity of the domain folds the nonconserved amino acids of the β2/β3 loop and the carboxyl-terminus over the surface of the β-sheet where the RNA lies. On this surface, as Figure 5 illustrates (van Gelder et al., 1994), both Tyr 13 and Phe 56 are solvent accessible, which is rather unusual for a phenylalanine residue (most of the conserved phenylalanines in the RBD form the hydrophobic core of the protein). These two aromatic residues are in a position to use stacking interactions to hold the RNA in place for subsequent contacts to be formed.

There remain intriguing questions of how the RBD selects its RNA target. As these results with mutant RNAs and proteins show, the network of interactions is pliable and can tolerate some loss of contacts. However, as binding of other RNA mutants to the RBD has also demonstrated (Hall, 1994), the interaction surface is not infinitely forgiving, and some losses cannot be compensated. Understanding the intricate relation of intramolecular and intermolecular contacts that comprise this binding surface, and understanding how this binding surface compares to those of the other 70 (potential) RBDs, will be a start to the goal of predicting and modifying the affinity and specificity of RNA binding domains.

MATERIALS AND METHODS

Protein purification

Amino acids 1–102 of the human U1A protein were subcloned into a plasmid for overexpression in *Escherichia coli*, as described previously (Hall & Stump, 1992). The protein was isolated by French-pressing the cells, followed by ammonium sulfate precipitation, and subsequent purification by cation-exchange chromatography. The protein is a monomer in solution, and when stored in 3 mM sodium cacodylate, pH 6, 50 mM NaCl, 10% glycerol, at 4 °C, is stable for months. Protein mutants were constructed by PCR oligo-directed mutagenesis of the shorter 1–95 RBD. All constructs were sequenced after introduction into the ptac vector used for overproduction (Hall & Stump, 1992).

RNA synthesis

RNA hairpins were enzymatically synthesized by either T7 or SP6 RNA polymerase from DNA oligonucleotides, as described (Milligan et al., 1987; Stump & Hall, 1993). The sequence of the wild-type and C13U sequence are shown in Figure 1. The C13U RNA was transcribed by T7 RNA polymerase; when 5-I-uridine (Sigma) was incorporated, the transcription mix contained 100% iodinated nucleotide, which had no effect on the transcription efficiency. The RNA was internally labeled with α-^{32}P-ATP (NEN) at varying specific activities, or doubly labeled with α-^{32}P-ATP and α-^{32}P-CTP for binding assays. Those reactions containing 5-I-U and subsequent purification of the RNA were done in reduced light. RNA products were used in crosslinking or binding assays within 24 h of their synthesis, to reduce the amount of nonspecific degradation (presumably due to the I-U).

Photochemical crosslinking

Typically, 20-μL samples containing RNA and protein were placed in 1.5-mL microtube caps and preincubated 20 min at 4 °C in 25 mM sodium cacodylate, pH 6, 50 mM NaCl, 1 mM $MgCl_2$, including 20 μg/mL BSA and 50 μg/mL tRNA. Exposure to the ultraviolet light source (Spectroline, λ_{max} = 312 nm) was at a distance of 2.5 cm, filtered through a polystyrene petri dish; light source and samples were in a refrigerator. Products were separated on 15% SDS-polyacrylamide gels with a 7% stacking gel.

In experiments designed to compare crosslinking efficiency, RNA and protein concentrations were 0.1 μM and 1 μM, respectively. For these experiments, 8 μL of sample was mixed with 8 μL of 2× Laemmli loading buffer and applied to the gel. Efficiencies were calculated from measuring the radiolabeled RNA free and bound to the protein by directly scanning the gel (Betagen Betascope Blot Reader, model 603).

Nuclease mapping

For the larger preparations used in localizing the RNA attachment site, BSA and tRNA were omitted from the crosslink-

ing reaction and the RNA and protein concentrations were increased 10×. Samples were illuminated for 20 min, and the α-^{32}P-ATP C13U:102A complex and the irradiated free RNA were recovered from an SDS-PAGE gel by electroelution (using a Schleicher & Schuell electroeluter) into TBE overnight in the dark. After concentration of the sample by speedvac at low temperatures, the samples were immediately digested with nuclease. To map the site on the RNA that was covalently bound to the protein, the bound RNA was digested with either RNase T1, which cleaves after rG, or with RNase A, which cleaves after pyrimidines. Complete digestion of 6,000 cpm of bound RNA was accomplished in 1 h at 65 °C with 10 units of RNase T1 (Boehringer Mannheim) or 5 μg of RNAse A (US Biochemical) in 10 μL Tris-EDTA, pH 8. The temperature of the digestion mix was slowly raised to 85 °C, at which point 10 μL of 7 M urea/1× TBE was added and the samples were loaded on a warm prerun 20% urea/polyacrylamide gel. Control samples contained nonirradiated 5-I-U-C13U RNA both intact and digested with each nuclease. Bands were visualized by autoradiography.

Sequencing

To prepare the complex for amino acid sequencing, the reactions included 10 μM RNA and 100 μM protein, again omitting BSA and tRNA. After 20 min irradiation, followed by electrophoretic separation of the complexed species, the samples were electroblotted onto a Millipore Immobilon membrane in CAPS buffer, using a Hoeffer apparatus. The blot was rinsed in water 1 min, dried, and the band corresponding to the crosslinked complex was cut out. Sequencing from this blot was done by automated Edman degradation on an Applied Biosystems sequencer (David McCourt, Midwest Analytical Inc., St. Louis, Missouri).

Binding assays

RNAs used for binding assays were enzymatically synthesized using T7 or SP6 RNA polymerase. Transcribed RNAs were internally α-^{32}P-ATP and α-^{32}P-CTP labeled in reactions containing 1 mM GTP and UTP, and 0.1 mM ATP and CTP, to produce high specific activity RNAs. All RNAs were purified by denaturing gel electrophoresis. Nitrocellulose filter binding experiments were done in 200 mM NaCl, 10 mM sodium cacodylate, pH 6, 1 mM $MgCl_2$, 20 μg/mL BSA, 0.5 μg/μL tRNA, 22 °C, with a 20-min preincubation followed by filtration, using a Schleicher & Schuell 0.2-μ supported nitrocellulose membrane presoaked in the binding buffer in a modified dot-blot apparatus (Wong & Lohman, 1993). The membrane was lightly blotted with a Kimwipe after soaking to remove the excess buffer (without such blotting, the samples applied tended to diffuse over the surface). After filtering, the underside of the membrane was blotted again to remove excess radioactive buffer that diffused across the surface. 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. Each experiment was done at least twice. The error is typically about 2–3× in the values of the binding constants and about 20% in the calculation of free energies.

ACKNOWLEDGMENTS

We thank J. Kranz for useful suggestions, his stability data for the Tyr 13 mutants, and his contribution of the Tyr 13 Phe protein for crosslinking experiments, and Prof. Tim Lohman for his insights and advice. This work was supported by the Lucille P. Markey Charitable Trust (#90-47) and the NIH (GM46318). K.B.H. is a Lucille P. Markey Scholar.

Received December 1, 1994; returned for revision December 19, 1994; revised manuscript received January 4, 1995

REFERENCES

Bandziulis RJ, Swanson MS, Dreyfuss G. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev 3*:431–437.

Bentley R, Keene JD. 1991. Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B″ protein and through interactions with U2 snRNP-A′ protein. *Mol Cell Biol 11*:1829–1839.

Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res 21*:5803–5816.

Boelens W, Jansen EJR, van Venrooij WJ, Stripecke R, Mattaj IW, Gunderson SI. 1993. The U1 snRNP-specific U1A protein inhibits its polyadenylation of its own pre-mRNA. *Cell 72*:881–892.

Boelens WC, Scherly D, Jansen EJR, Kolen K, Mattaj IW, van Venrooij W. 1991. Analysis of in vitro binding of U1A protein mutants to U1 snRNA. *Nucleic Acids Res 19*:4611–4618.

Dreyfuss G, Matunis MJ, Pinol-Roma S, Burd C. 1993. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem 62*:289–321.

Dreyfuss G, Swanson MS, Pinol-Roma S. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci 13*:86–91.

Garrett DS, Lodi PJ, Shamoo Y, Williams K, Clore GM, Gronenborn A. 1994. Determination of the secondary structure and folding topology of an RNA binding domain of mammalian hnRNP A1 protein using three-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry 33*:2852–2858.

Ghetti A, Padovani C, Di Cesare G, Morandi C. 1989. Secondary structure prediction for RNA binding domain in RNP proteins identifies $\beta\alpha\beta$ as the main structural motif. *FEBS Lett 257*:373–376.

Gorlach M, Burd CG, Dreyfuss G. 1994. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem 269*:23074–23078.

Gorlach M, Wittekind M, Beckman RA, Mueller L, Dreyfuss G. 1992. Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J 11*:3289–3295.

Gott JM, Willis MC, Koch TH, Uhlenbeck OC. 1991. A specific, UV-induced RNA–protein crosslink using 5-bromouridine-substituted RNA. *Biochemistry 30*:6290–6295.

Hall KB. 1994. Interactions of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry 33*:10076–10088.

Hall KB, Stump WT. 1992. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res 20*:4283–4290.

Hoffman DW, Query CC, Golden BL, White SW, Keene JD. 1991. RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc Natl Acad Sci USA 88*:2495–2499.

Howe PWA, Nagai K, Neuhaus D, Varani G. 1994. NMR studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J 13*:3873–3881.

Jessen TH, Oubridge C, Teo CH, Pritchard C, Nagai K. 1991. Identification of molecular contacts between the U1A small nuclear ribonucleoprotein and U1 RNA. *EMBO J 10*:3447–3456.

Keene JD, Query CC. 1991. Nuclear RNA-binding proteins. *Prog Nucleic Acid Res 41*:179–202.

Kenan DJ, Query CC, Keene JD. 1991. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem Sci 16*:214–220.

Lee AL, Kanaar R, Rio DC, Wemmer DE. 1994. Resonance assignments and solution structure of the second RNA-binding domain

of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry 33*:13775–13786.

Lutz-Freyermuth C, Query CC, Keene JD. 1990. Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc Natl Acad Sci USA 87*:6393–6397.

Mattaj IW. 1989. A binding consensus: RNA–protein interactions in splicing, snRNPs, and sex. *Cell 57*:1–3.

Merrill BM, Stone KL, Cobianchi F, Wilson SH, Williams KR. 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J Biol Chem 263*:3307–3313.

Milligan JF, Groebe DR, Witherall GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res 15*:8783–8798.

Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature 348*:515–520.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.

Query CC, Bentley R, Keene J. 1989. A common RNA recognition motif identified within a defined U1 RNA-binding domain of the 70K U1 snRNP protein. *Cell 57*:89–101.

Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990. Major determinants of the specificity and interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. *Nature 345*:502–506.

Scherly D, Boelens W, van Venrooij WJ, Dathan NA, Hamm J, Mattaj IW. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J 8*:4163–4170.

Scherly D, Kambach C, Boelens W, van Venrooij WJ, Mattaj IW. 1991. Conserved amino acids within and outside of the N-terminal RNP motif of the U1A snRNP protein involved in U1 RNA binding. *J Mol Biol 219*:577–584.

Stump WT, Hall KB. 1993. SP6 RNA polymerase efficiently synthesizes RNA from short dõuble-stranded DNA templates. *Nucleic Acids Res 21*:5480–5484.

van Gelder CW, Gunderson SI, Jansen E, Boelens W, Polycarpou-Schwartz M, Mattaj IW, van Venrooij WJ. 1993. A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A is required for regulation of polyadenylation. *EMBO J 12*:5191–5200.

van Gelder CWG, Leusen FJJ, Leunissen JAM, Noordik JH. 1994. A molecular dynamics approach for the generation of complete protein structures from limited coordinate data. *Proteins Struct Funct Genet 18*:174–185.

Willis MC, Hicke BJ, Uhlenbeck OC, Cech TR, Koch TH. 1993. Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science 262*:1255–1257.

Wittekind M, Gorlach M, Freidrichs M, Dreyfuss G, Mueller L. 1992. ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry 31*:6254–6265.

Wong I, Lohman TL. 1993. A double-filter method for nitrocellulose-filter binding: Application to protein–nucleic acid interactions. *Proc Natl Acad Sci USA 90*:5428–5432.