© 2002 Oxford University Press
Nucleic Acids Research, 2002, Vol. 30, No. 23 5269–5275

Substitution of an essential adenine in the U1A–RNA complex with a non-polar isostere

Jacob B. Tuite, Jerome C. Shiels and Anne M. Baranger*

Department of Chemistry, Wesleyan University, Middletown, CT 06459, USA

Received June 28, 2002; Revised and Accepted September 24, 2002

ABSTRACT

The RNA recognition motif (RRM) binds to single-stranded RNA target sites of diverse sequences and structures. A conserved mode of base recognition by the RRM involves the simultaneous formation of a network of hydrogen bonds with the base functional groups and a stacking interaction between the base and a highly conserved aromatic amino acid. We have investigated the energetic contribution of the functional groups involved in the recognition of an essential adenine, A6, in stem–loop 2 of U1 snRNA by the N-terminal RRM of the U1A protein. Previously, we found that elimination of individual hydrogen bond donors and acceptors on A6 destabilized the complex by 0.8–1.9 kcal/mol, while mutation of the aromatic amino acid (Phe56) that stacks with A6 to Ala destabilized the complex by 5.5 kcal/mol. Here we continue to probe the contribution of A6 to complex stability through mutation of both the RNA and protein. We have removed two hydrogen-bonding functional groups by introducing a U1A mutation, Ser91Ala, and replacing A6 with tubercidin, purine, or 1-deazaadenine. We find that the complex is destabilized an additional 1.2–2.6 kcal/mol by the elimination of the second hydrogen bond donor or acceptor. Surprisingly, deletion of all of the functional groups involved in hydrogen bonds with the U1A protein by substituting adenine with 4-methylindole reduced the binding free energy by only 2.0 kcal/mol. Experiments with U1A proteins containing mutations of Phe56 suggested that improved stacking interactions due to the greater hydrophobicity of 4-methylindole than adenine may be partly responsible for the small destabilization of the complex upon substitution of 4-methylindole for A6. The data imply that hydrophobic interactions can compensate energetically for the disruption of the complex hydrogen-bonding network between nucleotide and protein.

INTRODUCTION

RNA forms complex secondary and tertiary structures that provide diverse targets for RNA-binding proteins (1). Specific recognition of RNA by proteins occurs primarily at non-helical structures because the major groove of the A-form RNA helix is narrow and deep, limiting access by proteins, while other RNA structures present unique arrays of functional groups for sequence- and structure-specific binding (2). Since the RNA is not helical, interactions between the protein and RNA may be formed using functional groups that would normally be involved in maintaining the structure of the helix. For example, the base functional groups involved in Watson–Crick hydrogen bonds in the helix are available for hydrogen bonding to the protein and RNA bases are able to participate in stacking interactions with aromatic amino acid side chains without the energetic penalty of helix disruption. Thus, the complexes formed between RNA and proteins can involve more varied and extensive networks of interactions than are found in complexes formed between helical DNA and proteins (3,4).

The RNA recognition motif (RRM), also called the RNA-binding domain (RBD) or the RNA nucleoprotein (RNP) domain, is a ubiquitous RBD found in many proteins involved in gene expression (5). The RRM is comprised of 90–100 amino acids that form an antiparallel beta sheet supported by two alpha helices (6). This basic structure is modified by different RRMs to recognize diverse single-stranded RNA targets, often with high specificity. Three of the most highly conserved amino acids that contact RNA are aromatic (7). Structural studies of RRM–RNA complexes have shown that these conserved aromatic residues participate in stacking interactions with RNA bases, suggesting that stacking interactions are important contributors to non-specific RNA binding by the RRM (8–15). The bases that stack with the conserved aromatic amino acids also form hydrogen bonds with the RRM and often stack with an adjacent base. As a result, these bases contact the RRM using both highly conserved stacking interactions and hydrogen bond networks unique to each particular RRM–RNA complex.

U1A is a protein component of U1 snRNP, a subunit of the spliceosome that splices most eukaryotic pre-mRNA (16). The U1A protein contains two RRMs, but only the N-terminal RRM binds RNA (17,18). The U1A protein binds with high affinity and specificity to three target sites of nearly identical sequence, stem–loop 2 in U1 snRNA and two internal loop target sites in the 3′-untranslated region of the pre-mRNA of

*To whom correspondence should be addressed. Tel: +1 860 685 2739; Fax: +1 860 685 2211; Email: abaranger@wesleyan.edu

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

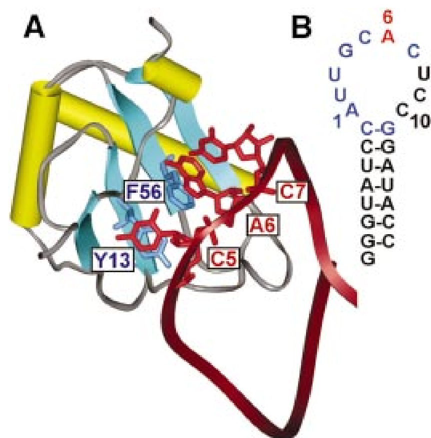

Figure 1. (**A**) Diagram of the complex formed between the N-terminal RRM of the U1A protein and stem–loop 2 RNA from the X-ray cocrystal structure (8). (**B**) Stem–loop 2 RNA used in these experiments. The nucleotides that form the binding site for the U1A protein are shown in blue and the adenine that stacks with Phe56 is shown in red.

the U1A protein (18,19). The U1A protein primarily associates with the loop regions of these target sites (Fig. 1) (8,9,20). The U1A protein has been structurally characterized alone and bound to stem–loop and internal loop target sites (8,9,20,21). The U1A protein contains two of the three highly conserved aromatic amino acids found in RRMs, Tyr13 and Phe56. Tyr13 stacks with C5 and Phe56 stacks with A6 in the RNA target sites (Fig. 1) (8,9,20). Both C5 and A6 are involved in numerous hydrogen bonds with the U1A protein. A6 also stacks with C7, while C5 is exposed to water on one face. In the free protein, Phe56 interacts with residues in the C-terminal helix of the U1A protein, and these interactions must be disrupted for the complex to form. Although the loop regions of the free RNA target sites are not well-ordered, molecular dynamics simulations and NMR spectroscopy have suggested that most of the bases are directed towards the inside of the RNA loop and undergo large conformational changes upon binding to the U1A protein in order to be placed external to the loop as they are found in the complex (8,9,20,22–25).

We have studied the binding of A6 in the U1A-stem–loop 2 complex as a model for the recognition of RNA bases by stacking interactions in RRM–RNA complexes. Previously, we determined that the conserved aromatic residue, Phe56, was essential for high binding affinity and contributed indirectly to specific target site recognition (26,27). A6 is also important for the stability of the complex (27–29). Mutation of A6 to any other base results in a 100 000-fold increase in the equilibrium dissociation constant of the complex (27,28). We replaced A6 with modified bases in which individual hydrogen-bonding functional groups were eliminated or replaced with hydrocarbon groups. These modifications led to a destabilization of the complex of between 0.8 and 1.9 kcal/mol (26,27). In addition to forming the interactions observed in the structure of the complex, the conserved aromatic amino acid and the A6 functional groups are likely to contribute to the stability of the free U1A protein and the free RNA target site and to participate in networks of cooperative interactions in both the free and bound components. The previous experiments described above measured all of the contributions of the protein or RNA modification to binding affinity and taken together, suggested that both the conserved aromatic amino acid and the hydrogen bond donor and acceptor groups on A6 are important for complex formation.

In this paper, we have eliminated the hydrogen bond network between A6 and the U1A protein in order to investigate the ability of the conserved stacking interaction alone to stabilize the U1A-stem–loop 2 complex. To do this we have substituted the non-polar base isostere 4-methylindole for A6 in stem–loop 2 RNA. Although the elimination of individual hydrogen-bonding functional groups destabilized the complex by 0.8–1.9 kcal/mol and the simultaneous elimination of a second hydrogen-bonding functional group resulted in a further destabilization of the complex of 1.2–2.6 kcal/mol, the substitution of A6 with 4-methylindole resulted in surprisingly little loss in binding energy, only 2.0 kcal/mol. Despite the modest destabilization of the complex upon incorporation of 4-methylindole into stem–loop 2 RNA, this substitution changed the relative binding affinities of a series of U1A proteins mutated at Phe56 considerably. These experiments suggested that the ability of 4-methylindole to participate in stacking interactions due to its hydrophobicity might compensate for the energetic penalty that results from eliminating functional groups from A6.

MATERIALS AND METHODS

RNA synthesis

1-Deazaadenosine was synthesized by published methods, while purine riboside and tubercidin were purchased from Sigma (30). The phosphoramidites were synthesized by published methods (31). 4-Methylindole deoxynucleoside phosphoramidite was purchased from Glen Research. The RNA stem–loop was chemically synthesized on an Applied Biosystems ABI 394 automated DNA/RNA synthesizer using Pac-A, *i*Pr-Pac-G, Ac-C and U phosphoramidites from Glen Research (31). To prevent oxidation of 4-methylindole, 0.02 M I_2 was used during RNA synthesis. The RNA was deprotected by published methods and purified using denaturing gel electrophoresis as reported previously (27). Concentrations were determined by UV at 260 nm. Correct composition was confirmed by MALDI mass spectrometry and enzymatic hydrolysis.

Peptide synthesis and purification

An expression vector for the N-terminal RRM of the U1A protein, U1A101, was obtained from Nagai (32). Ser91Ala, Thr11Val, Phe56His, Phe56Leu, Phe56Tyr, Phe56Trp and Phe56Ala mutations were introduced using standard Kunkel mutagenesis procedures (33). The proteins were expressed and purified as reported previously (27). The concentration of each protein was determined by amino acid analysis and the molecular weight by MALDI mass spectrometry.

Equilibrium binding assays

Peptide–RNA equilibrium dissociation constants were measured by gel mobility shift assays (27,34). Reactions were equilibrated at 25°C for at least 30 min in 10 mM Tris–HCl

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

(pH 7.4), 250 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 1 mg/ml tRNA, 25 pM ^{32}P-labeled stem–loop 2 RNA in a total volume of 10 µl. After addition of glycerol to a final concentration of 5%, the reactions were loaded on an 8% polyacrylamide gel (80:1 mono:bisacrylamide, 15 × 40 cm × 1.5 mm) in 100 mM Tris–borate (pH 8.3), 1 mM EDTA, and 0.1% Triton X-100 for 1 h at 350 V. The temperature of the gel was maintained at 25°C by a circulating water bath. Gels were analyzed on a Molecular Dynamics Storm phosphorimager. Fraction RNA bound versus protein concentration was plotted and curves were fit to the equation: fraction bound = $1/(1 + K_D/[P])$.

RESULTS

Elimination of two functional groups involved in the hydrogen-bonding network between A6 and the U1A protein

A6 of stem–loop 2 RNA participates in three hydrogen bonds with the U1A protein (Fig. 2). N1 hydrogen bonds to the side chain of Ser91, N7 forms a water-mediated hydrogen bond with Thr89, and the 6-amino group forms a hydrogen bond with the carbonyl of Thr89 (8). Previously we replaced A6 with the modified bases tubercidin, purine and 1-deazaadenine (c^{1}A), shown in Figure 3 (26,27). Two functional groups that form hydrogen bonds between A6 and the U1A protein were simultaneously modified by forming complexes between Ser91Ala and stem–loop 2 target sites containing A6purine or A6tubercidin substitutions. The Ser91Ala mutation should eliminate the hydrogen bond formed between N1 of A6 and the side chain of Ser91. The Ser91Ala-stem–loop 2 complex was destabilized by 2.2 kcal/mol compared with the wild type complex (Table 1). This decrease in binding affinity was nearly identical to that observed for the complex formed between wild type U1A protein and A6c^1 stem–loop 2 RNA (27). An identical binding constant was measured for the complex formed between Ser91Ala and A6c^1 stem–loop 2 RNA. Thus, the same destabilization was observed whether the hydrogen-bonding functional group was eliminated from the protein or the RNA or from both simultaneously.

The hydroxyl group of Ser91 also forms a hydrogen bond with Thr11 (Fig. 2) (8). The decrease in stability of the Ser91Ala-stem–loop 2 complex will have contributions from the elimination of the Ser91–Thr11 hydrogen bond. In order to learn more about the effect of Thr11 on the energetic contribution of A6 functional groups to binding, we substituted Val for Thr11 and measured the affinity of Thr11Val for stem–loop 2 RNAs containing modified adenines at position 6. The results of these experiments are reported in Tables 1 and 2. The Thr11Val-stem–loop 2 complex was 4.5 kcal/mol less stable than the wild type complex, consistent with previous reports (28,35). The Thr11Val-stem–loop 2 complex was destabilized by 1.1 kcal/mol upon substitution of 1-deazaadenine for A6, by 0.7 kcal/mol upon substitution of purine for A6, and by 0.8 kcal/mol upon substitution of tubercidin for A6. Although we do not know from our experiments the effect of the Thr11Val mutation on the structures of the free protein or complex, the destabilization of the complex with each base substitution suggests that each adenine functional group remains important for binding in the Thr11Val-stem–loop 2 complex.

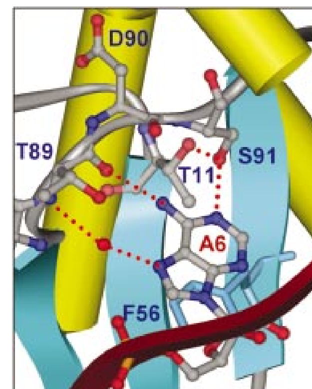

Figure 2. Diagram of the hydrogen bonds formed between A6 and Ser91 and Thr89 (8). The hydrogen bond between Thr11 and Ser91 is also shown.

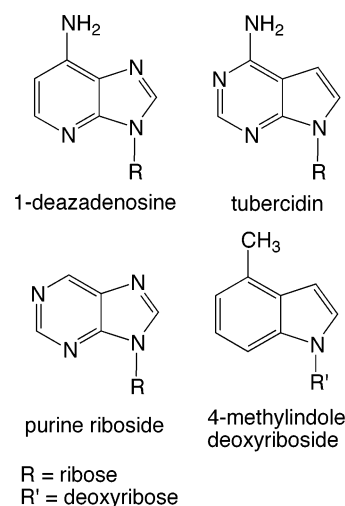

Figure 3. Modified bases incorporated into stem–loop 2 RNA to eliminate hydrogen-bonding functional groups in the U1A-stem–loop 2 complex.

The Ser91Ala-stem–loop 2 complex is destabilized by the incorporation of either tubercidin or purine at position 6 in stem–loop 2 RNA (Tables 1 and 2). Thus, the complexes in which two hydrogen-bonding functional groups have been eliminated are less stable than those in which only one hydrogen-bonding functional group has been eliminated. The destabilization that resulted from elimination of the 6-amino group on A6 was larger in the complex with the wild type protein than in the complex with Ser91Ala (Table 2). In contrast, the destabilization that resulted from elimination of N7 in A6 was smaller in the complex with the wild type protein than in the complex with Ser91Ala. The different magnitudes of destabilization that result from each base substitution in the complexes formed with the wild type and Ser91Ala U1A proteins is not surprising since cooperativity between functional groups in DNA–protein and RNA–protein interfaces is common (36–40). Therefore, the loss in binding free energy observed when two hydrogen-bonding groups are eliminated is rarely equal to the sum of the loss in binding

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

Table 1. Binding affinities of wild type and mutant proteins for wild type and modified stem–loop 2 RNA target sites

RNA		Protein Wild type[b]	Ser91Ala	Thr11Val
Wild type	K_D (M)	5 (±3) × 10^{-10}	2 (±1) × 10^{-8}	9 (±4) × 10^{-7}
	ΔG^0 (kcal/mol)[a]	–12.7 ± 0.3	–10.5 ± 0.3	–8.2 ± 0.3
A6purine	K_D (M)	1.0 (±0.5) × 10^{-8}	1.6 (±0.4) × 10^{-7}	3 (±1) × 10^{-6}
	ΔG^0 (kcal/mol)[a]	–10.9 ± 0.3	–9.3 ± 0.1	–7.5 ± 0.2
A6tubercidin	K_D (M)	2 (±1) × 10^{-9}	1.7 (±0.7) × 10^{-6}	4 (±3) × 10^{-6}
	ΔG^0 (kcal/mol)[a]	–11.9 ± 0.4	–7.9 ± 0.3	–7.4 ± 0.3
A6c[1]	K_D (M)	1.2 (±0.6) × 10^{-8}	1.2 (±0.3) × 10^{-8}	6 (±1) × 10^{-6}
	ΔG^0 (kcal/mol)[a]	–10.8 ± 0.3	–10.8 ± 0.2	–7.1 ± 0.1

[a]ΔG^0 is the free energy of association of the complex.
[b]Previously reported in Shiels *et al.* (27).

Table 2. Comparison of the destabilization ($\Delta\Delta G^0$) that results from A6purine, A6tubercidin, and A6c[1] base substitution in the complexes formed with the wild type, Ser91Ala and Thr11Val U1A proteins

RNA	Wild type[b] $\Delta\Delta G^0$ (kcal/mol)[a]	Ser91Ala $\Delta\Delta G^0$ (kcal/mol)[a]	Thr11Val $\Delta\Delta G^0$ (kcal/mol)[a]
A6purine	1.8	1.2	0.7
A6tubercidin	0.8	2.6	0.8
A6c[1]	1.9	–0.3	1.1

[a]$\Delta\Delta G^0$ is the difference in binding free energy between the complex formed with stem–loop 2 RNA containing the indicated base substitutions and the wild type stem–loop 2 RNA.
[b]Previously reported in Shiels *et al.* (27).

energy observed when each hydrogen-bonding functional group is eliminated individually. In addition, the dependence of the pK_a of N1 on the identity of the ring may contribute to the different effects of the A6 base substitutions on the stabilities of the complexes formed with the wild type and Ser91Ala U1A proteins. The pK_a of N1 in adenine is 3.5, while the pK_a of N1 in purine is 2.1, and the pK_a of N1 in tubercidin is 5.3 (41–44). The correlation between the acidity and basicity of hydrogen-bonding groups and the strength of the hydrogen bond formed has been investigated (45). Hydrogen bond strength is more sensitive to changes in pK_a in an aprotic solvent than in water. Within a folded protein structure the dependence of hydrogen bond strength on pK_a can resemble the dependence observed in aprotic solvents, rather than in water (46). It is likely that the environment around A6 in the U1A-stem–loop 2 complex is less polar than water. In the complex formed with wild type U1A protein, the incorporation of tubercidin should increase the strength of the hydrogen bond formed with N1, while incorporation of purine should decrease the strength of the hydrogen bond formed with N1. If a hydrogen bond is not formed with N1 in the complex with Ser91Ala, the pK_a of N1 should not have a large effect on binding affinity. Therefore, there is a correlation between the pK_a of N1 and the different destabilizations observed when purine and tubercidin were substituted for A6 in the complexes formed with the wild type and Ser91Ala proteins.

Substitution of the non-polar base isostere 4-methylindole for A6

All of the functional groups on A6 that hydrogen bond to the U1A protein were eliminated by substituting 4-methylindole for A6 (Fig. 3). Since 4-methylindole was incorporated as a deoxyriboside (dM), binding studies were performed with deoxyadenosine at position 6 (dA6) to demonstrate that removal of the 2′-OH did not destabilize the wild type complex (Table 3). The complex formed between wild type U1A protein and A6dM stem–loop 2 RNA was 2.0 kcal/mol less stable than that formed between wild type U1A protein and dA6 stem–loop 2 RNA. The substitution of Ala for Ser91 should be less destabilizing for the complex formed with A6dM stem–loop 2 RNA than for that formed with dA6 stem–loop 2 RNA because 4-methylindole cannot hydrogen bond to the hydroxyl group of Ser91. Indeed, Ser91Ala bound with identical affinity to A6dM and dA6 stem–loop 2 RNAs.

Since A6 stacks with Phe56 and C7 in the U1A–RNA complex, substitution of the hydrophobic 4-methylindole for adenine may change the interactions of this residue with Phe56 and C7 (Fig. 1). Previously we investigated the contribution of Phe56 to complex stability by substituting Trp, Tyr, His, Leu and Ala for Phe56 (26,27). Here we measured the ability of these mutant proteins to bind to stem–loop 2 RNA containing 4-methylindole at position 6 (Tables 3 and 4). The stabilities of the mutant protein-A6dM stem–loop 2 complexes were compared with the stabilities of the mutant protein-dA6 stem–loop 2 complexes (Table 3). The binding affinities of the mutant peptides for the dA6 and wild type stem–loop 2 target sites were similar (26,27). The Phe56Trp, Phe56Tyr and Phe56His proteins bound with as high affinity or nearly as high affinity as the wild type protein to dA6 stem–loop 2 RNA (Table 4). In contrast, the complexes of the Phe56Leu and Phe56Ala proteins were destabilized by 3.8 and 5.5 kcal/mol, respectively, compared with the wild type complex. The substitution of adenine with 4-methylindole changed the relative binding affinities of the proteins mutated at position 56 (Table 4). Only the destabilization resulting from substituting Phe56 with Tyr was comparable in the complexes of dA6 and A6dM stem–loop 2 RNAs. Substitution of Trp for Phe56 resulted in a 1.4 kcal/mol destabilization of the complex with A6dM stem–loop 2 RNA, while Phe56Trp bound dA6 stem–loop 2 RNA with as high affinity as the wild type protein. Substitution of Phe56 with His was also more destabilizing for the complex formed with A6dM than dA6 stem–loop 2 RNAs. In contrast, substitution of Phe56 with either Ala or Leu was less destabilizing for the complex formed with A6dM than dA6 stem–loop 2 RNAs. As a result, Phe56Leu bound with as high affinity as the wild type U1A protein to A6dM stem–loop 2 RNA.

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

Table 3. Binding affinities of wild type and mutant U1A proteins for stem–loop 2 RNAs containing dA6 and A6dM substitutions

Protein	RNA dA6 K_D (M)	ΔG^0 (kcal/mol)[a]	A6dM K_D (M)	ΔG^0 (kcal/mol)[a]	$\Delta\Delta G^0$ (kcal/mol)[b]
Wild type	$7 (\pm 3) \times 10^{-10}$	–12.5 ± 0.3	$1.8 (\pm 0.7) \times 10^{-8}$	–10.5 ± 0.3	2.0
Ser91Ala	$7 (\pm 4) \times 10^{-8}$	–9.8 ± 0.4	$6 (\pm 5) \times 10^{-8}$	–9.8 ± 0.5	0
Phe56Trp	$5 (\pm 1) \times 10^{-10}$	–12.7 ± 0.1	$2 (\pm 1) \times 10^{-7}$	–9.1 ± 0.3	3.6
Phe56Tyr	$1.0 (\pm 0.8) \times 10^{-8}$	–10.9 ± 0.3	$4 (\pm 2) \times 10^{-7}$	–8.7 ± 0.2	2.2
Phe56His	$3 (\pm 1) \times 10^{-9}$	–11.6 ± 0.2	$3 (\pm 1) \times 10^{-6}$	–7.5 ± 0.3	4.1
Phe56Leu	$4 (\pm 2) \times 10^{-7}$	–8.7 ± 0.3	$2 (\pm 1) \times 10^{-8}$	–10.5 ± 0.3	–1.8
Phe56Ala	$9 (\pm 5) \times 10^{-6}$	–6.9 ± 0.3	$1.5 (\pm 0.4) \times 10^{-6}$	–7.9 ± 0.2	–1.0

[a]ΔG^0 is the free energy of association of the complex.
[b]$\Delta\Delta G^0$ is the difference in binding free energy between the complexes containing A6dM and dA6 substitutions.

Table 4. Comparison of the destabilization that results from the mutation of Phe56 in the complexes formed with dA6 and A6dM stem–loop 2 RNAs

Protein	RNA dA6 $\Delta\Delta G^0$ (kcal/mol)[a]	A6dM $\Delta\Delta G^0$ (kcal/mol)[a]
Wild type	0	0
Ser91Ala	2.7	0.7
Phe56Trp	-0.2	1.4
Phe56Tyr	1.6	1.8
Phe56His	0.9	3.0
Phe56Leu	3.8	0
Phe56Ala	5.6	2.6

[a]$\Delta\Delta G^0$ is the difference in binding free energy between the indicated mutant U1A protein and the wild type U1A protein.

DISCUSSION

The modifications of A6 and U1A amino acids discussed in this paper are likely to affect the structures and stabilities of the free RNA and the U1A protein as well as the structure and stability of the complex. In addition to the direct effect of the modification on the interactions in which A6 or the U1A amino acid participates, the networks of cooperative interactions present in both the free and bound components should enable the modifications to indirectly affect binding affinity. Since the binding free energy is the difference in the free energies of the free and bound components, the change in binding free energy that results from functional group modification reflects all of the contributions, both direct and indirect, to the free energies of the free and bound components of the complex.

To probe the interdependence of the hydrogen bond donors and acceptors involved in hydrogen bonds between U1A and A6, two hydrogen-bonding functional groups were eliminated simultaneously, one from A6 (A6c[1], A6tubercidin and A6purine) and one from U1A (Ser91Ala). The A6c[1] and Ser91Ala modifications eliminated functional groups involved in the hydrogen bond formed between N1 and Ser91 and there was no additional energetic penalty for incorporating both modifications at the same time. However, simultaneous incorporation of A6purine or A6tubercidin and Ser91Ala resulted in 1.3 and 2.7 kcal/mol greater destabilization of the complex than when either the RNA modification or the protein mutation was incorporated separately (Table 2). The energetic coupling between the hydrogen-bonding functional groups is difficult to estimate because modification of the base changes the pK_a of the remaining functional groups. However, removing the functional groups involved in two hydrogen bonds is more destabilizing than removing those involved in only one hydrogen bond. These results confirm that correct recognition of A6 is essential for the stability of the U1A-stem–loop 2 complex. The sensitivity of binding to modifications of A6 or amino acids that interact with A6 in the complex suggests that the binding reaction is not able to compensate for changes in complex, free RNA, or free protein structure at this position.

The substitution of 4-methylindole for A6 eliminates all of the functional groups on A6 that participate in hydrogen bonds with the U1A protein. 4-Methylindole mimics as closely as possible the size and shape of adenine in order to minimize changes in steric interactions that may be important for binding (47). Since elimination of individual hydrogen-bonding functional groups alone or in pairs significantly destabilized the complex and substitution of any other base for adenine resulted in a 100 000-fold increase of the equilibrium dissociation constant of the complex, it is surprising that substitution of 4-methylindole for A6 only decreased binding affinity by 2 kcal/mol. Similar non-polar base isosteres incorporated opposite a natural base in a DNA helix have been found to be more destabilizing to the helix than a mismatch (48,49). This large destabilization has been proposed to result from the loss of the Watson–Crick hydrogen bonds and from the energetically unfavorable desolvation of the polar base, which cannot form any hydrogen bonds with the non-polar base isostere in the helix. In contrast, the substitution of 4-methylindole for A6 is considerably less destabilizing than any of the natural base substitutions. Unfavorable desolvation of the amino acids that form hydrogen bonds with A6 in the wild type complex may be reduced in the U1A-A6dM stem–loop 2 complex by the greater accessibility to water and the greater flexibility of the U1A amino acids than DNA bases in a helix.

Changes in the free energies of either the RNA or the complex as a result of the substitution of 4-methylindole for A6 could compensate for the loss of the functional groups from A6. The contribution of stacking interactions to RNA

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

and DNA helices is often estimated from experiments in which a base is incorporated into a nucleoside that is dangled from the end of the DNA or RNA helix (50–52). Thermal denaturation is then used to measure the effect of the dangling base on the stability of the helix. Since the base has no hydrogen-bonding partner, any increase in stability compared with a duplex without the dangling base is ascribed to stacking interactions. Using this technique, Kool and co-workers (52) compared the stacking interactions of the natural bases and a series of non-polar base analogs. In general, the non-polar base analogs stabilized the helix by more than the natural bases, presumably due to increased hydrophobic effects. In particular, incorporation of 4-methylindole at each end of a helix stabilized the helix by 3.1 kcal/mol, while adenine only stabilized the helix by 2.0 kcal/mol. Since 4-methylindole is removed from water on both faces in the U1A-stem–loop 2 complex, a comparable or greater increase in stacking interactions would be expected in the U1A-stem–loop 2 complex. Since the RNA loop is dynamic in the free RNA (22–24,53), it is likely that 4-methylindole is more effectively removed from water in the complex than in the free RNA. Therefore, increased association of 4-methylindole with the U1A protein due to the greater hydrophobicity of 4-methylindole than adenine may contribute to the unexpected affinity of the A6dM stem–loop 2 RNA for the U1A protein.

The experiments performed with U1A proteins containing mutations at Phe56, the residue that stacks with A6, largely support the compensation of the missing hydrogen-bonding functional groups by hydrophobic effects in the U1A-A6dM stem–loop 2 complex (Tables 3 and 4). Phe56Leu bound with 1.8 kcal/mol higher affinity to A6dM stem–loop 2 RNA than to dA6 stem–loop 2 RNA, implying the hydrophobic residue Leu associates more tightly with the non-polar 4-methylindole than the polar adenine. Phe56Ala also bound with higher affinity to A6dM than dA6 stem–loop 2 RNA, although the increase in binding affinity was less than that observed for Phe56Leu. In contrast, Phe56His bound with 4.1 kcal/mol lower affinity to A6dM stem–loop 2 RNA than to dA6 stem–loop 2 RNA, which suggests that placing the hydrophobic 4-methylindole next to the polar His is energetically unfavorable. However, the substitution of 4-methylindole for adenine was also destabilizing for the Phe56Trp-stem–loop 2 complex. Therefore, the destabilization of the Phe56His- and Phe56Trp-A6dM stem–loop 2 RNA complexes may arise from other sources besides stacking ability and hydrophobicity.

The results reported in this paper, along with those we reported previously, demonstrate that the binding of the U1A protein to stem–loop 2 RNA is sensitive to modifications of A6 and the amino acids that contact A6 in the complex. Despite this requirement for precise recognition of A6, the non-polar base isostere, 4-methylindole, may be substituted for A6 with relatively little loss in binding affinity. The experiments performed with U1A proteins containing Phe56 mutations suggested that hydrophobic effects compensate, in part, for the loss of the functional groups on A6. Since most RRMs do not bind RNA with as high affinity or specificity as the N-terminal RRM of the U1A protein, the individual functional groups that participate in hydrogen bonds are likely to contribute less binding energy than those we have eliminated from A6 and the U1A protein. As a result, the incorporation of non-polar base isosteres may be less disruptive to most RRM–RNA complexes than we observed for the U1A-stem–loop 2 complex. In addition, other non-polar base analogs, especially those with extended aromatic rings, may form stronger interactions with RRMs than 4-methylindole. Thus, our experiments using the U1A–RNA complex as a model system suggest that improving stacking interactions by increasing the hydrophobicity of the aromatic groups may provide a general approach for stabilizing RRM–RNA complexes.

ACKNOWLEDGEMENTS

We are grateful to Prof. K. Nagai for the expression vector for U1A101. Funding was provided by NIH, GM-56857. A.M.B. is an Alfred P. Sloan Research Fellow. J.C.S. and J.B.T. were supported by a NIH Training Grant in Molecular Biophysics (GM-08271).

REFERENCES

1. Varani,G. (1997) RNA–protein intermolecular recognition. *Acc. Chem. Res.*, **30**, 189–195.
2. Hermann,T. and Westhof,E. (1999) Non-Watson–Crick base pairs in RNA–protein recognition. *Chem. Biol.*, **6**, R335–R343.
3. Antson,A.A. (2000) Single stranded RNA binding proteins. *Curr. Opin. Struct. Biol.*, **10**, 87–94.
4. Draper,D.E. (1999) Themes in RNA–protein recognition. *J. Mol. Biol.*, **293**, 255–270.
5. Varani,G. and Nagai,K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 407–445.
6. Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.
7. Birney,E., Kumar,S. and Krainer,A.R. (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.
8. Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and 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.
9. Allain,F.H., Howe,P.W.A., Neuhaus,D. and Varani,G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5774.
10. Deo,R.C., Bonanno,J.B., Sonenberg,N. and Burley,S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, **98**, 835–845.
11. Price,S.R., Evans,P.R. and Nagai,K. (1998) Crystal structure of the spliceosomal U2B"-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645–650.
12. Ding,J., Hayashi,M.K., Zhang,Y., Manche,L., Krainer,A.R. and Xu,R.J. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.*, **13**, 1102–1115.
13. Handa,N., Nureki,O., Kuimoto,K., Kim,I., Sakamoto,H., Shimura,Y., Muto,Y. and Yokoyama,S. (1999) Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579–585.
14. Wang,X. and Tanaka Hall,T.M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nature Struct. Biol.*, **8**, 141–145.
15. Allain,R.H.T., Bouvet,P., Dieckmann,T. and Feigon,J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.*, **19**, 6870–6881.
16. Stark,H., Dube,P., Lührmann,R. and Kastner,B. (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature*, **409**, 539–542.
17. Lu,J. and Hall,K.B. (1995) An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol.*, **247**, 739–752.

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011

18. Scherly,D., Boelens,W., van Venrooij,W.J., Dathan,N.A., Hamm,J. and Mattaj,I.W. (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.
19. van Gelder,C.W.G., Gunderson,S.I., Jansen,E.J.R., Boelens,W.C., Polycarpou-Schwarz,M., Mattaj,I.W. and van Venrooij,W.J. (1993) A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.*, **12**, 5191–5200.
20. Varani,L., Gunderson,S.I., Mattaj,I.W., Kay,L.E., Neuhaus,D. and Varani,G. (2000) The NMR structure of the 38 kDa U1A protein-PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nature Struct. Biol.*, **7**, 329–335.
21. Avis,J.M., Allain,F.H.-T., Howe,P.W.A., Varani,G., Nagai,K. and Neuhaus,D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
22. Blakaj,D.M., McConnell,K.J., Beveridge,D.L. and Baranger,A.M. (2001) Molecular dynamics and thermodynamics of protein–RNA interactions: mutation of a conserved aromatic residue modifies stacking interactions and structural adaptation in the U1A–stem loop 2 RNA complex. *J. Am. Chem. Soc.*, **123**, 2548–2551.
23. Tang,Y. and Nilsson,L. (1999) Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.*, **77**, 1284–1305.
24. Reyes,C.M. and Kollman,P.A. (2000) Structure and thermodynamics of RNA–protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.*, **297**, 1145–1158.
25. Gubser,C.C. and Varani,G. (1996) Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3′-untranslated region and interaction with the U1A protein. *Biochemistry*, **35**, 2253–2267.
26. Nolan,S.J., Shiels,J.C., Tuite,J.B., Cecere,K.L. and Baranger,A.M. (1999) Recognition of an essential adenine at a protein-RNA interface: comparison of the contributions of hydrogen bonds and a stacking interaction. *J. Am. Chem. Soc.*, **121**, 8951–8952.
27. Shiels,J.C., Tuite,J.B., Nolan,S.J. and Baranger,A.M. (2002) Investigation of a conserved stacking interaction in target site recognition by the U1A protein. *Nucleic Acids Res.*, **30**, 550–558.
28. Zeng,Q. and Hall,K.B. (1997) Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA*, **3**, 303–314.
29. Hall,K.B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry*, **33**, 10076–10088.
30. Seela,F., Debelak,H., Usman,N., Burgin,A. and Beigelman,L. (1998) 1-Deazaadenosine: synthesis and activity of base-modified hammerhead ribozymes. *Nucleic Acids Res.*, **26**, 1010–1018.
31. Scaringe,S.A., Francklyn,C. and Usman,N. (1990) Chemical synthesis of biologically active oligoribonucleotides using β-cyanoethyl protected ribonucleoside phosphoramidites. *Nucleic Acids Res.*, **18**, 5433–5441.
32. Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
33. Kunkel,T.A., Bebenek,K. and McClary,J. (1991) Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.*, **204**, 125–139.
34. Rimmele,M. and Belasco,J. (1998) Target discrimination by RNA-binding proteins: Role of the ancillary protein U2A′ and a critical leucine residue in differentiating the RNA-binding specificity of spliceosomal proteins U1A and U2B". *RNA*, **4**, 1386–1396.
35. Jessen,T.-H., Oubridge,C., Teo,C.H., Pritchard,C. and Nagai,K. (1991) Identification of molecular contacts between the U1A small nuclear ribonuleoprotein and U1 RNA. *EMBO J.*, **10**, 3447–3456.
36. Mossing,M.C. and Record,M.T. (1985) Thermodynamic origins of specificity in the *lac* repressor-operator interaction. Adaptability in the recognition of mutant operator sites. *J. Mol. Biol.*, **186**, 295–305.
37. Lesser,D.R., Kurpiewski,M.R. and Jen-Jacobson,L. (1990) The energetic basis of specificity in the *Eco*RI endonuclease-DNA interaction. *Science*, **250**, 776–786.
38. Johansson,H.E., Dertinger,D., LeCuyer,K.A., Behlen,L.S., Greef,C.H. and Uhlenbeck,O.C. (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl Acad. Sci. USA*, **95**, 9244–9249.
39. Kranz,J.K. and Hall,K.B. (1998) RNA binding mediates the local cooperativity between the β-sheet and the C-terminal tail of the human U1A RBD1 protein. *J. Mol. Biol.*, **275**, 465–481.
40. Kranz,J.K. and Hall,K.B. (1999) RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J. Mol. Biol.*, **285**, 215–231.
41. Rosemeyer,H. and Seela,F. (1988) 171. Stereoselective synthesis of pyrrolo[2,3-*d*]pyrimidine α- and β-D-ribonucleosides from anomerically pure D-ribofuranosyl chlorides; solid-liquid phase-transfer glycosylation and ^{15}N-NMR spectra. *Helv. Chim. Acta*, **71**, 1573–1577.
42. Jardetzky,C.D. and Jardetzky,O. (1960) Investigation of the structure of purines, pyrimidines, ribose nucleosides and nucleotides by proton magentic resonance. II. *J. Am. Chem. Soc.*, **82**, 222–229.
43. Christensen,J.J., Rytting,J.H. and Izatt,R.M. (1970) Thermodynamic pK, $\Delta H°$, $\Delta S°$ and $\Delta C_p°$ values for proton dissociation from several purines and their nucleosides in aqueous solution. *Biochemistry*, **9**, 4907–4913.
44. Townsend,L.B. (1975) In Fasman,G.D. (ed.), *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd Edn. CRC Press, Cleveland, Vol. I., pp. 61–62, 388–389.
45. Perrin,C.L. and Nielson,J.B. (1997) 'Strong' hydrogen bonds in chemistry and biology. *Annu. Rev. Phys. Chem.*, **48**, 511–544.
46. Thorson,J.S., Chapman,E., Murphy,E.C., Schultz,P.G. and Judice,J.K. (1995) Linear free energy analysis of hydrogen bonding in proteins. *J. Am. Chem. Soc.*, **117**, 1157–1158.
47. Kool,E.T. (2001) Hydrogen bonding, base stacking and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 1–22.
48. Schweitzer,B.A. and Kool,E.T. (1995) Hydrophobic, non-hydrogen-bonding bases and base pairs in DNA. *J. Am. Chem. Soc.*, **117**, 1863–1872.
49. Kool,E.T., Morales,J.C. and Guckian,K.M. (2000) Mimicking the structure and function of DNA: insights into DNA stability and replication. *Angew. Chem. Int. Ed.*, **39**, 990–1009.
50. Petersheim,M. and Turner,D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp and ACCGGUp. *Biochemistry*, **22**, 256–263.
51. Freier,S.M., Sugimoto,N., Sinclair,A., Alkema,D., Neilson,T., Kierzek,R., Caruthers,M.H. and Turner,D.H. (1986) Stability of XGCGCp, GCGCYp and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry*, **25**, 3214–3219.
52. Guckian,K.M., Schweitzer,B.A., Ren,R.X.-F., Sheils,C.J., Tahmassebi,D.C. and Kool,E.T. (2000) Factors contributing to aromatic stacking in water: evaluation in the context of DNA. *J. Am. Chem. Soc.*, **122**, 2213–2222.
53. Howe,P.W.A., Nagai,K., Neuhaus,D. and 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.

Downloaded from nar.oxfordjournals.org at Periodical Department on July 8, 2011