doi:10.1016/j.jmb.2004.07.029 J. Mol. Biol. (2004) 342, 743–756

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Human UP1 as a Model for Understanding Purine Recognition in the Family of Proteins Containing the RNA Recognition Motif (RRM)

Jeffrey C. Myers and Yousif Shamoo*

Department of Biochemistry and Cell Biology, Rice University, 6100 S. Main Street-MS140, Houston TX 77005, USA

*Corresponding author

Heterogeneous ribonucleoprotein A1 (hnRNP A1) is a prototype for the family of eukaryotic RNA processing proteins containing the common RNA recognition motif (RRM). The region consisting of residues 1–195 of hnRNP A1 is referred to as UP1. This region has two RRMs and has a high affinity for both single-stranded RNA and the human telomeric repeat sequence $d(TTAGGG)_n$. We have used UP1's novel DNA binding to investigate how RRMs bind nucleic acid bases through their highly conserved RNP consensus sequences. Nine complexes of UP1 bound to modified telomeric repeats were investigated using equilibrium fluorescence binding and X-ray crystallography. In two of the complexes, alteration of a guanine to either 2-aminopurine or nebularine resulted in an increase in K_d from 88 nM to 209 nM and 316 nM, respectively. The loss of these orienting interactions between UP1 and the substituted base allows it to flip between *syn* and *anti* conformations. Substitution of the same base with 7-deaza-guanine preserves the O6/N1 contacts but still increases the K_d to 296 nM and suggests that it is not simply the loss of affinity that gives rise to the base mobility, but also the stereochemistry of the specific contact to O6. Although these studies provide details of UP1 interactions to nucleic acids, three general observations about RRMs are also evident: (1) as suggested by informatic studies, main-chain to base hydrogen bonding makes up an important aspect of ligand recognition (2) steric clashes generated by modification of a hydrogen bond donor–acceptor pair to a donor–donor pair are poorly tolerated and (3) a conserved lysine position proximal to RNP-2 ($\mathbf{K}_{106}$-IFVGGI) orients the purine to allow stereochemical discrimination between adenine and guanine based on the 6-position. This single interaction is well-conserved in known RRM structures and appears to be a broad indicator for purine preference in the larger family of RRM proteins.

Keywords: RRM; RNA–protein interactions; hnRNP A1; hydrogen bonding; UP1

Introduction

The RNA recognition motif (RRM), also referred to as the RNA binding domain (RBD), is comprised of ~90 amino acid residues and is made up of a four-stranded β-sheet and two α-helices (Figure 1).[1] RRM-containing proteins are typically involved in all aspects of RNA processing including splicing, alternative splice site selection, transport, and turnover.[2–4] Many proteins such as nucleolin and poly(A)-binding protein have more than one RRM. At present, the InterPro Database at EMBL lists 455 human proteins containing putative RRMs, making it the most common RNA binding motif.[5] Understanding the atomic basis for RRM specificity is an important step towards piecing together the puzzle of how numerous RRM-containing proteins compete for similar RNA sequences during alternative splice site selection and RNA processing. In

Abbreviations used: 2AP, 2-amino-purine; 6-MI, 8-(2-deoxy-β-ribofuranosyl)isoxanthopteridine; 7deazaA, 7-deaza-adenine; 7deazaG, 7-deaza-guanine; ade, adenine; gua, guanine; hnRNP A1, heterogeneous ribonucleoprotein A1; ino, inosine; neb, nebularine; RRM, RNA recognition motif.

E-mail address of the corresponding author: shamoo@rice.edu

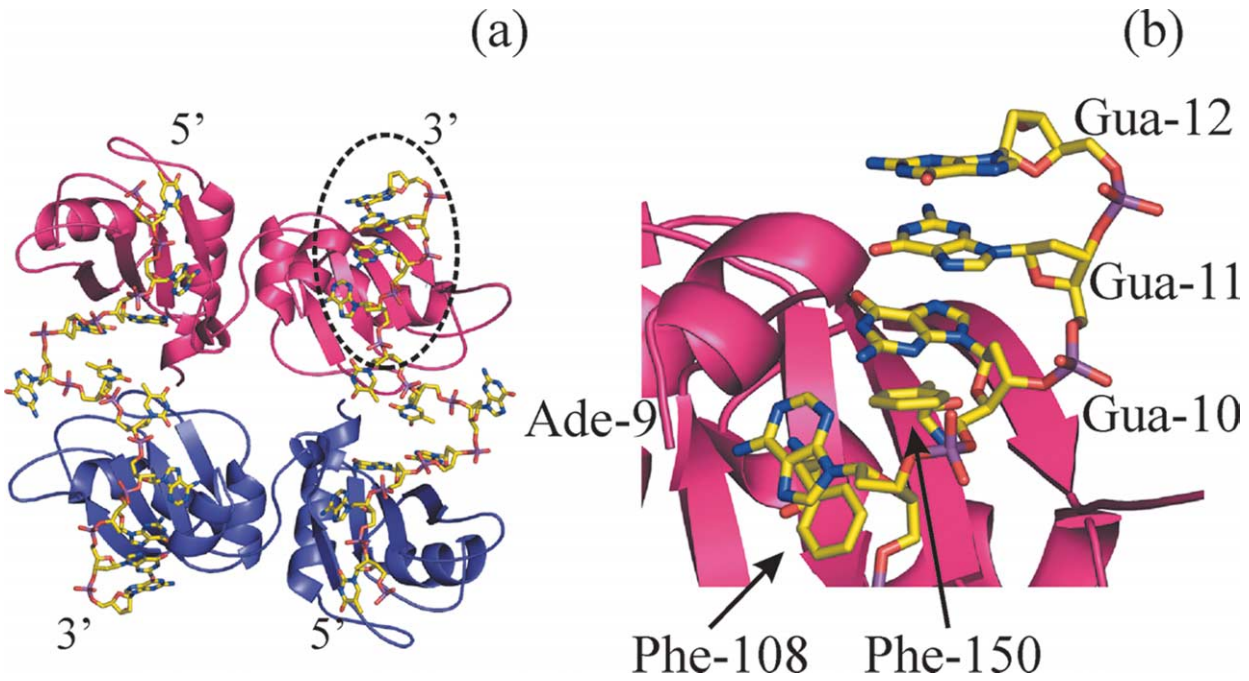

Figure 1. (a) Structure of UP1 bound to the sequence d(TTAGGG)$_2$. Two copies of UP1 form a crystallographic dimer to which the DNA binds in an antiparallel manner to RRM1 (residues 1–92) of the first copy of UP1 and then to RRM2 (residues 93–195) of the second copy of UP1. The circled region highlights the bases selected for these studies. (b) A closer view of RRM2 showing the general positioning of Ade9, Gua10 and Gua11 in relation to the highly conserved phenylalanine residues (Phe108 and 150) of the RNP consensus sequences found in RRM-containing proteins.

addition, a fuller understanding of RRM–RNA interactions will provide the basis for engineering specificity in these motifs and permit their development as tools for investigating patterns of splicing and gene expression.[6]

Heterogeneous ribonucleoprotein A1 (hnRNP A1) is a member of the abundant hnRNP family of nuclear proteins.[7,8] HnRNP A1 has multiple roles *in vivo* and has been directly implicated in pre-mRNA storage, alternative splicing, and RNA

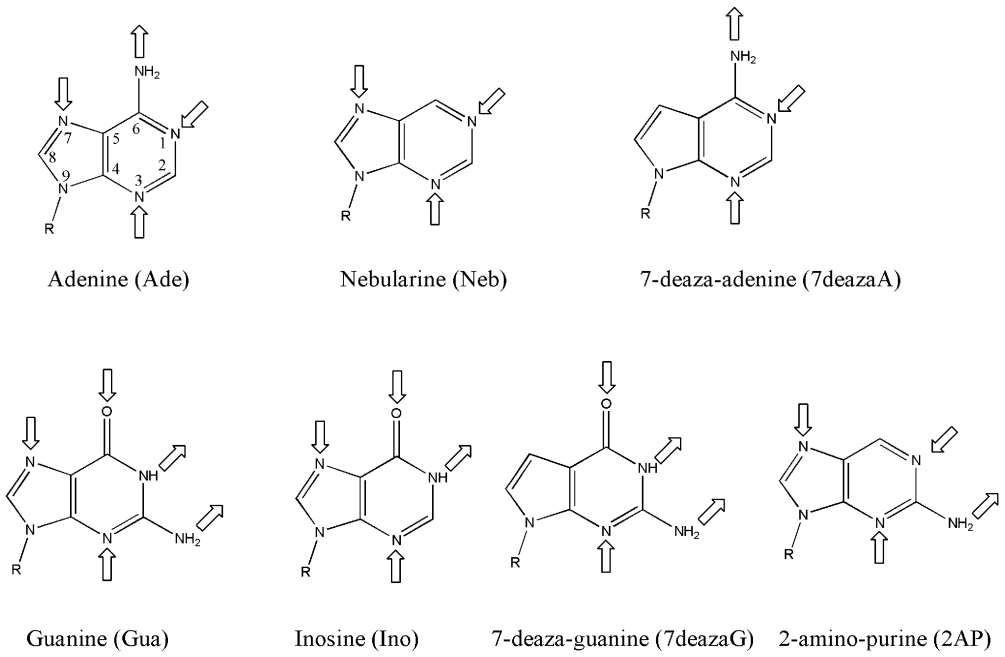

Figure 2. Guide to modified bases used in these studies. Adenine (ade), guanine (gua), 7-deaza-guanine (7deazaG), 7-deaza-adenine (7deazaA), nebularine (neb), inosine (ino), 2-aminopurine (2AP). Arrows indicate positions that are good hydrogen bond donors or acceptors. The nomenclature of each modified oligonucleotide starts with the base to be substituted, followed by its position from the 5′ end of the sequence 5′-d(TTAGGGTTAGGG)-3′, and then by the base substitution.

transport to the cytoplasm.[2,9–14] The N-terminal two-thirds of hnRNP A1 contains two RRMs that have been shown to bind with high affinity to single-stranded RNAs.[15,16] Partial proteolysis studies of hnRNP A1 showed that residues 1–195 (referred to as UP1) retain high affinity for single-stranded nucleic acids and led to the elucidation of an approximately 90 amino acid residue repeat RNA binding motif now commonly called the RRM.[7,8,17,18]

More recently, both UP1 and hnRNP A1 have been shown to bind the human telomeric $d(TTAGGG)_n$ and mouse minisatellite $d(GGCAG)_n$ repeats through their N-terminal RRMs.[19–23] These G-rich DNA sequences form stable G-quadruplexes under a broad range of solution conditions and studies have shown that both hnRNP A1 and UP1 are able to denature both types of G-quadruplexes.[22,23] The structure of UP1 bound to $d(TTAGGG)_2$ has been determined and shows that UP1 binds the human telomeric repeat as an extended single-stranded structure (Figure 1).[24] The ability to bind DNA in a sequence-specific fashion through two RRMs appears to be unique to hnRNP A1 and presents an opportunity to study nucleic acid recognition *via* modified DNA bases.

We have begun to investigate the nucleic acid recognition surface of UP1 using modified purines to assess the mechanisms of ligand recognition (Figure 2). Hydrogen bonding from protein main-chain carbonyl and amide groups to nucleic acid bases are abundant in RNA–protein interactions but cannot be addressed by site-directed mutagenesis of the protein. Informatic analysis of available structures has indicated that main-chain interactions are a prominent feature of RNA recognition and make up about one-third of base-specific hydrogen bonding interactions.[25–27] Main-chain atoms interact with bases by forming a close fit, which may aid in building a nucleotide recognition pocket based both on hydrogen bonding and the shape of the base. The RRM also has a pair of consensus sequences, RNP2 (L-**F**/**Y**-V/I-G-N/D-L) and RNP1 (G-F-G-**F**-V/I-polar-**F**) whose stacking interactions between aromatic amino acid residues and RNA bases are highly conserved. These sequences are also essential to building a conserved nucleotide recognition surface that puts the base edge in position to make specific interactions to the protein.

In addition to examining a role for main-chain recognition in RRMs, we have examined the broader role of specific hydrogen bonding interactions from protein side-chains. Molecular dynamics has suggested that certain highly ionized bidentate interactions, such as lysine or arginine to the N7/O6 position of guanine, are likely to be much stronger than non-ionized interactions, such as asparagine to the N7/O6 of guanine.[28,29] Both Tregor *et al*. and our own studies have shown that ionized hydrogen bonding interactions from lysine and arginine to the guanine N7/O6 are statistically over-represented.[25] The role of these highly specific interactions in positioning and binding nucleic

Table 1. Data collection and structure determination statistics of UP1–oligonucleotide complexes

	G(11)2AP	G(11)Ino	G(10)Neb	G(10)2AP	G(10)Ino	G(10)7-deazaG	A(9)Neb	A(9)7deazA
Resolution range (Å)	20–1.8	20–2.0	20–2.1	20–1.9	20–1.8	20–2.0	20–2.0	20–2.0
R_{merge} (%)	5.1	7.7	8.5	5.6	3.2	8.1	6.2	7.2
Completeness over range (%)	95	96	95	99	97	96	91	89
I/σ average	13.3	12.3	9.6	7.8	46.5	7.9	9.9	9.2
I/σ highest resolution bin (0.1 Å)	3.3	5.1	3.9	1.8	15.3	1.6	1.7	2.3
R_{calc}/R_{free}	24.0/26.5	23.1/27.3	21.3/25.8	23.8/28.4	20.7/25.8	24.1/27.3	23.3/27.4	23.4/26.8
Estimated coordinate error (Å), Luzzatti plot (5 Å cutoff)[a]	0.26	0.28	0.25	0.27	0.21	0.31	0.28	0.32
Overall r.m.s. deviation (C^{α}) to wild-type structure (Å)	0.5	0.5	0.5	0.5	0.6	0.6	0.4	0.7
PDB	1U1R	1U1O	1U1N	1U1P	1U1Q	1U1M	1U1L	1U1K

[a] The estimated upper limit in uncertainty of coordinate error.

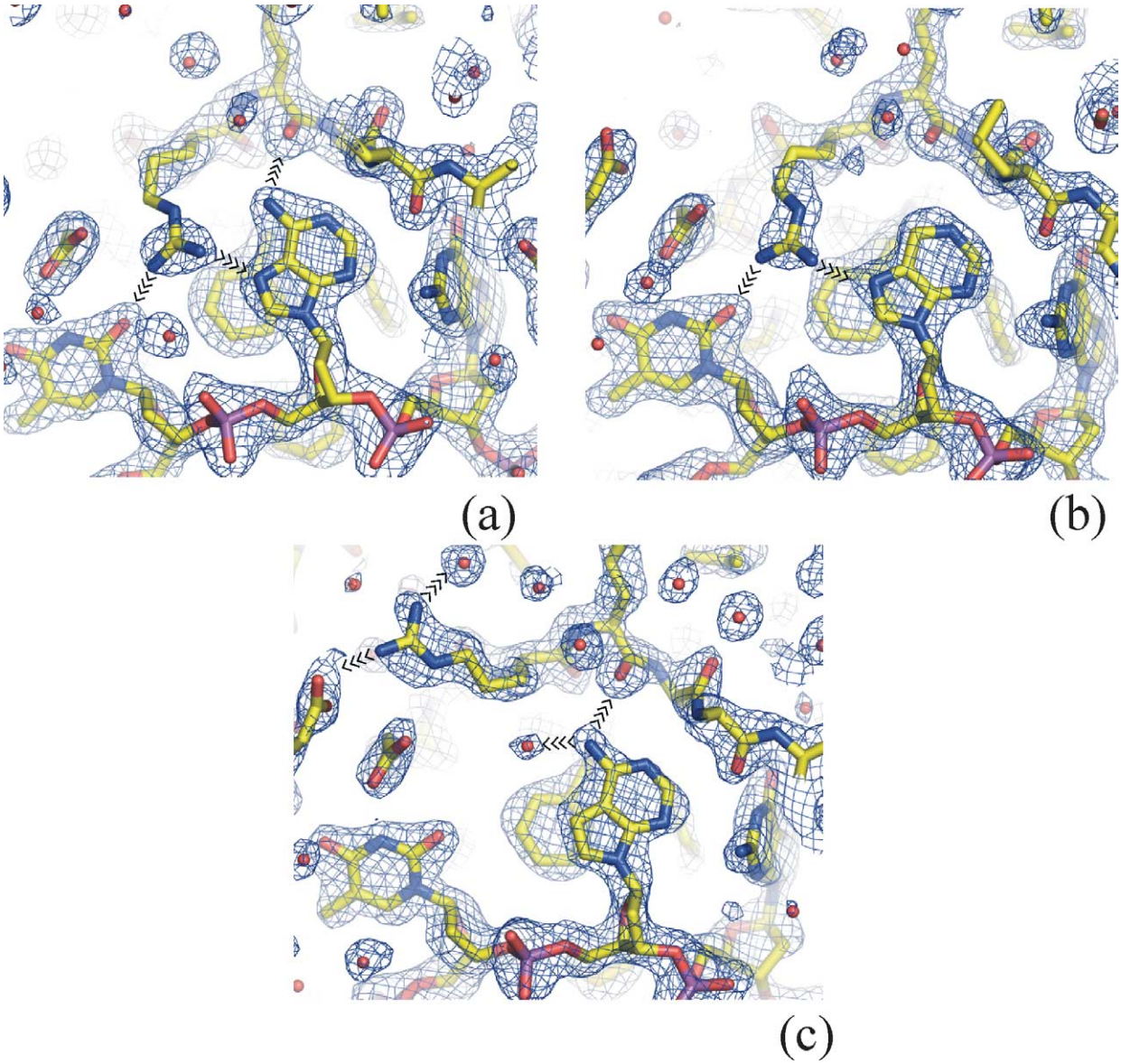

Figure 3. Structures of UP1–oligonucleotide complexes for substitution of adenine 9. $2F_o - F_c$ composite omit electron density maps contoured at 1.25 σ. Chevrons indicate the hydrogen bonding network. (a) Wild-type structure with Ade9 shown making hydrogen bonding contacts to the Arg178 guanidinium group through its N7 (2.7 Å) and from the main-chain carbonyl of Lys179 (3.0 Å). Ade9 is stacked directly over the conserved Phe108 of the RNP2 consensus sequence. (b) The UP1–A(9)Neb structure showed no substantive changes in either protein or DNA structures. The absence of the N6 amino group allowed the base to move slightly closer to Arg178 (2.5 Å). (c) UP1–A(9)7deazaA structure shows a large conformational rearrangement of the Arg178 side-chain. The Arg178 guanidinium group shifted 9.1 Å away from its position in the wild-type structure where it makes contacts to the O2 of Thy8 and N7 of Ade9 to make a new set of contacts to Glu93. All electron density Figures were made using PYMOL (DeLano Scientific, CA).

acids is important to understanding RNA recognition. Experimental investigation into these interactions provides both validation and further insight for future studies into the mechanism of ligand recognition for this important class of proteins.

Here, we have used modified purines to investigate how bases are positioned and how binding affinity is correlated to the proper alignment of sequence-specific contacts. The atomic basis for nucleic acid recognition was examined by a combination of X-ray crystallographic and equilibrium fluorescence binding studies. This combination proved invaluable for evaluating how subtle changes, such as the loss of a single hydrogen bonding interaction, affect nucleic acid recognition. In addition to specific hydrogen bonding patterns, studies have shown that steric exclusion[30] and protein dynamics are also important for RNA recognition.[31,32] Recent work in our laboratory and others' suggests that UP1, along with other well understood RRM proteins such as U1A,[32] can provide excellent experimental systems for

deconvolution of sequence-specific nucleic acid interactions in the RRM family of proteins.

Results

Determination of UP1–oligonucleotide structures

X-ray crystallography was used to determine the structures of UP1 bound to modified telomeric DNA sequences. As shown in Table 1, we were able to obtain moderately high-resolution structures for eight of the nine complexes. All eight structures showed only localized changes to atomic positions in either the protein or DNA (Table 1) and indicated that the modifications did not alter the global structure of the complex. As shown in Figures 3–5, $2F_o - F_c$ composite omit maps were used to identify changes in structure at the region where the modified nucleotides contact the protein surface. Clearly interpretable electron density was obtained for all the modified oligonucleotides except in the case of UP1–G(10)Neb where the modified nucleoside had electron density consistent with high mobility and multiple conformers. The UP1–G(10)2AP structure clearly demonstrated a conformational change from *syn* to *anti* when 2-amino-purine is substituted for guanine.

Binding studies of UP1 complexed to modified oligonucleotides using fluorescence of 6-methyl-8-(2-deoxy-β-ribofuranosyl)-isoxanthopteridine

The affinity of UP1 for modified DNAs was determined by a competition assay in which UP1 bound to a fluorescent oligonucleotide containing 6-methyl-8-(2-deoxy-β-ribofuranosyl)isoxanthopteridine, or 6-MI, d(TTAGG(6-MI)TTAGGG) was challenged by non-fluorescent oligonucleotides of interest.[23] In a previous study, we have shown that incorporation of 6-MI into the telomeric repeat sequence at position 6 puts the fluorescent base into a completely solvent-exposed position with no base to UP1 contacts.[23] Formation of the UP1–d(TTAGG(6-MI)TTAGGG) complex resulted in a ~200% increase in emission at 430 nm when excited at 340 nm. The strength of the emission is advantageous for recording the moderately weaker binding isotherms we expected from our modified oligonucleotides.

Using the native $d(TTAGGG)_2$ sequence as a competitor we have measured an affinity of 88 nM in 300 mM NaCl, 20 mM Hepes (pH 7.4) (Figure 6), which is in good agreement with earlier measurements of UP1 affinity for telomeric DNA sequences.[22,23] The results of the competition binding assays of UP1 for each modified sequence are shown in Figure 6 and Table 2. The range of measured affinities varied from an increase in affinity for G(10)Ino (59 nM) to a nearly tenfold decrease for G(10)A. Few general trends can be ascertained from these binding data without an appropriate structure of the complex upon which to base them. Given that the affinity of UP1 for the sequence $d(TTAGGG)_2$ is approximately 1000-fold stronger than that for a comparable non-specific single-stranded sequence[15] and involves many contacts to the DNA bases, it is not surprising that our intentionally modest changes have not produced dramatic changes in affinity. The range of affinities observed is consistent with the effects expected from the loss of one to two non-ideal hydrogen bonding interactions.

Circular dichroism (CD) of modified oligonucleotides

The telomeric repeat $d(TTAGGG)_n$ forms a stable G-quadruplex structure under a wide range of pH and ionic conditions.[23,33–36] Although UP1 is a potent destabilizer of these G-quadruplex structures, their formation and stability would be a potential challenge when comparing dissociation constants from modified oligonucleotides, because each would have varying degrees of quadruplex stability. Since we wished to compare the relative affinities of our modified oligonucleotides, we used circular dichroism to quantify the ability of each oligonucleotide to form stable G-quadruplexes in the concentration range of our fluorescence studies.[34] If an oligonucleotide formed a G-quadruplex, then melting of the structure would be required for UP1 binding and could result in underestimation of the affinity, compared to an oligonucleotide that did not form a stable quadruplex.

The telomeric G-quadruplex formed by $d(TTAGGG)_n$ has a characteristic spectrum (Figure 4) and is readily distinguished from single-stranded DNA. The human telomeric repeat $d(TTAGGG)_2$ forms a G-quadruplex with characteristic positive peaks at 295 nm and 250 nm, and a small negative peak at 266 nm.[34] Heat denaturation of the G-quadruplex at 45 °C (Figure 7(a)) induces changes in the CD spectra that reduce the positive peak at 295 nm to near baseline and shift the peak at 250 nm to around 260 nm. As shown in Figure 7, all of our oligonucleotides with modifications at Gua10 and 11 abrogated G-quadruplex formation, thus allowing a reliable comparison of their relative affinities. These data, as well as a previous study, have shown that incorporation of 6-MI into position 6 also destabilizes the G-quadruplex structure.[23]

The structure of the human telomeric repeat has been determined and suggests that Ade9 is not directly involved in the formation of the planar G-quadruplex interactions, but rather is looped out into solvent.[36] Our CD studies agree with the structure, in that modifications to Ade9 did not prevent formation of a stable G-quadruplex (Figure 7(b)), whereas modifications of Gua6, 10 and 11 do. Substitution of Ade9 did reduce 295 nm peak intensity to approximately half of that seen for $d(TTAGGG)_2$ and suggests that the modifications at

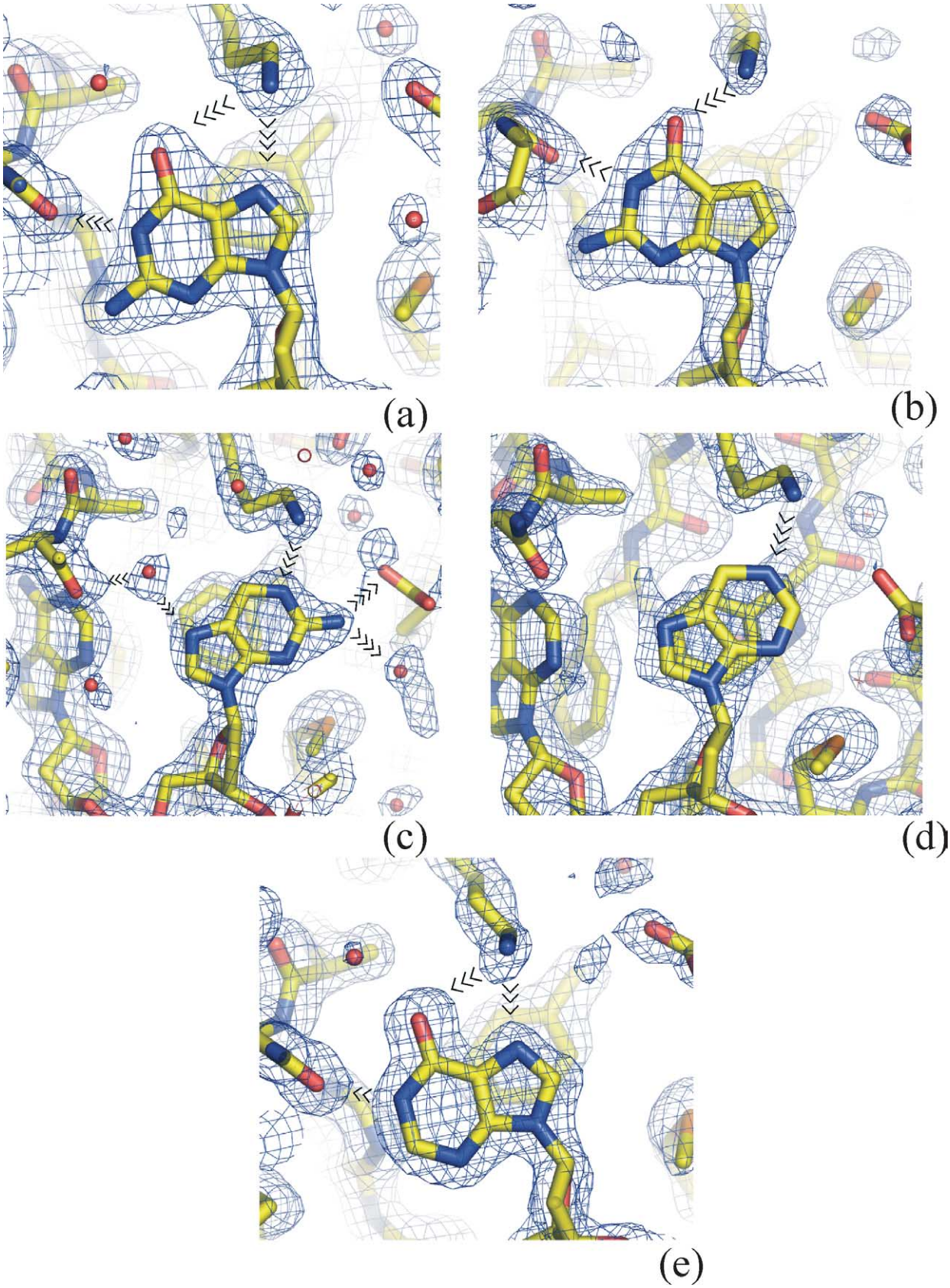

Figure 4. Structures of UP1–oligonucleotide complexes for substitution of guanine 10. $2F_o - F_c$ composite omit electron density maps contoured at 1.25 σ. (a) Wild-type structure with Gua10 shown making a bidentate hydrogen bond from the amino group of Lys106 to N7 (3.0 Å) and O6 (3.5 Å) and from the carbonyl group of Leu181 to N1 (2.6 Å). (b) UP1–G(10)7deazaG shows no overall change in structure, with Lys106 in good position to contact O6 (2.7 Å). (c) The UP1–G(10)2AP structure shows that the base rotates nearly 180° from the *syn* to *anti* conformation. The amino group of Lys106

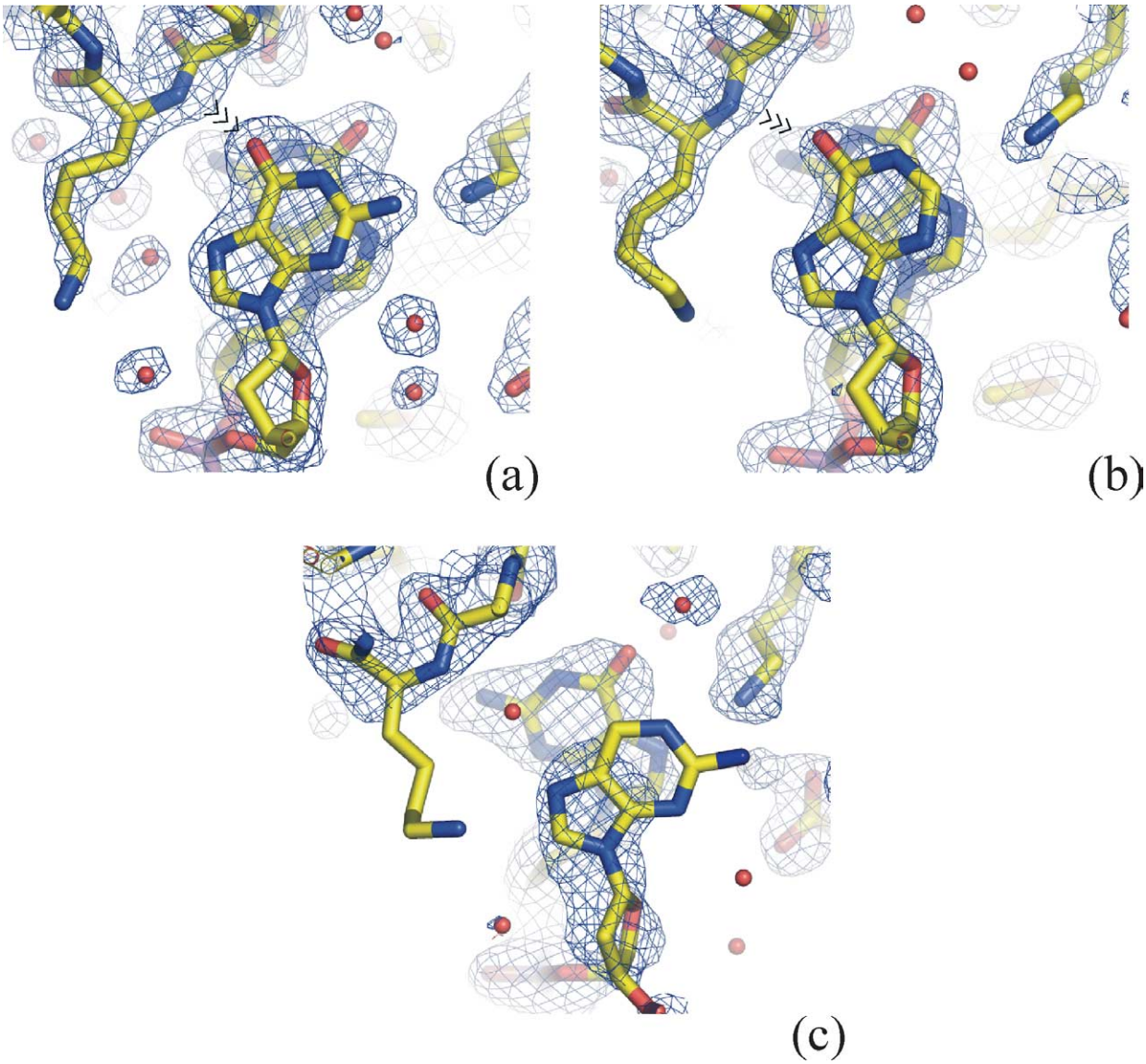

Figure 5. Structures of UP1–oligonucleotide complexes for substitution of guanine 11. $2F_o - F_c$ composite omit electron density maps contoured at 1.25 σ ((a) and (b)) and 1.0 (c). Gua11 stacks between Gua10 and 12 and has only one direct contact to UP1 through the Lys183 main-chain amide to O6 (2.7 Å). (a) Wild-type structure. (b) UP1–G(11): Ino remains stacked and maintains high affinity. (c) UP1–G(11)2AP also remains stacked between adjacent guanine bases despite loss of the one contact from O6 to the main-chain amide, but the mobility of both positions 11 and 12 have increased as demonstrated by weakened electron density and commensurately increased temperature factors. Gua11 is positioned by the favorable stacking interactions with Gua10 and 12 but loss of the main-chain interaction results in significantly increased base mobility and an overall decrease in affinity.

Ade9 have somewhat reduced the stability of the G-quadruplex.

Discussion

The complex and varied contacts responsible for sequence-specific RNA recognition by the RRM family of proteins are a rich source of insights into the atomic basis for RNA recognition by proteins. The secondary structure of the RRM is characterized by a βαββαβ-fold in which the four β-strands make a relatively flat, anti-parallel β-sheet that forms most of the nucleic acid binding surface (Figure 1). Many of the base-specific contacts made between RRMs and nucleic acids are from β-strand 4 and residues at the C terminus. One of the characteristic features of the RRM is a highly conserved set of residues that form the RNP consensus sequence found in β-strands 1 (RNP2:

makes a new contact to the N1 position and all contacts to the main-chain carbonyl of Leu181 are lost. (d) UP1–G(10)Neb shows greatly weakened electron density for nebularine and indicates increased base mobility consistent with the *syn* and *anti* conformers. (e) The UP1–G(10)Ino structure is essentially unchanged compared to wild-type complex.

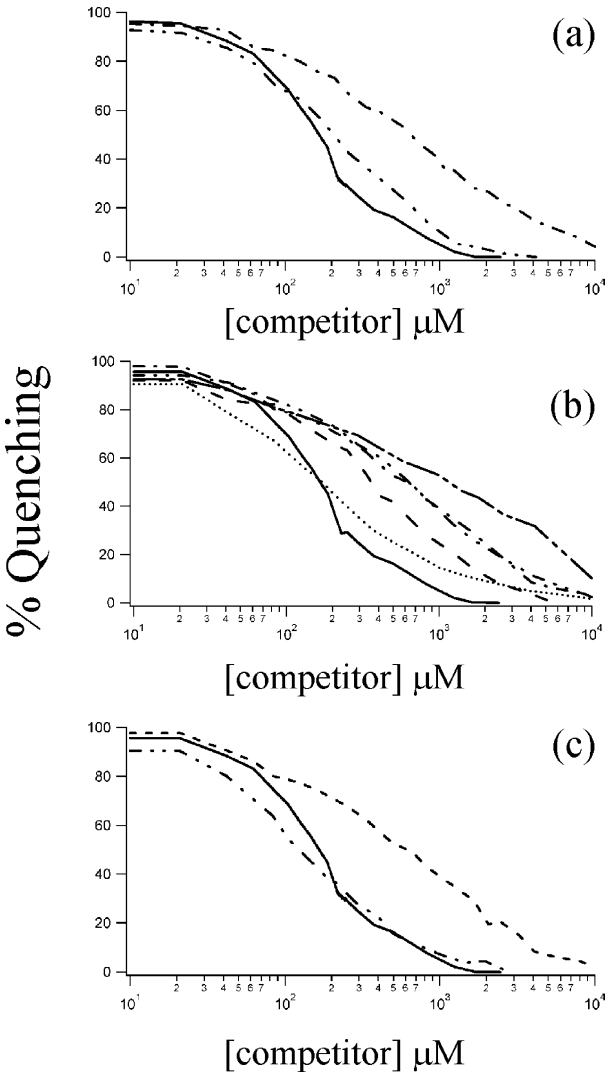

Figure 6. Competition binding of modified oligonucleotides *versus* the fluorescently labeled oligonucleotide d(TTAGG(6-MI)TTAGGG). Competitor oligonucleotides were titrated into a cuvette containing 150 nM UP1 and 150 nM d(TTAGG(6MI)TTAGGG) in 300 mM NaCl, 20 mM Hepes (pH 7.4). The continuous line in each panel shows a binding isotherm for the wild-type $d(TTAGGG)_2$ sequence (—). The data shown are the change in fluorescence as a function of competitor oligonucleotide concentration. (a)–(c) Binding isotherms for the modified oligonucleotides at each position. (a) Ade9 substitutions: A(9)7deazaA (– · –), A(9)Neb (– · · –); (b) Gua10 substitutions: G(10)7deazaG (– · –), G(10)Ino (·), G(10)2AP (– –), G(10)Neb (– · · –), G(10)A (– – –); (c) Gua11 substitutions: G(11)Ino (– · · –), G(11)2AP (– –). Data were fit using IgorPro (Wavemetrics Inc, CA) as described in Materials and Methods and by Myers *et al*.[23]

F/**Y**-V/I-G-N/D-L) and 3 (RNP1: G-**F**-V/I-polar-**F**). Two of the most highly conserved residues within the RNP consensus sequences are solvent-exposed aromatic residues that make stacking interactions with nucleic acid bases (Figure 1(b)). Alignment of putative RRMs in the Conserved Domain Database[37] reveal that the conserved stacking aromatic amino acid residues in RNP2 are either

Table 2. Dissociation constants for UP1 binding to oligonucleotides from competition assays

Oligonucleotide	K_d (nM)
Wild-type	88
A(9)7deazaA	319
A(9)Neb	107
G(10)7deazaG	296
G(10)Ino	76
G(10)2AP	209
G(10)Neb	316
G(10)A	~500–1000[a]
G(11)Ino	59
G(11)2AP	229

[a] Very weak binding of this oligonucleotide precluded an accurate estimate of affinity at 0.3 M NaCl.

phenylalanine (42%) or tyrosine (30%) and in RNP1 is phenylalanine (71%). These stacking interactions are conserved independent of target RNA or DNA sequence and play an important role in positioning the base as part of the ligand recognition surface.[24,38–41] Since these highly conserved stacking interactions are found in all the available RRM structures, they suggest a common and clearly important aspect of RNA recognition.

Adenine 9 recognition

In the native UP1:$d(TTAGGG)_2$ structure, the base of Ade9 is stacked over Phe108 and makes two hydrogen bonds from the N7 to the Arg178 guanidinium group (2.7 Å) and from N6 to the main-chain carbonyl of Lys179 (3.0 Å) (Figure 3(a)). Arg178 also makes a hydrogen bond to the O2 of Thy8.

Substitution of Ade9 with nebularine removes the N6 amino group and changes N1 from a hydrogen bond acceptor to a donor (Figure 3(b)). This substitution has only modest effects on structure and UP1 affinity. The loss of the N6 to Lys179 main-chain carbonyl interaction results in a small reduction in affinity to 109 nM. Any interpretation of the change in affinity must also take into account the observation that loss of the N6 amino group allows the nebularine to make a closer contact with Arg178 (2.5 Å) and presumably helps compensate for loss of the main-chain interaction.

Very different results were obtained upon substitution of Ade9 with 7-deaza-adenine. 7-Deaza-adenine changes the N7 to a carbon and eliminates the ability of residue 9 to serve as a hydrogen bond acceptor (Figure 2). This substitution changes the binding affinity from 88 nM to 319 nM and moves the Arg178 side-chain away from the base into an alternate interaction with Glu93 more than 9.5 Å away. The void left by the movement of Arg178 is filled, in part, by a water (Figure 3(c)). Globally, the position of the 7-deaza-adenine is nearly identical to that of the Ade9 wild-type structure (Figure 3(c)). More of the affinity appears to be associated with the Arg178 to adenine N7, than the N6 to Lys179 carbonyl hydrogen bond. However, it is clear that

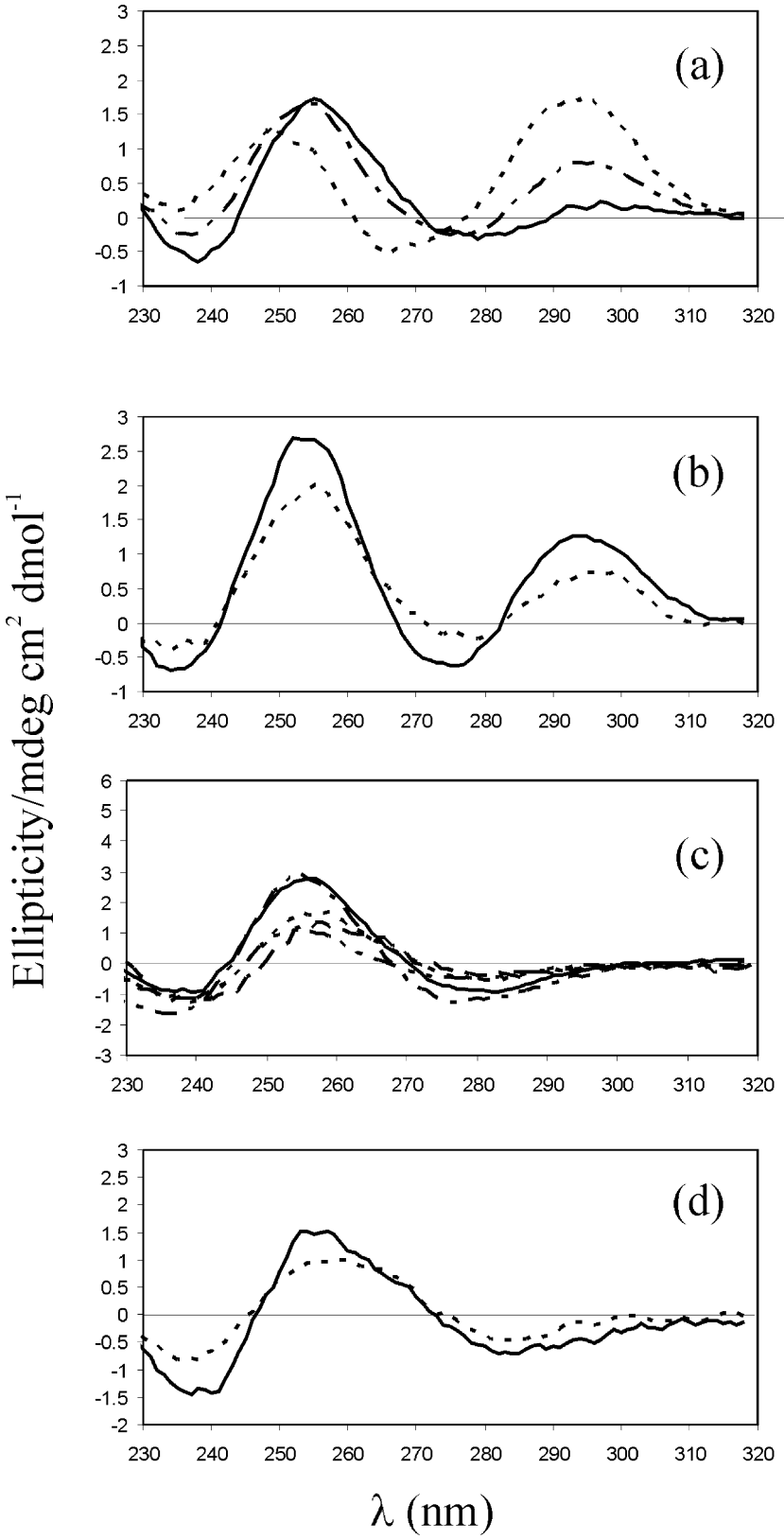

Figure 7. Circular dichroism spectra of oligonucleotides. The human telomeric repeats d(TTAGGG)$_n$ have a strong tendency to form G-quadruplex structures. The G-quadruplexes formed by these repeats have a characteristic spectra with strong positive peaks at 295 nm and 250 nm, as well as a shallow negative region at 266 nm. As shown in (a) the G-quadruplex can be denatured at high temperature to a single-stranded species with a single positive peak at 260 nm and loss of the peak at 295 nm. 25 °C (– –), 35 °C (– - –), 45 °C (—). (b)–(d) Spectra obtained for each modified oligonucleotide by position. (b) Adenine 9 substitutions: A(9)7deaza A(– –), A(9)Neb (—); (c) guanine 10 substitutions: G(10)7deazaG (—), G(10)Ino (– - –), G(10)2AP (– - - –), G(10)Neb (– –), G(10)A (— —); and (d) guanine 11 substitutions: G(11)Ino (– –), G(11)2AP (—).

the combination of main-chain to N6 and stacking of Ade9 on Phe108 are sufficient to restrain the base into its wild-type position without the interaction of Arg178. These results point out the importance of the stacking and main-chain interactions for ligand positioning.

The dramatic repositioning of Arg178 suggests that a steric clash between the hydrogen of the 7-deaza carbon and Arg178 guanidino group, together with the loss of the favorable interaction to the adenine N7, induces Arg178 to find a new and more favorable interaction with Glu93.

Guanine10 recognition

The nucleotide recognition surface of UP1 for Gua10 is comprised of a stacking interaction from Phe150 and two sequence-specific hydrogen bond contacts (Figure 4). The ε-amino group of Lys106 makes bidentate hydrogen bonds to N7 (3.0 Å) and O6 (3.5 Å) of Gua10. Gua10 also makes a hydrogen bond from N1 to the main-chain carbonyl of Leu181 (2.6 Å) (Figure 4). The carbonyl of Leu181 could also make a more distant (3.3 Å) interaction to the N2 but the geometry is not optimal. Although Lys106 is proximal to both the O6 and N7 of Gua10, a lower map contour shows that stronger electron density exists for the O6 to Lys106 hydrogen bond (not shown).

The structure and binding studies of UP1 bound to G(10)7deazaG showed a substantial decrease in affinity (296 nM), but no significant change in position of the bases or UP1 (Figure 4(b)). The ε-amino group of Lys106 is better poised to make contact with the O6 (2.7 Å from 3.5 Å in the wild-type structure), but the loss of the bidentate interaction to the purine N7 has led to a measurable decrease in affinity (Table 2). Despite the lowered affinity, the combination of the monodentate Lys106 interaction to O6, Leu181 main-chain to N1, and stacking interaction with Phe150 are able to maintain the position of G(10)7deazaG.

The UP1 to G(10)2AP and G(10)Neb structures both exhibit dramatic changes in base position (Figure 4(c) and(d)). In the case of G(10)2AP, loss of the guanine O6 and change of N1 to a hydrogen bond acceptor causes 2AP to flip into an *anti* conformation. Despite rotating nearly 180° about the glycosidic linkage, binding is down only 2.4-fold largely because the ε-amino group of Lys106 can now make a good hydrogen bond to the N1 of 2-aminopurine (2.7 Å). All potential contacts from N1 and N2 to the carbonyl of Leu181 are lost as a new set from the N2 to Glu135 are made (2.7 Å). A water-mediated hydrogen bond from N7 to the carbonyl of Leu181 is also formed. Substitution of Gua10 with nebularine addresses the potential contribution of the N1, N2 and O6 for positioning Gua10. The UP1–G(10)Neb structure suggests that the nebularine base has become substantially more mobile than G(10)2AP and may flip between *syn* and *anti* conformers. Figure 4(d) shows that the electron density for nebularine is substantially disordered and indicates increased base mobility. The G(10)Neb has a 3.6-fold decrease in binding, presumably because there is no O6 for Lys106, or N1 for Leu181, to contact as previously shown in the *syn* conformation. Like G(10)2AP, the nebularine base does have a viable N1 hydrogen bond acceptor and shows a similar positioning of the ε-amino group of Lys106. Steric clashes make it unlikely that Lys106 could contact an adenine N1 *in vivo*. Atomic clashes between the adenine N6 amino group and the N1 position in the *anti*, or the N7 position in the *syn* conformation, would seriously limit adenine binding to the RRM at this position. Although we did not produce crystals of the UP1–G(10)A complex, we were able to show that the increase in K_d for G(10)A is higher than that of any of the modified bases. Therefore Lys106 is an excellent candidate to discriminate against adenine in either the *syn* or *anti* conformer.

Substitution of Gua10 with inosine removes the potential Leu181 carbonyl to N2 interaction. Our studies showed a modest increase in affinity (59 nM), but no effect on overall base position, which suggests that loss of the putative N2 contact (3.3 Å) has only a minimal affect (Figure 4(e)). It is likely that the N1 position which is closer (2.6 Å) and in a better geometric position to contact the carbonyl of Leu181, is more important to recognition than N2. The increase in affinity might be due to the inability of G(10)Ino to form stable G-quadruplexes. A shift in the G(10)Ino to a more single-stranded population would undoubtedly facilitate UP1 binding to the ligand and account for the higher affinity.

Guanine 11 recognition

Instead of stacking upon the conserved aromatic residues of the RNP motif, Gua11 stacks between Gua10 and 12 (Figure 1(b)). The only direct contact between Gua11 and UP1 is from the main-chain amide of Lys183 to the O6 atom of the base (2.7 Å) (Figure 5). Removal of the potential main-chain to O6 hydrogen bond was investigated by two substitutions; inosine and 2-aminopurine. G(11):Ino had no discernable changes in overall position, and an increased affinity for UP1. Again, the inosine substitution abrogated the G-quadruplex formation normally found in the $d(TTAGGG)_n$ and may increase the observed affinity slightly (Figure 7). The UP1–G(11)2AP structure shows a clear void where the O6 was present in the wild-type structure. The stacking of Gua11 between Gua10 and 12 partly stabilizes its position in the absence of any direct contacts from UP1. However, the electron density for 2-amino-purine is much weaker, and loss of this single main-chain hydrogen bond allows significant base mobility. This is readily seen in the average temperature factors for the 2-amino-purine (B_{aver} = 61.3 Å^2) compared to the wild-type guanine (B_{aver} = 28.4 Å^2). Electron density for Gua12 in the G(11)2AP structure is also significantly weaker, and

suggests that Gua11 has become more mobile and, as a consequence, partially disrupts stacking to the neighboring base. The binding of G(11)2AP was reduced (229 nM), presumably from the loss of the O6 to main-chain amide interaction, and makes clear that non-ionized main-chain contacts can be important determinants in binding.

General conclusions

Sequence-specific contacts are made *via* complementary hydrogen bond donor–acceptor relationships and by steric exclusions that position van der Waals contacts close to the base. Using base modifications, we have probed the RRM recognition surface to determine the essential attributes of purine discrimination for UP1 in this important class of RNA binding proteins.

Several conclusions can be drawn readily from our studies. As expected, complementary hydrogen bonding donor–acceptor relationships are key components of purine discrimination. The potential importance of main-chain contacts to nucleic acids, especially RNA, has been suggested by informatic analyses of high resolution structures.[25,26] Our experimental results with G(11)2AP clearly support an important role for these contacts that agrees with informatic and molecular modeling studies.[25,28,29] Main-chain atoms are generally not as mobile as side-chains and thus moving the polypeptide main-chain to avoid a clash is likely to be more difficult than a side-chain. The potentially higher energetic cost to moving a region of main-chain may provide additional base discrimination.

We have found that in addition to appropriate complementation by hydrogen bond donor–acceptor groups, a key feature to discrimination is the steric repulsion that often precludes positioning of two hydrogen bond donors proximal to each other. The clash of the 7-deaza hydrogen with those of Arg178 in the UP1–G(10)7deazaG structure is similar to the type of clashes generated by positioning the guanine N1 (a donor) into a position normally occupied by the adenine N1 (an acceptor). The frequency of adenine N1 recognition is second only to that of the N6 position, and is usually made by a main-chain amide (47.5%).[25] In contrast, the guanine N1 is poorly utilized (fourth in frequency), in part, due to being less accessible, but perhaps also because pairing a hydrogen bond acceptor–acceptor pair in proximity, though unfavorable, does not have the power of a steric clash for discrimination. A good example of steric exclusion in another RRM can be seen in the structure of poly(A)-binding protein, where the positioning of close contacts to C2 of three of the adenine bases would preclude binding of guanine (pdb: 1CVJ).

The most dramatic effects on base mobility and affinity for UP1 were seen in substitutions that removed all specific hydrogen bonding interactions to UP1 such as G(10)Neb, G(10)2AP and G(11)2AP. In these cases, base stacking between the conserved aromatics of the RNP and non-polar contacts were unable to position the base. In contrast, substitution of A(9)7deazA and G(10)deazaG resulted in reduced affinity, but the base still maintained its position because at least one specific contact was maintained.

Our results are in good qualitative agreement with *ab initio* quantum chemical predictions that suggest ionized amino acid to base interactions have substantially higher interaction energies than those involving uncharged side-chain and main-chain atoms.[29] *Ab initio* interaction energies calculated between 28 possible bidentate interactions to four unpaired RNA bases have suggested that the lysine to guanine N7/O6 is the strongest interaction and is ~two to threefold higher than comparable bidentate interactions made between uncharged side-chains such as asparagine residues.[29] This is consistent with our previous informatic studies that showed lysine is second in frequency only to arginine in making contacts to guanine.[25] Our results for G(10)7deazaG show that loss of an amino to N7 contact, while maintaining the O6 hydrogen bond, reduced affinity 3.4-fold and corresponded to a loss of about ~1.0 kcal/mol. Conversion of a bidentate interaction at Gua10 to a monodentate interaction to O6 for G(10)7deazaG has therefore induced a loss of affinity on par with the loss of a single non-ionized main-chain interaction in G(11)2AP. This supports computational studies suggesting that ionized bidentate interactions are particularly useful both for positioning bases and for affinity.[29]

One of the most interesting general findings from these studies is a potentially conserved role for Lys106, two residues N-terminal to the stacking aromatic of RNP-2 (K_{106}-I**F**VGGI). In RRM2 of UP1, Lys106 contacts Gua10 and was shown to be a critical determinant in specificity. The equivalent position in RRM1 is Lys15 and contacts the Gua4 O6. It is interesting to note that the analogous positions for Lys15 and 106 in poly(A)-binding protein are occupied by Ser10 and Asn100, which contact the N6 of adenine. It would seem that this position, two amino acid residues before the first highly conserved aromatic residue of RNP2, may signal the preferred purine. Based on sequence alignment, lysine and arginine are present at this position 37% of the time, whereas asparagine is found 21%, and supports the notion that this position provides a key side-chain for purine discrimination in RRMs.[1] Human U1A, U2B″ and *Drosophila melanogaster* SXL all bind pyrimidines at the position equivalent to Gua10 and we could not discern any readily usable consensus for the proteins that bind pyrimidines.

Analysis of purine recognition in UP1 by X-ray crystallography and equilibrium fluorescence has proven to be a useful experimental approach for the testing and validation of the principles governing nucleic acid recognition in the RRM-containing family of proteins. Many of the predictions from informatic and molecular dynamics studies are eminently testable in UP1. Taken together with

comparable biophysical studies using a variety of approaches with other RRM–proteins such as U1A,[32] we should expect significant progress towards engineering RRMs for use as tools for dissecting pathways of splicing and gene regulation in eukaryotes.

Materials and Methods

Purification of UP1 and oligonucleotides

UP1 was expressed and purified as described.[23] Oligonucleotides that did not contain modified bases were obtained commercially from either IDT or MWG Biotech Inc. and those that contained modified bases were obtained from TriLink Biotechnologies. Oligonucleotide purification was performed as described by Myers *et al*.[23]

Crystallization conditions

Crystals of UP1–DNA complexes were grown from initial protein concentrations of 0.8 mM. The UP1–DNA molar ratio was unique for each crystal but was generally higher for modified complexes (1:1.5–1:1.7) than for the native UP1–d(TTAGGG)$_2$ complex (1:1–1:1.2). Oligonucleotides were mixed with UP1 and incubated on ice for 1.5 hours prior to crystallization setup. Crystals were grown using the hanging-drop vapor diffusion method at 10 °C and the initial ratio of complex to precipitant in the drop was typically ≈1:3.0 for the modified oligonucleotides. The precipitant used in the UP1–A(9)7deazaA complex was 28–32% PEG 1550, 50 mM Tris (pH 8.1), 50 mM NaCl, while all other complexes mentioned here crystallized in 1.8–2.2 M dibasic ammonium phosphate, 15% (v/v) glycerol, 100 mM Tris (pH 8.5). All drops required streak seeding 48 hours after setup and diffraction quality crystals took from 3 to 5 days to grow.

Data collection and structure solution

Diffraction data for the UP1–A(9)7deazaA, UP1–A(9)Neb, and UP1–G(10)7deazaG complexes were collected on a Rigaku Jupiter™ CCD detector with Osmic Mirrors™. Data for UP1–G(10)Neb, UP1–G(11)Ino, and UP1–G(10)2AP were collected on a Rigaku R-Axis IV++™. Data for UP1–G(10)Ino and UP1–G(11)2AP were collected at the CHESS A1 beam line on an ADSC Quantum IV Detector. All crystals diffracted to at least 2.2 Å with at least 89% overall completeness. Table 1 shows statistical evaluations of the data and molecular model quality. All complexes crystallized in space group $P4_32_12$ with similar unit cell dimensions of $a=b=52$ Å, $c=172$ Å and two copies of the complex in the asymmetric unit.

Data were indexed and scaled using either d*TREK[42] or DENZO and SCALEPACK[43] and the structures solved in CNS[44] by molecular replacement using the structure described by Ding *et al*. (2UP1) as a search model.[24] Modified bases were built starting from energy minimized structures[45,46] and parameterized for CNS using XPLO2D.[45]

Competition fluorescence titrations

Competition binding experiments were performed using an SLM 8100 spectrometer (SLM Instruments) with polarizing filters (0° excitation and 90° emission). UP1 and d(TTAGG(6MI)TTAGGG), a fluorescent analog of the hTR sequence that uses 8-(2-deoxy-β-ribofuranosyl)isoxanthopteridine in the Gua6 position,[23,47,48] were mixed at equimolar concentrations of 150 nM in 20 mM Hepes (pH 7.4), 300 mM NaCl, and incubated on ice for at least 1 hour prior to the experiment and allowed to warm to 25 °C in a 1 cm path length stirring quartz cuvette. Competitor oligonucleotides were diluted from stock concentrations of 5 mM to 150 μM and brought to a final volume of 400 μl in 20 mM Hepes (pH 7.4), 300 mM NaCl with 150 nM UP1–d(TTAGG(6MI)TTAGGG) mixture so that the total concentration of UP1 and d(TTAGG(6MI)TTAGGG) did not change during the titrations. The competitor oligonucleotides were titrated into the cuvette to compete with the fluorescent oligonucleotide from UP1 and the loss of signal was recorded.[23] Excitation and emission wavelengths were 340 nm and 430 nm, respectively, and all four emission and excitation band-passes were set to 8 nm. Measurements were taken using a 10 seconds time base with five measurements for each titration and 450 seconds between titrations.

The average of the five measurements taken at each titration point was used to fit the Hill equation using the program IgorPro (Wavemetrics, CA). The apparent K_d of the competitor oligonucleotide was determined by identifying the IC_{50} point and from the fitted Hill equation where total binding was the signal of the UP1–d(TTAGG(6MI)TTAGGG) complex without the competitor, and non-specific binding was the saturated signal at the end of the titration. Apparent K_d of the competitor oligonucleotide was determined using K_d of the equilibrium binding constant for the UP1–d(TTAGG(6MI)TTAGGG) complex. This $K_d=130$ nM was determined using a "forward titration" as described.[23,49]

Circular dichroism

Data were collected on an Aviv Model 62A DS Circular Dichroism Spectrometer. Samples for CD were made up at 2 μm in 300 mM NaCl, 20 mM Hepes (pH 7.4) at 25 °C in a 1 cm path length quartz cell. Spectra were taken over the wavelength range 220–320 nm in 1 nm steps using a 10 seconds averaging time per point. Three measurements were made at each step and averaged. The samples were corrected for buffer effects by subtraction of a matched 20 mM Hepes (pH 7.4), 300 mM NaCl solution.

Acknowledgements

This work was supported by a fellowship from The Robert A. Welch Foundation (C-1584) to J.M. and a grant from the American Cancer Society (RSG GMC-104925) to Y.S. The authors thank S. Sun and J. Bruning for data collection and M. Szebenyi for technical expertise at CHESS.

References

1. Birney, E., Kumar, S. & Krainer, A. R. (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucl. Acids Res.* **21**, 5803–5816.
2. Nasim, F. U., Hutchison, S., Cordeau, M. & Chabot, B. (2002). High-affinity hnRNP A1 binding sites and

duplex-forming inverted repeats have similar effects on 5′ splice site selection in support of a common looping out and repression mechanism. *RNA*, **8**, 1078–1089.
3. Faustino, N. A. & Cooper, T. A. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* **17**, 419–437.
4. Lorsch, J. R. (2002). RNA chaperones exist and DEAD box proteins get a life. *Cell*, **109**, 797–800.
5. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A. *et al*. (2003). The InterPro Database, 2003 brings increased coverage and new features. *Nucl. Acids Res.* **31**, 315–318.
6. Hall, K. B. (2002). RNA-protein interactions. *Curr. Opin. Struct. Biol.* **12**, 283–288.
7. Pinol-Roma, S. & Dreyfuss, G. (1993). hnRNP proteins: localization and transport between the nucleus and the cytoplasm. *Trends Cell Biol.* **3**, 151–155.
8. Dreyfuss, G., Kim, V. N. & Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nature Rev. Mol. Cell. Biol.* **3**, 195–205.
9. Mayeda, A. & Krainer, A. R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, **68**, 365–375.
10. Mayeda, A., Munroe, S. H., Xu, R. M. & Krainer, A. R. (1998). Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA*, **4**, 1111–1123.
11. Pinol-Roma, S. & Dreyfuss, G. (1992). Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature*, **355**, 730–732.
12. Michael, W. M., Choi, M. & Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell*, **83**, 415–422.
13. Hou, V. C., Lersch, R., Gee, S. L., Ponthier, J. L., Lo, A. J., Wu, M. *et al*. (2002). Decrease in hnRNP A/B expression during erythropoiesis mediates a pre-mRNA splicing switch. *EMBO J.* **21**, 6195–6204.
14. Blanchette, M. & Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**, 1939–1952.
15. Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F. *et al*. (1991). Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry*, **30**, 2968–2976.
16. Burd, C. G. & Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204.
17. Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H. & Williams, K. R. (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.
18. Shamoo, Y., Abdul-Manan, N., Patten, A. M., Crawford, J. K., Pellegrini, M. C. & Williams, K. R. (1994). Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry*, **33**, 8272–8281.
19. Ishikawa, F., Matunis, M. J., Dreyfuss, G. & Cech, T. R. (1993). Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)$_n$. *Mol. Cell. Biol.* **13**, 4301–4310.
20. LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M. R., Wellinger, R. J. & Chabot, B. (1998). Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nature Genet.* **19**, 199–202.
21. Dallaire, F., Dupuis, S., Fiset, S. & Chabot, B. (2000). Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication *in vitro*. *J. Biol. Chem.* **275**, 14509–14516.
22. Fukuda, H., Katahira, M., Tsuchiya, N., Enokizono, Y., Sugimura, T., Nagao, M. & Nakagama, H. (2002). Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc. Natl Acad. Sci. USA*, **99**, 12685–12690.
23. Myers, J. C., Moore, S. A. & Shamoo, Y. (2003). Structure-based incorporation of 6-methyl-8-(2-deoxy-beta-ribofuranosyl)isoxanthopteridine into the human telomeric repeat DNA as a probe for UP1 binding and destabilization of G-tetrad structures. *J. Biol. Chem.* **278**, 42300–42306.
24. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. & Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115.
25. Allers, J. & Shamoo, Y. (2001). Structure-based analysis of protein-RNA interactions using the program ENTANGLE. *J. Mol. Biol.* **311**, 75–86.
26. Treger, M. & Westhof, E. (2001). Statistical analysis of atomic contacts at RNA-protein interfaces. *J. Mol. Recognit.* **14**, 199–214.
27. Hoffman, M. M., Khrapov, M. A., Cox, J. C., Yao, J., Tong, L. & Ellington, A. D. (2004). AANT: the amino acid-nucleotide interaction database. *Nucl. Acids Res.* **32**, D174–181.
28. Cheng, A. C., Chen, W. W., Fuhrmann, C. N. & Frankel, A. D. (2003). Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol.* **327**, 781–796.
29. Cheng, A. C. & Frankel, A. D. (2004). *Ab initio* interaction energies of hydrogen-bonded amino acid side-chain [bond] nucleic acid base interactions. *J. Am. Chem. Soc.* **126**, 434–435.
30. Nobeli, I., Laskowski, R. A., Valdar, W. S. & Thornton, J. M. (2001). On the molecular discrimination between adenine and guanine by proteins. *Nucl. Acids Res.* **29**, 4294–4309.
31. Castrignano, T., Chillemi, G., Varani, G. & Desideri, A. (2002). Molecular dynamics simulation of the RNA complex of a double-stranded RNA-binding domain reveals dynamic features of the intermolecular interface and its hydration. *Biophys. J.* **83**, 3542–3552.
32. Showalter, S. A. & Hall, K. B. (2004). Altering the RNA-binding mode of the U1A RBD1 protein. *J. Mol. Biol.* **335**, 465–480.
33. Williamson, J. R. (1994). G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 703–730.
34. Balagurumoorthy, P. & Brahmachari, S. K. (1994). Structure and stability of human telomeric sequence. *J. Biol. Chem.* **269**, 21858–21869.
35. Phan, A. T. & Mergny, J. L. (2002). Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucl. Acids Res.* **30**, 4618–4625.

36. Parkinson, G. N., Lee, M. P. & Neidle, S. (2002). Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
37. Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C., Fedorova, N. D., Geer, L. Y., He, S. *et al.* (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucl. Acids Res.* **31**, 383–387.
38. Shiels, J. C., Tuite, J. B., Nolan, S. J. & Baranger, A. M. (2002). Investigation of a conserved stacking interaction in target site recognition by the U1A protein. *Nucl. Acids Res.* **30**, 550–558.
39. Tuite, J. B., Shiels, J. C. & Baranger, A. M. (2002). Substitution of an essential adenine in the U1A-RNA complex with a non-polar isostere. *Nucl. Acids Res.* **30**, 5269–5275.
40. Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallog.* **24**, 958–961.
41. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & 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.
42. Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallog. sect. D, Biol. Crystallog.* **55**, 1718–1725.
43. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
44. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography NMR system: a new software suite for macromolecular structure determination. *Acta Crystalog. sect. D*, **54**, 905–921.
45. Kleywegt, G. J. (1995). Dictionaries for Heteros. *CCP4/ESF-EACBM Newsletter Protein Crystallog.* **31**, 45–50.
46. van Aalten, D. M. F., Bywater, R., Findlay, J. B. C., Hendlich, M., Hooft, R. W. W. & Vriend, G. (1996). PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput. Aided Mol. Des.* **10**, 255–262.
47. Hawkins, M. E., Pfleiderer, W., Balis, F. M., Porter, D. & Knutson, J. R. (1997). Fluorescence properties of pteridine nucleoside analogs as monomers and incorporated into oligonucleotides. *Anal. Biochem.* **244**, 86–95.
48. Hawkins, M. E. (2001). Fluorescent pteridine nucleoside analogs: a window on DNA interactions. *Cell Biochem. Biophys.* **34**, 257–281.
49. Lohman, T. M. & Bujalowski, W. (1991). Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: spectroscopic approaches to monitor binding. *Methods Enzymol.* **208**, 258–290.

Edited by J. Doudna

(Received 23 June 2004; accepted 14 July 2004)