Molecular Recognition of an RNA Trafficking Element by Heterogeneous Nuclear Ribonucleoprotein A2†

Michael J. Landsberg,‡ Kim Moran-Jones,§ and Ross Smith*

School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

Received November 30, 2005; Revised Manuscript Received January 31, 2006

ABSTRACT: Heterogeneous nuclear ribonucleoprotein (hnRNP) A2 is a multitasking protein involved in RNA packaging, alternative splicing of pre-mRNA, telomere maintenance, cytoplasmic RNA trafficking, and translation. It binds short segments of single-stranded nucleic acids, including the A2RE11 RNA element that is necessary and sufficient for cytoplasmic transport of a subset of mRNAs in oligodendrocytes and neurons. We have explored the structures of hnRNP A2, its RNA recognition motifs (RRMs) and Gly-rich module, and the RRM complexes with A2RE11. Circular dichroism spectroscopy showed that the secondary structure of the first 189 residues of hnRNP A2 parallels that of the tandem $\beta\alpha\beta\beta\alpha\beta$ RRMs of its paralogue, hnRNP A1, previously deduced from X-ray diffraction studies. The unusual GRD was shown to have substantial β-sheet and β-turn structure. Sedimentation equilibrium and circular dichroism results were consistent with the tandem RRM region being monomeric and supported earlier evidence for the binding of two A2RE11 oligoribonucleotides to this domain, in contrast to the protein dimer formed by the complex of hnRNP A1 with the telomeric ssDNA repeat. A three-dimensional structure for the N-terminal, two-RRM-containing segment of hnRNP A2 was derived by homology modeling. This structure was used to derive a model for the complex with A2RE11 using the previously described interaction of pairs of stacked nucleotides with aromatic residues on the RRM β-sheet platforms, conserved in other RRM−RNA complexes, together with biochemical data and molecular dynamics-based observations of inter-RRM mobility.

The heterogeneous nuclear ribonucleoproteins (hnRNPs)[1] make up a structurally and functionally diverse family of mosaic proteins formed from combinations of nucleic acid-binding and auxiliary domains. Some of these domains are present in several hnRNPs, but their functions are often not conserved and may be modulated in cis by other domains (*1*−*4*). The hnRNP A/B family includes paralogues A0−A3, the last three of which have multiple isoforms that arise from alternative splicing. HnRNP A2 is a predominantly nuclear protein involved in RNA packaging and splicing, telomere biogenesis, cytoplasmic RNA trafficking, and cap-dependent translation (*5*−*9*). It has a high affinity for several short oligonucleotide sequences, including telomeric DNA repeats (*5*, *10*−*13*), a RNA sequence with homology to a region found in 3′ splice site selection elements (*10*), an AU-rich element from glucose transporter 1 mRNA (*14*−*16*), the vitamin D-regulating response element (*17*), and the A2RE/A2RE11 cytoplasmic RNA trafficking element (*9*, *18*−*20*).

The hnRNP A/B proteins are composed of two ∼90-residue N-terminal RNA recognition motifs (RRMs) followed by an ∼150-residue Gly-rich C-terminal domain (GRD). The sequences of human hnRNP A1 and A2 paralogues are 85% identical in the RRM domains and 48% identical in the GRD. The RRM has an overall $\beta\alpha\beta\beta\alpha\beta$-fold, with the four strands forming an antiparallel β-sheet and the two helices packing at right angles to each other, and adjacent to the sheet (*21*). RRMs are distinguished by the presence of two highly conserved sequences of eight and six amino acids (RNP-1 and RNP-2, respectively) located on the central strands of the β-sheet. The RRM module is also present in one or multiple copies in numerous other proteins that bind mRNA, pre-mRNA, small nuclear RNA, and ribosomal RNA. In the few structurally characterized examples of RNA−RRM complexes, the ligand is bound on the exposed β-sheet surface by a network of interactions (*46*−*49*). The variable loops joining β-strands are thought to determine binding specificity. In vivo proteolysis of hnRNP A1 yields a segment containing only the two RRMs (residues 1−196), also known as unwinding protein 1 (UP1). The three-dimensional structures of these and other RRMs have been the focus of several studies (*21*−*25*).

The GRD of rat hnRNP A2 consists of more than 40% glycine residues and includes several Arg-Gly-Gly (RGG)

† This work was supported by a grant to R.S. from the Australian National Health and Medical Research Council.

* To whom correspondence should be addressed: School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia. Telephone: +61 7 3365 4627. Fax: +61 7 3365 4699. E-mail: ross.s@uq.edu.au.

‡ Present address: Institute for Molecular Bioscience, Queensland Biosciences Precinct, The University of Queensland, Brisbane, Queensland 4072, Australia.

§ Present address: The Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Glasgow, Scotland G61 1BD.

[1] Abbreviations: A2RE, 21-nucleotide hnRNP A2 RNA trafficking response element; A2RE11, 11-nucleotide hnRNP A2 RNA trafficking response element; CD, circular dichroism; GRD, glycine-rich domain; hnRNP, heterogeneous nuclear ribonucleoprotein; HuD, human antigen D protein; RRM, RNA recognition motif; SCR, structurally conserved region; Sxl, *Drosophila* Sex-lethal protein; UP1, unwinding protein 1.

10.1021/bi052440e CCC: $33.50
Published on Web 03/04/2006

repeats, which are thought to constitute an independently recognized, RNA-binding, RGG box domain (*26*). Additionally, GRDs are thought to be capable of participating in protein−protein interactions (*27*). In other classes of protein, including intermediate filament proteins, GRDs are thought to regulate protein oligomerization and anchoring to desmosomes (*28*, *29*).

We present here an analysis of the structure of hnRNP A2 and its complex with A2RE11, an oligoribonucleotide found in many eukaryotic RNAs that appears to act as a signal for microtubule-dependent cytoplasmic mRNA transport (*9*, *19*, *30*). Our circular dichroism (CD) spectra suggest that the three-dimensional structure of the first 189 residues of hnRNP A2 parallels that of the tandem RRMs of hnRNP A1 deduced from X-ray crystallographic data (*1*) and indicate that the GRD has partial β-sheet structure. These observations are consistent with an earlier proposal that the GRD of nucleolin forms an unusual supersecondary structure (*31*). The CD results also lend further support to a model (*2*, *9*) in which two sites for binding oligoribonucleotides are generated by the tandem RRMs acting in concert (*2*, *9*). Our sedimentation equilibrium results are consistent with hnRNP A2 being monomeric when bound to A2RE11, in contrast to the dimeric protein observed in X-ray crystallographic studies of the complex of the UP1 fragment of hnRNP A1 with the telomeric ssDNA repeat (*1*). Finally, by using homology modeling, biochemical data, and molecular dynamics-based observations of inter-RRM mobility and by imposing a set of conserved RNA−protein interactions identified in a survey of known three-dimensional structures of RRM−RNA complexes as structural restraints, we have derived a plausible model for the complex formed between the RRMs of hnRNP A2 and A2RE11.

MATERIALS AND METHODS

Protein Expression and Purification. Plasmids expressing either wild-type hnRNP A2 (pET9c-A2), the 189 N-terminal residues (pET30ΔTh-A2$^{1-189}$), or the 152 C-terminal residues (pET30-A2$^{190-341}$) were transformed into *Escherichia coli* BL21-DE3 cells. pET9c-A2-transfected cells were grown in standard Luria broth (LB) at 37 °C to an A_{600} of 0.6 prior to induction with 1 mM isopropyl thio-β-D-galactoside (IPTG) for 4 h. Recombinant hnRNP A2 was purified to homogeneity from the cleared cell lysate in three steps using DEAE-cellulose (Whatman, Maidstone, U.K.) ion-exchange chromatography, Sephacryl S-300 (Amersham Biosciences, Buckinghamshire, U.K.) size-exclusion chromatography, and C4 reverse-phase (Vydac, Hesperia, CA) HPLC (*9*). A2$^{1-189}$ fusion protein was expressed using a fed-batch fermentation. The recombinant protein was purified to homogeneity using TALON chromatography, on-resin hexa-His tag cleavage, and Sephacryl S-300 size-exclusion chromatography. A2$^{190-341}$ was expressed in standard LB supplemented with 1% glucose and recovered from the insoluble fraction by extraction with 4 M urea, prior to being bound to TALON Co^{2+}-chelating resin (Clontech, Palo Alto, CA). It was refolded on-resin using a 12 h gradient from 4 to 0 M urea. Refolded protein was eluted with 150 mM imidazole and further purified on a Sephacryl S-300 column. It eluted from the column in a single peak.

RNA Binding Analysis. Binding of recombinant proteins to A2RE11 or randomized RNA sequences was assessed by a pull-down assay described elsewhere (*19*), using streptavidin-coated magnetic particles (Roche Diagnostics, Basel, Switzerland) bound to biotinylated RNA (Oligos Etc., Wilsonville, OR). Approximately 10 mg of porcine intestinal heparin (Sigma, St. Louis, MO) was added prior to addition of the protein to the beads.

For UV-cross-linking assays, 1.5 pmol of each oligoribonucleotide, ^{32}P-labeled with T4 polynucleotide kinase (New England Biolabs, Beverly, MA), and 10 pmol of each protein were mixed in binding buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 4% (w/v) glycerol] in a final volume of 15 μL and incubated for 20 min on ice. The reaction mixtures were then irradiated with 500 mJ of 254 nm light in a Bio-Rad GS Genelinker UV chamber. These samples were electrophoresed on SDS−15% polyacrylamide gels and autoradiographed. In competition experiments, 75 pmol of the cold competitor and radiolabeled oligoribonucleotide were added prior to addition of the protein.

CD Spectroscopy. CD experiments were performed at room temperature on a J-710 spectropolarimeter (Jasco, Tokyo, Japan). The protein concentrations were determined by triplicate RP-HPLC-based amino acid analysis using norleucine as an internal control in hydrolysates produced by digestion with 6 M HCl at 110 °C in vacuo for 20 h. Spectra were collected with a 1 nm bandwidth at a scan speed of 20 nm/min for wavelengths above 195 and 2 nm/min for wavelengths below 195 nm. The spectrometer was purged with 5 L of high-purity nitrogen per minute above 195 nm and 20 L of high-purity nitrogen per minute below this wavelength. The buffer solution was either 50 mM Tris (pH 7.5) (protein alone) or 250 mM potassium phosphate (pH 6.6) (for the complex). Spectra recorded with buffer alone were subtracted from the corresponding protein or RNA spectra. Data were collected down to 178 nm where possible to improve the accuracy of secondary structure estimates (*32*). Folding was monitored by following the ellipticity at 225 nm. The level of secondary structure was determined by the method of variable selection incorporated in VARSLC1 (*33*).

Analytical Ultracentrifugation. Sedimentation equilibrium measurements on a solution containing a 2.3:1 molar ratio of synthetic A2RE11 RNA (5′-GCCAAGGAGCC-3′; Dharmacon Research Inc., Lafayette, CO) and purified A2$^{1-189}$ were performed at 20 °C using an Optima XL-I analytical ultracentrifuge (Beckman Instruments, Palo Alto, CA) equipped with interference optics. The equilibrium protein distribution was measured at 14 000 rpm in 250 mM potassium phosphate buffer (pH 6.6). Centrifugation was continued until the protein distribution in the cell was constant (~24 h). Solvent density was measured using an Anton Paar densitometer (Graz, Austria), and the partial specific volume of the protein (0.728 mL/g) was calculated from the amino acid composition (*34*). A value of 0.524 mL/g was used for the partial specific volume of ssRNA (*35*). Molecular masses were calculated either from the slope of a plot of interference fringe displacement versus the square of the radial distance or by Ψ-function analysis (*36*). The protein concentration at the meniscus at sedimentation equilibrium was calculated by a standard iterative technique using the initial concentration (9.03 fringes) and the final fringe pattern, assuming conservation of mass.

Molecular Modeling. Homology modeling of the RRM concatemer of hnRNP A2 ($A2^{9-177}$) was based on the principles described by Greer (*37*) using the HOMOLOGY and DISCOVER modules of INSIGHT II (Accelrys, San Diego, CA). Structurally conserved regions (SCRs) were identified by alignment of the X-ray crystal structures of the N-terminal proteolytic fragment of hnRNP A1 (PDB entry 1UP1) (*38*) and the equivalent domain from *Drosophila* Sex-lethal, Sxl (PDB entry 3SXL) (*39*). The coordinates of the SCRs were transferred from the hnRNP A1 structure as it exhibited a markedly higher degree of sequence conservation than Sxl with hnRNP A2. Energy minimization was performed using an iterative, stepwise procedure on loop regions and SCR side chains, using the DISCOVER consistent valence force field (CVFF), starting with the steepest descent algorithm and transferring to a conjugate gradient algorithm, introducing Morse potentials and cross-terms in later steps. The quality of the model was assessed using PROFILES 3D (*40*) and PROCHECK (*41*). DISCOVER molecular dynamics analysis of the $A2^{9-177}$ model was performed at 300 K for 110 ps using a 1 fs time step to observe any hinge bending motions in the model.

Previous experiments have shown that only one nucleotide may be deleted from the ends of the 11-nucleotide A2RE11 without markedly reducing the level of binding to hnRNP A2 (*13*). The structure of the A2RE11 RNA sequence motif was initially modeled in three fragments. The 5′ three and 3′ four nucleotides were modeled on the respective regions of the UP1-bound telomeric DNA sequence (PDB entry 2UP1) (*1*) that interacted with the β-sheet surface of each RRM, using the atomic coordinates of the phosphate backbone as a framework. The central four nucleotides were built using the default extended-strand properties in the BIOPOLYMER module of INSIGHT II. The three fragments were then ligated, and the whole molecule was subjected to energy minimization. Using GOLD (*42*), we prescreened the A2RE11−$A2^{9-177}$ complex for favorable interactions between pairs of adjacent RNA nucleotide bases and the respective, aromatic RNA binding pockets located on the central β-strands of each RRM. Having identified the optimal RNA−protein contacts as ring stacking interactions between C2/C3 and Phe103/Phe145, and between A8/G9 and Phe12/Phe54, we then docked the complete A2RE11 onto the $A2^{9-177}$ RNA binding surface using DISCOVER molecular dynamics.

RESULTS

RNA Binding of hnRNP A2 Domains. Recombinant proteins were expressed in *E. coli* and purified (Figure S1 of the Supporting Information). Recombinant $A2^{1-189}$ and full-length hnRNP A2 were functional as displayed by their ability to specifically bind magnetic particles bearing biotinylated A2RE11 RNA (data not shown) and to bind sequence-specifically in competition UV-cross-linking assays (Figure 1). In these assays, a 50-fold excess of unlabeled A2RE11 eliminated binding of labeled A2RE and NS1. By contrast, a 50-fold excess of the unlabeled NS1 oligonucleotide eliminated NS1 binding but only partly displaced the binding of A2RE11. These results are in accord with earlier studies that have shown that hnRNP A2 and $A2^{1-189}$ have two sites for binding oligoribonucleotides, one of which is sequence-specific (*2*).

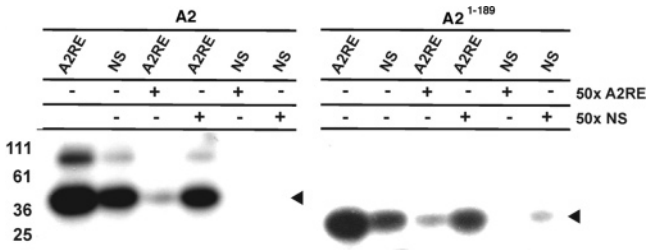

FIGURE 1: Association of hnRNP A2 and $A2^{1-189}$ with A2RE11. UV-cross-linking assays with ^{32}P-labeled A2RE11 and NS1, in the presence or absence of a 50-fold excess of unlabeled oligoribonucleotide, and purified recombinant hnRNP A2 or $A2^{1-189}$. The arrowheads indicate the positions of A2 or $A2^{1-189}$, with the bound probe. The molecular masses, in kilodaltons, of standard proteins are indicated at left. NS1 is an oligoribonucleotide with the same composition as A2RE but with a scrambled sequence (CAA GCA CCG AAC CCG CAA CUG).

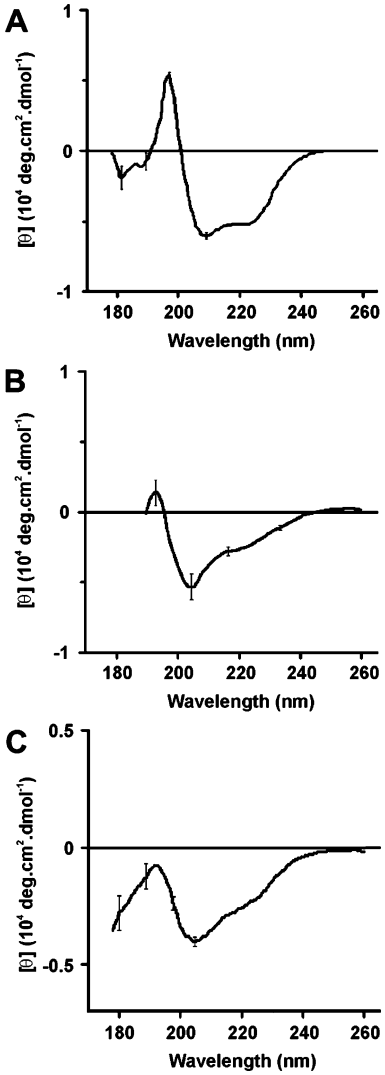

FIGURE 2: CD analysis of hnRNP A2 structural domains. CD spectra recorded at 22 °C for purified recombinant proteins $A2^{1-189}$ (A), $A2^{190-341}$ (B), and wild-type hnRNP A2 (C). The proteins were dissolved in 50 mM Tris at pH 7.5 (A), 8.5 (B), or 7 (C) at protein concentrations of 1 (A and C) and 0.01 mg/mL (B).

Secondary Structure Analysis. The structures of the hnRNP A2 modules were analyzed by CD spectroscopy. Spectra were averaged over five scans using 0.106 mm (Figure 2A,C) or 1 cm (Figure 2B) path length cells. Calculation of the level of secondary structure by spectral deconvolution using

Table 1: Secondary Structures of hnRNP A2 Structural Domains from CD Spectral Deconvolution

	secondary structure	% composition[a]	no. of residues[a]	model[b]
$A2^{1-189}$	α-helix	23 ± 2	43 ± 4	38
	β-sheet (antiparallel)	21 ± 3	37 ± 6	49
	β-sheet (parallel)	3 ± 1	6 ± 2	0
	β-turn	19 ± 2	36 ± 4	36
	random coil	35 ± 1	66 ± 2	66
A2 (wild type)	α-helix	9 ± 1	31 ± 3	
	β-sheet (antiparallel)	29 ± 2	99 ± 7	
	β-sheet (parallel)	6 ± 1	20 ± 3	
	β-turn	21 ± 1	72 ± 3	
	random coil	33 ± 1	113 ± 3	

[a] The standard deviation of the secondary structure estimate is given. [b] The number of residues of secondary structure present in the $A2^{9-177}$ homology model (Figure 5A).

the variable selection algorithm yielded values for the α-helical, antiparallel β-sheet, parallel β-sheet, β-turn, and random coil structures in $A2^{1-189}$ (Table 1) in accord with the content of these secondary structures observed in the crystal structure of the tandem RRMs of hnRNP A1 (*22*)

The CD spectra of the Gly-rich C-terminus of hnRNP A2 could be reproduced in several different refolding experiments. However, the secondary structure of this domain was more difficult to determine as this region is not homologous to any established structural motif. Direct CD measurements of the isolated $A2^{190-341}$ polypeptide at wavelengths below 190 nm, which are required to accurately determine β-structure, were limited by the high concentration of buffer salts required to maintain its solubility. However, analysis of spectra obtained down to 190 nm (Figure 2B) suggested the presence of β-structure, based on previous observations which correlate a low absolute value of negative ellipticity at 200 nm with β-turn content (*43*, *44*). Qualitatively, the spectrum also resembles that of a Gly-rich sequence from nucleolin, previously shown to contain regions of β-structure (*31*).

The CD spectrum was recorded down to 178 nm for wild-type hnRNP A2 (Figure 2C) and deconvoluted by variable selection. The α-helical secondary structure content deduced from the CD data for the full-length protein was 31 ± 3 residues compared with 43 ± 4 residues for $A2^{1-189}$ (Table 1). This difference suggests that there are few, if any, helical residues in the GRD. By contrast, even allowing for some uncertainty in the calculation of the numbers of residues in β-sheets and β-turns, we find there is strong evidence for a substantial content of such residues in the GRD. The sum of the number of β-sheet and β-turn residues in the $A2^{1-189}$ polypeptide was calculated to be 79, compared with 191 such residues in the wild-type protein. The "coiled" residues are approximately evenly distributed between the two segments of the protein.

CD of the $A2^{1-189}$−A2RE11 Complexes. CD spectra were recorded for the A2RE11 RNA and for complexes of A2RE11 bound by $A2^{1-189}$ with RNA−protein stoichiometries of 1:1, 2:1, and 3:1 (Figure 3A). The spectrum of $A2^{1-189}$ showed no optical activity above 240 nm (data not shown), suggesting that changes in the spectra of the complexes above this wavelength primarily reflect changes in the RNA component. In the A2RE11 spectrum, a positive CD peak was observed over the range from 250 to 295 nm,

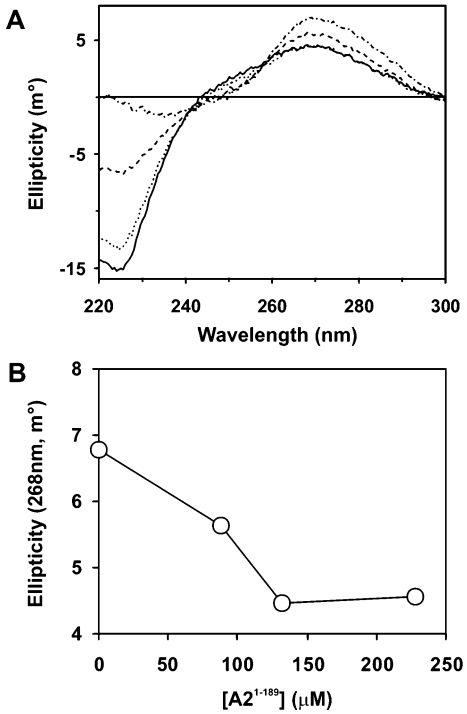

FIGURE 3: (A) Formation of an A2RE11−$A2^{1-189}$ complex monitored by CD. CD spectra were recorded for free A2RE11 (-•-) and A2RE11−$A2^{1-189}$ complexes. The complexes were dissolved in a solution of 250 mM potassium phosphate (pH 6.6) at a RNA concentration of 265 μM. Protein was added to achieve a RNA:protein molar ratio of 3:1 (- - -), 2:1 (•••), or 1.16:1 (—). Data were collected as described for Figure 2A. (B) The change in CD at 268 nm, influenced by inter- and intramolecular base stacking, plateaus at a protein concentration equivalent to an A2RE11:$A2^{1-189}$ molar ratio of 2:1.

with a maximum at 268 nm. Upon addition of $A2^{1-189}$, this peak became more symmetrical, extending through to 245 nm, and the magnitude of the maximum decreased (Figure 3A). Such changes have been observed in other RRM−RNA complexes and can be attributed to the formation of intermolecular base stacking interactions (reviewed in ref *44*), such as those formed between aromatic amino acid side chains on the protein and RNA nucleotide bases.

The changes in the CD spectra near 270 nm are also consistent with the proposed stoichiometry of the $A2^{1-189}$−A2RE11 complex. Titration of the free RNA solution with $A2^{1-189}$ showed that no further change in the magnitude of the CD at 270 nm occurs beyond a RNA:protein ratio of 2:1 (Figure 3B). We have previously shown by other methods that hnRNP A2 binds A2RE11 with a K_d near 50 nM for the sequence-specific site and ∼250 nM for the nonspecific site (*2*). Therefore, at the RNA and protein concentrations used in these CD experiments, the data support our previously proposed two-site binding model based on resonant mirror biosensor and UV-cross-linking electrophoretic mobility shift assays (*2*).

Sedimentation Equilibrium of the A2RE11−$A2^{1-189}$ Complex. The stoichiometry of the A2RE11−$A2^{1-189}$ complex was further investigated using equilibrium analytical ultracentrifugation (Figure 4). An iterative calculation was used to calculate the equilibrium meniscus concentration and an M_r for the A2RE11−$A2^{1-189}$ complex of 30.1 ± 1.2 kDa. This M_r was in agreement with the presence of a heterotrimer consisting of one molecule of $A2^{1-189}$ and two molecules of

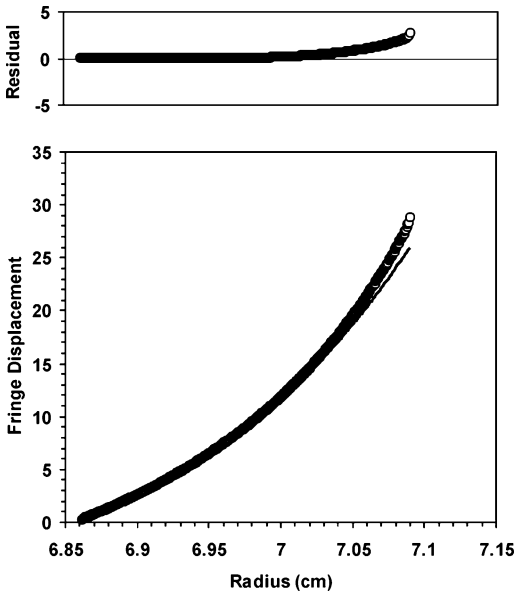

FIGURE 4: Sedimentation equilibrium analysis of the A2RE11−$A2^{1-189}$ complex. The complex was formed by addition of 2.3 mol of A2RE11/mol of $A2^{1-189}$. The solid line in the bottom panel indicates the best fit to a single-molecular mass species in solution. The top panel indicates the residuals for the best fit curve.

the A2RE11 (theoretical mass of 29.3 kDa). We also applied Ψ-function analysis (*36*) to these data and determined that the predominant species (90%) in a 2.3:1 (molar ratio) solution of A2RE11 and $A2^{1-189}$ had an M_r of 29.0 ± 1.2 kDa.

We found no evidence of the formation of dimers or higher-order aggregates for either the A2RE11−$A2^{1-189}$ complex or the $A2^{1-189}$ protein in the absence of RNA, which was determined by sedimentation analysis to have an M_r of 21 ± 0.8 kDa (theoretical mass of 20.8 kDa; data not shown). Additionally, these experiments, along with earlier data (*2*), indicate that little of the protein or A2RE11 is misfolded or modified to an extent that interferes with complex formation.

Homology Modeling and Molecular Dynamics Simulations. In the absence of any experimental three-dimensional structural information for hnRNP A2, we used a modeled structure of the $A2^{1-189}$ fragment (Figure 5A) to further investigate structural determinants expected to dictate the mode of recognition for A2RE11. The model was built using homology modeling and a three-dimensional template structure of the human UP1 fragment of hnRNP A1 (*1*, *22*, *38*, *45*), the sequence of which is approximately 80% identical with that of the corresponding segment of hnRNP A2. The three-dimensional structure of *Drosophila* Sex-lethal protein was also used to identify conserved structural elements. As a result of the high level of sequence identity between UP1 and $A2^{1-189}$, many of the structural elements were conserved in the A2 model, including turns and hairpins located in the interconnecting loop regions of the RRMs. Residues 1−8 and 178−189 could not be modeled because of a lack of sequence identity with known three-dimensional structural fragments.

The secondary structure observed in the three-dimensional model of the hnRNP A2 RRM concatemer agreed well with the structure predicted for this fragment by deconvolution of the CD spectra (Table 1). All secondary structure contents calculated from the CD spectra were within two standard deviations of the structure content observed in the model.

Determinants of A2RE11 Recognition. Ding et al. (*1*) previously identified two pairs of amino acid residues in the UP1-telomeric DNA repeat crystal structure that form four inter-RRM salt bridges. These residues are conserved in hnRNP A2 (Arg70, Arg83, Asp150, and Asp152; see Figure 5B) and in our model form similar contacts. We performed

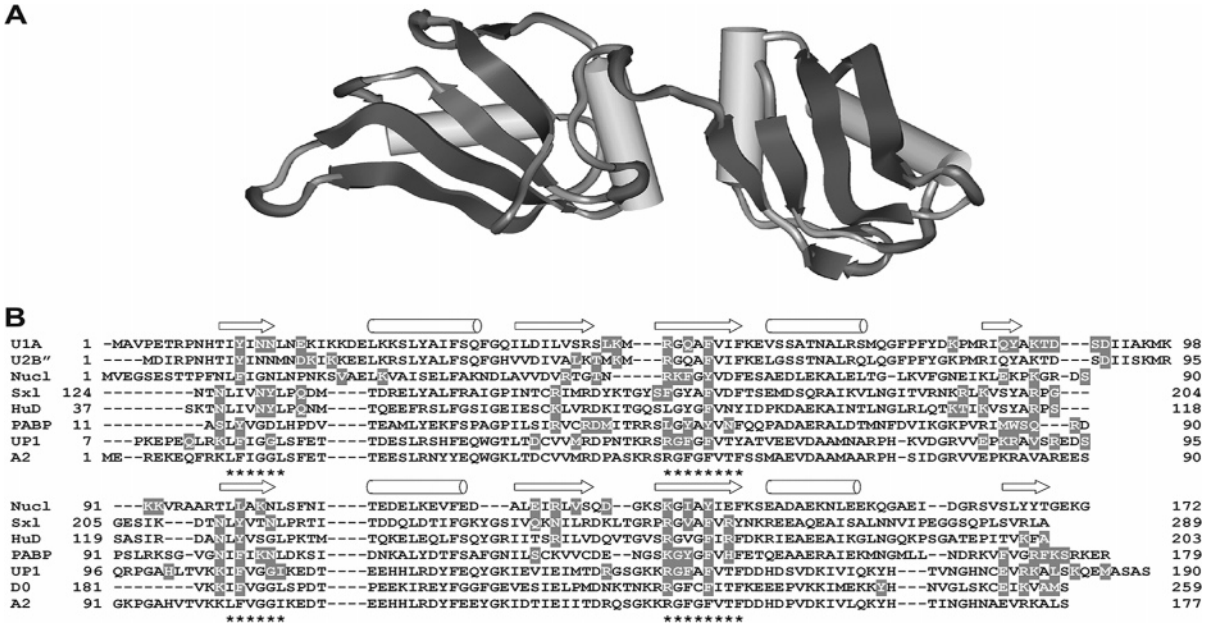

FIGURE 5: Three-dimensional structure of the A2RE11-binding domain of hnRNP A2 and conserved RNA-binding residues. (A) The homology model of the N-terminal, A2RE11-binding region of hnRNP A2 contains two copies of the RRM. In this orientation, the RNA-binding β-sheet platform faces out of the page. (B) Sequence alignment of RRMs that have been cocrystallized with short oligonucleotide ligands. Asterisks denote the RNP-1 and RNP-2 consensus sequences. Gray highlighting denotes residues that interact with the cognate nucleic acid in the structure of each complex. Cartoons above the amino acid sequences represent secondary structure. The PDB entries are as follows: 1URN for U1A, 1A9N for U2B″, 1FJE for nucleolin, 1B7F for Sex-lethal, 1G2E for HuD, 1CVJ for poly-A binding protein, and 2UP1 for unwinding protein 1.

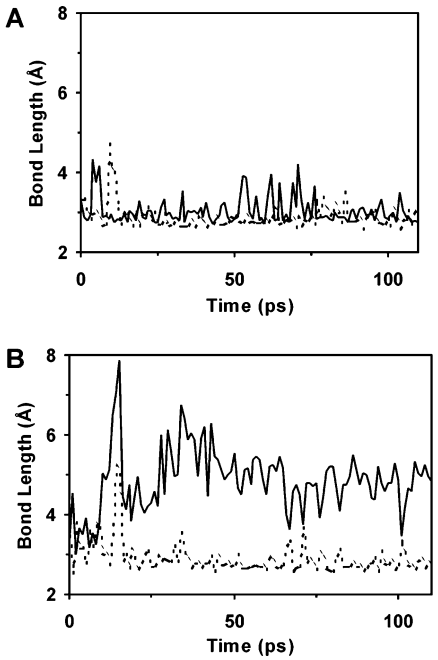

FIGURE 6: In silico analysis of inter-RRM mobility. (A) Asp150−Arg70 electrostatic interactions: Oδ-1−Nϵ (—) and Oδ-2−NH-2 (- - -) bond lengths. (B) Asp152−Arg83 electrostatic interactions: Oδ-2−NH-1 (—) and Oδ-1−NH-2 (- - -) bond lengths.

a molecular dynamics simulation to determine the relative stability of these interactions, as they were predicted to significantly restrain inter-RRM movement. Three of the four bonds that were monitored (Figure 6) were found to be relatively stable and maintained an appropriate bond length for electrostatic interaction (typically 2.5−2.8 Å), suggesting that there is restricted inter-RRM mobility in the structure and that substantial rearrangement of the two RRMs upon RNA binding is unlikely. By contrast, some other cocrystal structures of tandem RRM−RNA complexes such as those involving HuD and Sex-lethal, in which the two corresponding salt bridge Arg residues on the N-terminal RRM are not conserved, show significant relative reorientation of their RRMs upon ligand binding (*46*, *47*).

A further search for conserved nucleic acid-binding residues revealed that only two interactions are conserved across almost all RRM−RNA complexes (Figure 5B). A pair of hydrophobic (usually aromatic) amino acid residues juxtaposed on strands 1 and 3 of the β-sheet surface form base stacking interactions with the nucleic acid. In tandem RRM−RNA complexes, these interactions have been observed for both RRMs, with the two hydrophobic side chains on both RRMs forming dinucleotide-binding pockets. In all but one case (*48*), these hydrophobic residues bind two dinucleotides separated by a sequence, conserved in length, of four nucleotides (Table S1 of the Supporting Information). The model is consistent with the conclusion drawn from the CD spectra, recorded over 250−300 nm, that association of $A2^{1-189}$ with A2RE11 is accompanied by an increase in the level of intermolecular base stacking,

Model for the Specific Recognition of A2RE11 by A2. We have proposed a model structure for the complex formed between $A2^{1-189}$ and a single A2RE11 oligomer bound in the sequence-specific binding site (Figure 7). Initial in silico experiments using Genetic Optimisation of Ligand Docking (*42*) indicated that binding of the RNA in the 3′ to 5′ direction (with the protein oriented from the N- to C-terminus) was preferred. Furthermore, the intermolecular base stacking interactions described above were most energetically favorable when RNP-1 and RNP-2 aromatic side chains stacked with the A8G9 and C2C3 dinucleotides for RRMs-1 and -2, respectively (Figure 7). Interaction of an AG dinucleotide with RRM-1 was also observed in the UP1−telomeric DNA complex, further justifying the positioning of the A8G9 dinucleotide in the Phe12/Phe54 binding pocket. The C2C3 and A8G9 dinucleotides in this complex are also separated by a four-nucleotide linker sequence, consistent with the observations presented in Table S1, and in accord with the positioning of the C2C3 dinucleotide in the Phe103/Phe145 binding pocket. Thus, sequence-specific binding of A2RE11 by hnRNP A2 appears to involve base stacking between A8G9 and Phe12/Phe54 on RRM-1 and between C2C3 and Phe103/Phe145 on RRM-2.

In the model, the RNA molecule lies across the β-sheet platform and forms extensive contacts with the inter-RRM linker. This observation is consistent with several previous examples in which inter-RRM linkers adopt a defined conformation as a result of such ligand−linker interactions (*46*−*49*). The binding platform is relatively electropositive (Figure 7A) as expected for a surface binding negatively charged RNA. Significantly, the structure proposed here differs from those previously described for complexes of Sex-lethal and HuD with their cognate RNAs. In these structures, strands 4 and 2 of RRM-1 and -2, respectively, lie closest to one another. By contrast, in the structures of hnRNPs A1 and A2, the second RRM is rotated 180° with respect to the first such that the same strands (strand 4) of the two RRMs are juxtaposed. This arrangement prevents the inter-RRM linker from looping away to one side of the β-sheet binding platform, and instead, the RNA is positioned beside part of the inter-RRM linker. This difference in binding mode, in addition to the presence of inter-RRM salt bridges, is consistent with the reduced degree of conformational change anticipated for hnRNP A1 (from the cocrystal structure) and hnRNP A2 (from molecular dynamics), compared to HuD and Sex-lethal (*46*, *47*), on binding oligonucleotides.

DISCUSSION

The sequence-specific recognition of the cis-acting A2RE11 by hnRNP A2 is necessary and sufficient to direct the localization of myelin basic protein mRNA in oligodendrocytes (*9*). A2RE-like elements have also been found in a number of other eukaryotic mRNAs (*30*), and the localization of A2RE-containing transcripts has also been demonstrated in neurons (*20*), suggesting that hnRNP A2 may form part of a generic cytoplasmic RNA trafficking pathway.

Sequence analysis identifies two RRMs in the N-terminus of hnRNP A2, in accord with the CD and homology modeling data we have presented. We have previously shown that this region of hnRNP A2 (with the inclusion of 10 amino acids immediately C-terminal to the second RRM, $A2^{1-189}$) retains full binding capacity for A2RE11 (*2*). This protein segment possesses two A2RE11-binding sites: one site is sequence-specific and the second accommodates any RNA sequence (*2*, *9*). The former site has a K_d near 50 nM for A2RE11 binding, while the latter has a K_d near 200 nM.

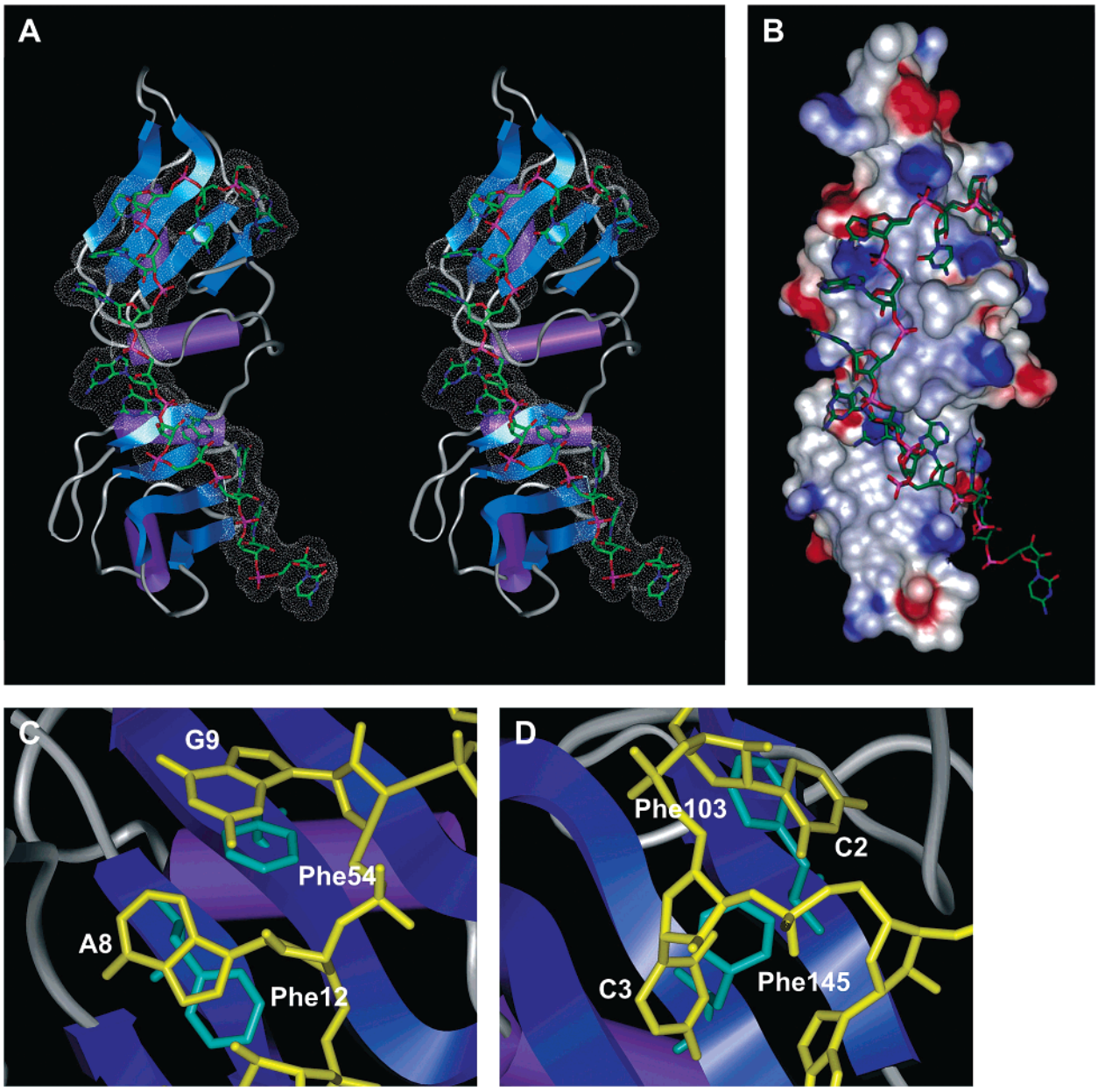

FIGURE 7: Computational three-dimensional model of an A2RE11−A2$^{1-189}$ complex. (A) Stereogram of the complex with the protein heavy atom backbone and the Connolly surface of the RNA (dotted surface). The secondary structure rendering of the protein highlights the location of the extended β-sheet platform, as well as the central β-strands where the conserved dinucleotide binding pockets reside. (B) The same complex with the protein rendered as a Connolly surface, colored according to electrostatic potential, where red indicates electronegativity and blue electropositivity. The orientation and scale of the complex in panels A and B are approximately the same. (C and D) The phenylalanine residues of hnRNP A2 at positions 12 and 54 of RRM-1 and positions 103 and 145 of RRM-2 are highly conserved and are the only residues that bind RNA in all structurally characterized complexes. In this model, the RRM-1 pocket (C) accommodates nucleotides A8 and G9 of A2RE11 and the RRM-2 pocket (D) accommodates C2 and C3.

We have now examined the stoichiometry of the complex formed between A2$^{1-189}$ and A2RE11 using CD and sedimentation analysis and confirmed that two oligonucleotide molecules are bound per protein molecule, forming a heterotrimer. There was no evidence of formation of larger protein aggregates, in contrast to the cocrystal structure of UP1 with telomeric repeat DNA in which the protein formed dimers that bound two oligonucleotides (*1*). Thus, these complexes differ despite the high degree of sequence identity (~80%) and structural similarity (see above) of the RRM domains of the hnRNP A1 and A2 paralogues, and the expectation that the binding determinants in the UP1 structure would not be influenced by substituting the DNA for RNA (*1*). This difference between hnRNP A1 and A2 parallels other properties: they have some in common and others that differ. Examples of the latter are their differing affinities for A2RE/A2RE11 (*2*) and effects on cell proliferation (*50*).

We have identified several likely influences on the sequence-specific recognition of A2RE11 RNA by A2$^{1-189}$. Comparative analysis of all RRM−RNA complexes of known three-dimensional structure revealed two interactions in each RRM that are common to all structures. These involve a pair of conserved hydrophobic (usually aromatic) amino acid residues residing on β-strands 1 and 3 stacking with nucleotide base rings on the RNA. These two amino acid side chains frequently combine with variable surrounding side chains to form a dinucleotide-binding pocket that forms the basis for recognition of the RNA. Such an

interaction is feasible for the hnRNP A2−A2RE11 complex. Our CD data (Figure 3) also suggest that base stacking contacts are strongly involved in the intermolecular recognition of A2RE11 by $A2^{1-189}$.

In proteins where tandemly arranged RRMs recognize single-stranded RNAs, the pairs of nucleotides bound in the pockets of each RRM are in almost all cases separated by a spacer of four nucleotides. We imposed base stacking interactions with the conserved phenylalanines (residues 12, 54, 103, and 145) as structural restraints to generate a model for the sequence-specific recognition of A2RE11 by $A2^{1-189}$. We identified a pair of relatively stable inter-RRM salt bridges, as assessed by molecular dynamics analysis, which suggests that gross changes in the relative positions of the RRMs are unlikely to occur with binding of RNA. This is in contrast to other dual RRM−RNA complexes that lack these salt bridges (*46*, *47*) and change structure appreciably on binding RNA but parallels the UP1−telomeric DNA crystal structure in which the salt bridges are conserved (*1*). We also observed that in our model the RNA is accommodated along an electropositive surface of the protein and lies close to the inter-RRM linker, features that are common to most complexes of RNA with tandem RRMs for which three-dimensional structures have been reported.

Residues 180−189 of hnRNP A2, which are immediately C-terminal to RRM-2, are required for realization of the full binding capacity for A2RE11 (*2*). Our model encompasses residues 1−177, and while we were unable to reliably model residues 180−189, the shortened C-terminus observed in our model extends away from the β-sheet surface and is in the proximity of the RNA. It is feasible that the addition of 12 extra amino acids at the C-terminus of the homology model could introduce a molecular latch capable of stabilizing the bound RNA, as observed in the C-terminally extended construct of U1A (*51*).

Little is known about the structure of GRDs. We have presented experimental evidence that supports the possibility that these domains fold into recognizable secondary structures. Our CD data for both wild-type hnRNP A2 and the isolated GRD ($A2^{190-341}$) are consistent with the structure of nucleolin (*31*): for both, the GRDs are largely devoid of helix but contain significant proportions of β-structure. The GRDs can potentially self-assemble into more complex supersecondary structures (*29*, *31*). The role of this domain in hnRNP A2 remains unknown; however, in other RNA binding proteins, GRDs are thought to be involved in protein−protein (*27*) or protein−RNA (*26*) interactions.

In summary, we have explored the structures of hnRNP A2, its RRM and Gly-rich modules, and their complexes with the A2RE11. We have provided evidence from CD spectra and homology modeling that the structure of the tandem RRM domains of hnRNP A2 parallels that of its paralogue, hnRNP A1, previously deduced from X-ray crystallography. We have shown the GRD to have a secondary structure dominated by β-sheet and β-turns. However, our sedimentation equilibrium and CD data are consistent with the tandem RRM region being monomeric, in contrast with the UP1 region of hnRNP A1 complexed with the telomeric ssDNA repeat. These data also provide independent evidence of the binding of two A2RE11 oligoribonucleotides to the tandem RRMs, again in contrast to hnRNP A1, which exhibits a 1:1 tandem RRM:oligonucleotide stoichiometry. Finally, the three-dimensional model presented herein is consistent with the published structures of other tandem RRM−RNA complexes, our in silico experiments, and other biophysical data presented here and elsewhere (*2*, *9*, *13*).

ACKNOWLEDGMENT

We thank Dr. Adrian Krainer (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) for his kind gift of pET9c-A2 plasmid and Dr. B. Hawkins, Prof. D. Winzor, and Dr. L. Carrington (The University of Queensland) for assistance with the analytical ultracentrifugation experiments. $A2^{1-189}$ protein was produced at the University of Queensland Protein Expression Facility. We also thank Dr. N. Cowieson for assistance in preparing the figures.

SUPPORTING INFORMATION AVAILABLE

A SDS−polyacrylamide gel showing the purity of the recombinant proteins and a table of single-stranded RNA ligands and their cognate tandem RRM-containing proteins. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and 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.
2. Shan, J., Moran-Jones, K., Munro, T. P., Kidd, G. J., Winzor, D. J., Hoek, K. S., and Smith, R. (2000) Binding of an RNA trafficking response element to heterogeneous nuclear ribonucleoproteins A1 and A2, *J. Biol. Chem. 275*, 38286−38295.
3. Shahied-Milam, L., Soltaninassab, S. R., Iyer, G. V., and LeStourgeon, W. M. (1998) The heterogeneous nuclear ribonucleoprotein C protein tetramer binds U1, U2, and U6 snRNAs through its high affinity RNA binding domain (the bZIP-like motif), *J. Biol. Chem. 273*, 21359−21367.
4. Lerga, A., Hallier, M., Delva, L., Orvain, C., Gallais, I., Marie, J., and Moreau-Gachelin, F. (2001) Identification of an RNA binding specificity for the potential splicing factor TLS, *J. Biol. Chem. 276*, 6807−6816.
5. McKay, S., and Cooke, H. (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat $TTAGGG_n$, *Nucleic Acids Res. 20*, 6461−6464.
6. Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. (1993) hnRNP proteins and the biogenesis of mRNA, *Annu. Rev. Biochem. 62*, 289−321.
7. Mayeda, A., Munroe, S. H., Caceres, J. F., and Krainer, A. R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins, *EMBO J. 13*, 5483−5495.
8. Kwon, S., Barbarese, E., and Carson, J. H. (1999) The *cis*-acting RNA trafficking signal from myelin basic protein mRNA and its cognate *trans*-acting ligand hnRNP A2 enhance cap-dependent translation, *J. Cell Biol. 147*, 247−256.
9. Munro, T. P., Magee, R. J., Kidd, G. J., Carson, J. H., Barbarese, E., Smith, L. M., and Smith, R. (1999) Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking, *J. Biol. Chem. 274*, 34389−34395.
10. Ishikawa, F., Matunis, M. J., Dreyfuss, G., and 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.
11. Kamma, H., Fujimoto, M., Fujiwara, M., Matsui, M., Horiguchi, H., Hamasaki, M., and Satoh, H. (2001) Interaction of hnRNP A2/B1 isoforms with telomeric ssDNA and the *in vitro* function, *Biochem. Biophys. Res. Commun. 280*, 625−630.
12. Myers, J. C., and Shamoo, Y. (2004) Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM), *J. Mol. Biol. 342*, 743−756.

13. Moran-Jones, K., McLure, L., Kennedy, D., Reddel, R. R., Sara, S., and Smith, R. (2005) hnRNP A2, a potential ssDNA/RNA molecular adapter at the telomere, *Nucleic Acids Res. 33*, 486−496.
14. Tsukamoto, H., Boado, R. J., and Partridge, W. M. (1996) Differential expression in glioblastoma multiforme and cerebral hemangioblastoma of cytoplasmic proteins that bind two different domains within the 3′-untranslated region of the human glucose transporter 1 (GLUT1) messenger RNA, *J. Clin. Invest. 97*, 2823−2832.
15. Hamilton, B. J., Nichols, R. C., Tsukamoto, H., Boado, R. J., Pardridge, W. M., and Rigby, W. F. (1999) hnRNP A2 and hnRNP L bind the 3′UTR of glucose transporter 1 mRNA and exist as a complex *in vivo*, *Biochem. Biophys. Res. Commun. 261*, 646−651.
16. Griffin, M. E., Hamilton, B. J., Roy, K. M., Du, M., Willson, A. M., Keenan, B. J., Wang, X. W., and Nichols, R. C. (2004) Post-transcriptional regulation of glucose transporter-1 by an AU-rich element in the 3′UTR and by hnRNP A2, *Biochem. Biophys. Res. Commun. 318*, 977−982.
17. Chen, H., Hu, B., Allegretto, E. A., and Adams, J. S. (2000) The vitamin D response element-binding protein. A novel dominant-negative regulator of vitamin D-directed transactivation, *J. Biol. Chem. 275*, 35557−35564.
18. Ainger, K., Avossa, D., Morgan, F., Hill, S. J., Barry, C., Barbarese, E., and Carson, J. H. (1993) Transport and localization of exogenous myelin basic protein mRNA microinjected into oligodendrocytes, *J. Cell Biol. 123*, 431−441.
19. Hoek, K. S., Kidd, G. J., Carson, J. H., and Smith, R. (1998) hnRNP A2 selectively binds the cytoplasmic transport sequence of myelin basic protein mRNA, *Biochemistry 37*, 7021−7029.
20. Shan, J., Munro, T. P., Barbarese, E., Carson, J. H., and Smith, R. (2003) A molecular mechanism for mRNA trafficking in neuronal dendrites. *J. Neurosci. 23*, 8859−8866.
21. 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.
22. Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R., and Steitz, T. A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution, *Nat. Struct. Biol. 4*, 215−222.
23. Lee, A. L., Kanaar, R., Rio, D. C., and Wemmer, D. E. (1994) Resonance assignments and solution structure of the second RNA-binding domain of Sex-lethal determined by multidimensional heteronuclear magnetic resonance, *Biochemistry 33*, 13775−13786.
24. Allain, F. H., Gilbert, D. E., Bouvet, P., and Feigon, J. (2000) Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target, *J. Mol. Biol. 303*, 227−241.
25. Oberstrass, F. C., Auweta, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L., and Allain, F.-H. (2005) Structure of PTB bound to RNA: Specific binding and implications for splicing regulation, *Nature 309*, 2054−2057.
26. Kumar, A., and Wilson, S. H. (1990) Studies of the strand-annealing activity of mammalian hnRNP complex protein A1, *Biochemistry 29*, 10717−10722.
27. Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins, *J. Mol. Biol. 259*, 337−348.
28. Kouklis, P. D., Hutton, E., and Fuchs, E. (1994) Making a connection: Direct binding between keratin intermediate filaments and desmosomal proteins, *J. Cell Biol. 127*, 1049−1060.
29. Parry, D. A., Marekov, L. N., Steinert, P. M., and Smith, T. A. (2002) A role for the 1A and L1 rod domain segments in head domain organization and function of intermediate filaments: Structural analysis of trichocyte keratin, *J. Struct. Biol. 137*, 97−108.
30. Ainger, K., Avossa, D., Diana, A. S., Barry, C., Barbarese, E., and Carson, J. H. (1997) Transport and localization elements in myelin basic protein mRNA, *J. Cell Biol. 138*, 1077−1087.
31. Ghisolfi, L., Joseph, G., Amalric, F., and Erard, M. (1992) The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties, *J. Biol. Chem. 267*, 2955−2959.
32. Johnson, W. C. J. (1990) Protein secondary structure and circular dichroism: A practical guide, *Proteins 7*, 205−214.
33. Johnson, W. C. J. (1988) Secondary structure of proteins through circular dichroism spectroscopy, *Annu. Rev. Biophys. Biophys. Chem. 17*, 145−166.
34. Cohn, E. J., and Edsall, J. T. (1943) *Proteins, amino acids and peptides as ions and dipolar ions*, Reinhold, New York.
35. Durchschlag, H. (1986) *Thermodynamic data for biochemistry and biotechnology*, Springer-Verlag, Berlin.
36. Wills, P. R., Jacobsen, M. P., and Winzor, D. J. (1996) Direct analysis of solute self-association by sedimentation equilibrium, *Biopolymers 38*, 119−130.
37. Greer, J. (1990) Comparative modeling methods: Application to the family of the mammalian serine proteases, *Proteins 7*, 317−334.
38. Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs *Structure 5*, 559−570.
39. Crowder, S. M., Kanaar, R., Rio, D. C., and Alber, T. (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal, *Proc. Natl. Acad. Sci. U.S.A. 96*, 4892−4897.
40. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles, *Nature 356*, 83−85.
41. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
42. Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking, *J. Mol. Biol. 267*, 727−748.
43. Smith, J. A., and Pease, L. G. (1980) Reverse turns in peptides and proteins, *CRC Crit. Rev. Biochem. 8*, 315−399.
44. Woody, R. (1995) Circular dichroism, *Methods Enzymol. 246*, 34−71.
45. Vitali, J., Ding, J., Jiang, J., Zhang, Y., Krainer, A. R., and Xu, R. M. (2002) Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1, *Nucleic Acids Res. 30*, 1531−1538.
46. Handa, N., Nureki, O., Kurimoto, 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.
47. Wang, X., and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein *Nat. Struct. Biol. 8*, 141−145.
48. 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.
49. Allain, F. H., 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.
50. He, Y. (2005) Roles of heterogeneous nuclear ribonucleoproteins A and B in cell proliferation, *J. Cell Sci. 118*, 3173−3183.
51. Avis, J. M., Allain, F. H., Howe, P. W., 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.

BI052440E