Tertiary Structure of RBD2 and Backbone Dynamics of RBD1 and RBD2 of the Human U1A Protein Determined by NMR Spectroscopy[†,‡]

Jirong Lu and Kathleen B. Hall*

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

Received April 25, 1997; Revised Manuscript Received June 11, 1997[⊗]

ABSTRACT: The human U1A protein has two putative RNA binding domains, one at the N-terminal region of the protein (RBD1) and the other at the C-terminal end (RBD2). RBD1 binds tightly and specifically to one of the stem loops of the U1 snRNA, as well as to its own 3′-UTR. In contrast, RBD2 does not appear to associate with any RNA. The two domains share 25% amino acid identity, and both have the same $\beta\alpha\beta-\beta\alpha\beta$ secondary structure fold. In this work, $^{13}C/^{15}N/^{1}H$ multidimensional NMR methods were used to obtain side-chain assignments for RBD2, and then the tertiary structure was calculated using a distance geometry/simulated annealing algorithm that employs pairwise Gaussian metrization. RBD2 is shown to fold into an α/β sandwich with a four-stranded antiparallel β-sheet, which is the typical global topology of these domains. Specific structural features of RBD2 include a β-bulge in β2, N-capping boxes for both α-helices, and an extremely shallow twist of its β-sheet. The ^{15}N backbone dynamics of these two structurally homologous RBDs are significantly different, compared using order parameters and T_2 exchange terms in the Lipari and Szabo model-free formalism. Conformational exchange observed in RBD1, which is absent in RBD2, may correlate to the mechanism of RNA binding.

The RNA binding domain (RBD),[1] also known as RNA recognition motif (RRM), or ribonucleoprotein motif (RNP), is the most widely found and best characterized RNA binding motif. This motif consists of about 90 amino acids and is identified by the sequence and position of two consensus sequences, the RNP-2 hexamer and RNP-1 octamer. It is present as a single or multiple copy in proteins that bind to a variety of RNAs. These proteins play important roles in pre-mRNA splicing and posttranscriptional regulation of genes (Dreyfuss et al., 1988, 1993; Bandziulis et al., 1989; Mattaj, 1989, 1993; Keene & Query, 1991; Kenan et al., 1991; Burd & Dreyfuss, 1994a). Several of these motifs have been characterized structurally, including the N-terminal RBD of the human U1A snRNP protein (Nagai et al., 1990; Hoffman et al., 1991; Howe et al., 1994; Avis et al., 1996), the hnRNP C protein (Wittekind et al., 1992), the second RBD of *Drosophila* Sxl protein (Lee et al., 1994), and hnRNP A1 (Garrett et al., 1994). All contain a $\beta\alpha\beta-\beta\alpha\beta$ secondary structure, which folds into an α/β sandwich with a four-stranded antiparallel β-sheet packed against two α-helices. There are a number of conserved hydrophobic residues interspersed throughout the sequence (Burd & Dreyfuss, 1994a; Birney et al., 1993); most are located in the interior of the protein, where they contribute to structure and folding of this motif. Two highly conserved aromatic residues, one in RNP-2 and the other in RNP-1, located on the surface of the β-sheet, have been shown to be in direct contact with RNA (Merrill et al., 1988; Oubridge et al., 1994; Stump & Hall, 1995).

The human U1A protein contains two RBDs (RBD1 and RBD2) joined by a flexible linker region rich in proline and lysine; their sequence alignment based on secondary structure is shown in Figure 1. The structures of the free RBD1 and the RBD1:RNA complex have been solved by X-ray crystallography (Nagai et al., 1990; Oubridge et al., 1994) and NMR (Avis et al., 1996; Allain et al., 1996). RBD1 binds to stem/loop II of the U1 snRNA (Scherly et al., 1989) and to a short RNA hairpin with very high affinity and specificity (Jessen et al., 1991; Hall & Stump, 1992; Hall, 1994). It also binds to its own mRNA 3′-UTR (Boelens et al., 1993; van Gelder et al., 1993). The C-terminal RBD (RBD2) adopts a stable conformation and has a secondary structure similar to that found in the other member of this family but does not appear to interact with any RNA substrate (Lu & Hall, 1995).

The apparent topological similarity of many RBDs raises a fundamental question. What features of the protein direct specific RNA association? While specific amino acid side chains are obviously critical determinants for RNA recognition, other more subtle differences in protein structure and/or energetics can also control RNA association, as for example through protein dynamic motions, structural stability, or side-chain interactions on the binding surface. In these experiments, we have determined the tertiary structure of the human U1A C-terminal RBD2 using NMR spectroscopy and compared its structure with those of other RBDs. The backbone dynamics of U1A RBD1 and RBD2 are compared using amide ^{15}N relaxation measurements, looking for similarities and differences between these two structurally homologous and functionally different proteins.

[†] This work is supported by the Lucille P. Markey Charitable Trust (90-47) (K.B.H.), the NIH (GM46318 to K.B.H. and postdoctoral fellowship GM16739 to J.L.), and the Council for Tobacco Research.

[‡] The coordinates have been deposited in the Brookhaven Protein Data Bank (code name 2u1a).

* Address correspondence to this author. Phone: 314-362-4196. Fax: 314-362-7183. E-mail: hall@bionmr3.wustl.edu.

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1997.

[1] Abbreviations: RBD, RNA binding domain; NOE, nuclear Overhauser enhancement; snRNP, small nuclear ribonucleoprotein particle; UTR, untranslated region; 3D, three dimensional; 2D, two dimensional; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; HCCH-TOCSY, proton−carbon−carbon−proton correlation using carbon total correlated spectroscopy; CSI, chemical shift index; T_1, longitudinal relaxation time; T_2, transverse relaxation time; RMSD, root mean square difference.

S0006-2960(97)00981-1 CCC: $14.00

```
                    β1                  α1
RBD1 (1-36)    MAVPETRPNHTIYINNLNEKIKKDELKKSLYAIFSQ
RBD2 (195-230) MAPAQPLSENPPNHILFLTNLPEET    NELMLSMLFNQ
                  β2          β3        α2
RBD1 (37-72)     FGQILDILVSRSLKMRGQAFVIFKEVSSATNALRSM
RBD2 (231-264)   FPGFKEVRLV  PGRHDIAFVEFDNEVQAGAARDAL
                    β4
RBD1 (73-95)     QGFPFYDKPMRIQYAKTDSDIIA
RBD2 (265-282)   QGFKITQNNAMKISFAKK
```

FIGURE 1: Sequence alignment of RBD1 and RBD2 of human U1A protein based on secondary structure. The secondary structure elements are denoted on the top, and RNP-2 and RNP-1 consensus regions are shown in bold.

MATERIALS AND METHODS

Sample Preparation. The genes encoding the C-terminal RBD of human U1A (Sillikens et al., 1988) from residue 195 to 282 (referred to as U195A or RBD2) and the N-terminal RBD from residue 1 to 95 (referred to as RBD1) were PCR subcloned into an *Escherichia coli* expression vector under control of the tac promoter, and protein was expressed and purified. ^{15}N and ^{15}N/^{13}C uniformly labeled samples of RBD2 were prepared for NMR experiments as described previously (Lu & Hall, 1995).

NMR Spectroscopy. All NMR spectra were recorded at 25 °C with a Varian Unity-500 spectrometer equipped with a triple resonance Nalorac 5 mm probe with Z-gradients. For assignments, samples contained 2 mM ^{15}N or ^{15}N/^{13}C uniformly labeled proteins in 100 mM KCl, 30 mM potassium phosphate, and 0.02% NaN_3, pH 5.5. A 3D HCCH-TOCSY (Bax et al., 1990) experiment was collected to obtain side-chain ^{13}C and ^{1}H assignments. ^{15}N-Edited NOESY-HMQC and ^{13}C-edited NOESY-HSQC experiments based on 3D ^{13}C-edited NOESY-HMQC experiments (Ikura et al., 1990) were performed to obtain NOE distance constraints. When these NOESY spectra were collected at 1.0 mM protein concentration, more cross-peaks were observed in addition to those present at 2 mM protein concentration. These spectra were used to generate distance constraints for structure calculations. In the ^{15}N-NOESY-HMQC, the ^{15}N spectral width was 1520 Hz with the carrier frequency set at 120.2 ppm. In the 3D HCCH-TOCSY and ^{13}C-NOESY-HSQC experiments, the ^{1}H spectral width was 3499.87 ppm with the carrier set at 2.95 ppm, and the ^{13}C spectral width was 2765 Hz with the carrier at 46.0 ppm. The spectrum was folded, and each ^{13}C plane corresponds to three different ^{13}C chemical shifts. Mixing times for 3D ^{15}N-edited NOESY-HMQC, 3D ^{13}C-edited NOESY-HSQC, and HCCH-TOCSY were 150, 150, and 26 ms, respectively. Chemical shifts of ^{1}H and ^{13}C were referenced to TSP.

^{15}N T_1, T_2, and NOE spectra were recorded at 25 °C using pulse sequences described by Farrow et al. (1994). ^{15}N uniformly labeled protein at 0.5 mM in 100 mM KCl, 30 mM potassium phosphate, and 0.02% NaN_3, pH 5.5, was used. T_1 values were measured from the spectra recorded with seven different delay times: T = 5.6, 56, 112, 224, 448, 728, and 1120 ms. T_2 values were determined from spectra recorded with delays of T = 16, 32, 48, 80, 112, 144, and 176 ms. Duplicate spectra were recorded for T = 5.6 and 448 ms (T_1) and T = 16 and 112 ms (T_2) to estimate the noise levels. ^{1}H−^{15}N steady-state NOE values are obtained by recording spectra with (NOE) and without (NONOE) the use of ^{1}H saturation applied before the start of the experiment. All of these spectra were recorded with 128 × 512 complex matrices with 64 scans per t_1 point. Spectral widths of 1520 and 6499.8 Hz were employed in F_1 and F_2, respectively. All spectra were collected with the hypercomplex method (States et al., 1982) and were processed identically with Lorentzian−Gaussian functions applied in both dimensions, using Varian software on a Sun Sparc Workstation. Analysis and peak picking were carried out using NMR Compass (Molecular Simulation Inc.) on an SGI Indigo.

Structure Calculation. DISTGEOM, a distance geometry program using pairwise Gaussian metrization, was used to compute all structures, following the calculation protocol developed by Hodsdon et al. to solve the structure of a larger (15.4 kDa) protein (Hodsdon et al., 1996; Hodsdon & Cistola, 1997).

Structures were initially equilibrated at 200 °C for 1000 steps with a time step of 0.04 ps, followed by 10000 dynamics steps of cooling to 0 °C in steps of 0.02 ps. In these calculations, no general terms for van der Waals energies, torsional energies, or electrostatic energies were included to refine the structures. The penalty function for the molecular mechanics calculations consisted of a series of atom-based energy terms to enforce local bond geometry, prevent nonbonded contacts, and implement the experimental constraints. A flat-well parabolic penalty function restrained the ϕ torsion angle between −80° and −40° and ψ between −60° and −20° for two α-helices (Glu221-Asn229 and Glu254-Leu264) and ϕ between −190° to −80° and ψ between 90° to 180° for four β-strands (Ile209-Thr213, Glu236-Arg238, Ile246-Phe251, and Lys276-Ser278), weighted such that a 10° deviation in either direction resulted in a 1 kcal pseudoenergy. A similar penalty function restrained the NOE distances to their upper and lower bounds such that 1 Å violation resulted in a 1 kcal contribution to the penalty function. Except where noted, all NOE constraints were assigned with an upper bound of 5 Å and a lower bound of 1.86 Å. Restraints involving methylene protons, single methyl and geminal methyl protons, and aromatic ring protons required a correction to their upper bounds in order to account for a correlation to either of their prochiral atoms. The upper bounds were increased by 1.76, 4.66, and 4.30 Å for methylene and single methyl, geminal methyl, and aromatic ring protons, respectively. A separate restraint was entered into the bounds matrix for each of the prochiral resonances with a weighting in the penalty function equal to the inverse of its degeneracy. For example, a single NOESY correlation involving a methylene proton was considered as two separate distance constraints with upper bounds of 6.76 Å and each weighted by 0.5 in the penalty function. In those cases where restraints could be identified to resolved prochiral resonances, the corrections can be ignored, since both of the prochiral protons must be within 5 Å of the coupling partner. In that case, the single restraint was replaced with a pair of restraints involving the specific prochiral protons.

Backbone Dynamic Data Processing and Analysis. The peak heights of 2D spectra were obtained using NMR Compass. T_1 and T_2 values were obtained from the fit of the intensity as a function of time to the following equation using Kaleidegraph software:

$$I(t) = I(0)\exp(-t/T_{1,2}) \quad (1)$$

where $I(t)$ and $I(0)$ are the intensity of the peak after a delay of t and 0, respectively. The steady-state NOE values were

determined from the ratios of the intensities with or without proton saturation.

To describe the measured T_1, T_2, and NOE relaxation times in terms of the motion of the backbone $^{15}N-^{1}H$ vector, the Lipari and Szabo (1982a,b) model-free form of the spectral density function was used to determine the motional parameters (Dellwo & Wand, 1989). For particular residues, the extended form of the spectral density was also used (Clore et al., 1990b). The assumption in these analyses is that the overall tumbling is isotropic, and the internal motions can be separable into discrete uncorrelated contributions to the spectral density. In its simplest form, the model-free spectral density is

$$J(\omega) = (2/5)(S^2\tau_m/(1 + \omega^2\tau_m{}^2) + (1 - S^2)\tau/(1 + \omega^2\tau^2)) \quad (2)$$

where τ_m is the rigid body rotational correlation time of the (globular) protein; $1/\tau = 1/\tau_m + 1/\tau_e$, where τ_e is an effective correlation time for more rapid local motion of the N−H vector; and S^2 is the order parameter, reporting on the spatial distribution due to the internal motion of the N−H vector (Lipari & Szabo, 1982a,b). An extended form of the model-free spectral density function, which includes internal motions with two distinct time scales differing by at least an order of magnitude, was also used in some cases (Clore et al., 1990a,b; Farrow et al., 1994). This spectral density function is expressed as

$$J(\omega) = (2/5)(S^2\tau_m/(1 + \omega^2\tau_m{}^2) + (S_f{}^2 - S^2)\tau/(1 + \omega^2\tau^2)) \quad (3)$$

with order parameter $S^2 = S_s{}^2S_f{}^2$, where S_s and S_f represent slow and fast internal motions, respectively. The effective correlation time for the slow motion, τ_s, is included in the relationship $1/\tau = 1/\tau_s + 1/\tau_m$.

These expressions for the spectral density were used to fit the measured T_1, T_2, and NOE relaxation times. Thus, for example, $1/T_1 \propto J(\omega)$, where $J(\omega)$ is either (2) or (3). For some amides, an additional chemical exchange term was added when fitting transverse relaxation rates:

$$1/T_2 = 1/T_2{}^{DD} + 1/T_2{}^{CSA} + R_{ex} \quad (4)$$

in which DD and CSA indicate dipole−dipole and chemical shift anisotropy contributions to T_2 relaxation rates. The exchange term, R_{ex}, was not always required to fit the data, and as discussed in Clore et al. (1990a), the assumption is made that T_2 relaxation of a given amide is fit by including terms for either exchange broadening or additional (τ_s) motion, not both. From this expression, it is also clear that the motional regime reported by R_{ex} (microsecond to millisecond) is separable from the nanosecond/picosecond motion of the N−H vector reported by the order parameter, S^2.

Data were analyzed with software generously provided by Prof. Lewis Kay. To derive τ_m, the ratio of T_1/T_2 was calculated for each residue. For those residues with T_1/T_2 values within one standard deviation of the mean, relaxation parameters were fit assuming a fixed value of τ_m with the model-free parameter S^2 and allowed to vary until the residual χ^2

$$\chi^2 = \sum(T_1{}^{exp} - T_1{}^{c})^2/\sigma_1{}^2 + (T_2{}^{exp} - T_2{}^{c})^2/\sigma_2{}^2 + (NOE^{exp} - NOE^{c})^2/\sigma_3{}^2$$

was minimized. This global χ^2 target function is the summation over all amino acid residues; the superscript indicates experimental (exp) and calculated (c) values. The standard deviations (σ) of the T_1, T_2, and NOE values for each peak are estimated from experimental data.

The experimental relaxation data were fit with the following five different models, using the three formulations of the model-free approach described above with τ_m fixed, to determine the best fit to the data: (1) S^2 with τ_e fixed to zero; (2) S^2 and τ_e; (3) S^2 and R_{ex}; (4) S^2, τ_e, and R_{ex}; (5) S^2, τ_s, and $S_f{}^2$.

RESULTS

Side-Chain Assignments of RBD2. The assignment of the backbone resonances and determination of secondary structure have been described previously (Lu & Hall, 1995). The side-chain ^{1}H and ^{13}C resonances were assigned from ^{15}N-separated 3D TOCSY-HMQC and ^{13}C-separated 3D HCCH-TOCSY experiments. Complete spin systems can be obtained with 3D HCCH-TOCSY experiments due to the high efficiency of magnetization transfer as shown in Figure 2. A single spin system will appear at different planes corresponding to the chemical shifts of the side-chain carbons, allowing sequence-specific assignments for nearly all side chains. In cases where side-chain carbons have similar chemical shifts, ^{15}N TOCSY-HMQC experiments with shorter mixing times were used to assign the positions of protons in the side chain. A few additional backbone assignments were obtained after analyzing ^{13}C-edited NOESY-HSQC spectra. In addition, strong NOE cross-peaks were observed between the H_α protons of the residues preceding Pro205, Pro216, Pro232, and Pro241 and the H_δ protons of those proline rings, allowing assignments of the proline residues. These strong NOEs also indicate that peptide bonds between proline residues and preceding residues are in the *trans* configuration.

NOE Assignment and Distance Constraints. To calculate the tertiary structure of this domain, a total of 3305 NOE cross-peaks were obtained from 3D ^{15}N-separated NOESY-HMQC and 3D ^{13}C-separated NOESY-HSQC experiments. NOE cross-peaks were assigned by matching NOE cross-peak frequencies against existing assignments. For symmetry-related cross-peaks, only one distance constraint is used. Those that were ambiguous in the initial stage were further assigned using the procedure described.

The distances defined by these NOEs were described by an upper bound of 5.0 Å and a lower bound of 1.86 Å and were not otherwise differentiated. Appropriate upper bounds corrections were made for restraints involving nonstereospecifically assigned resonances as described in Materials and Methods. At a concentration of 2 mM RBD2, there was unexpectedly rapid relaxation of resonances, with the result that fewer NOEs than predicted were observed. This behavior is consistent with anomalously slow tumbling due to nonspecific self-association, high viscosity of associated solvent, or viscous drag from the disordered tail. It is important to note that there were no changes in chemical shift with protein concentration, and thus the apparent association appears to be nonspecific. At 1 mM RBD2, relaxation properties of the system were normal, and the greater number of NOEs observed were typical for a protein of this size. These data were used to generate structures,

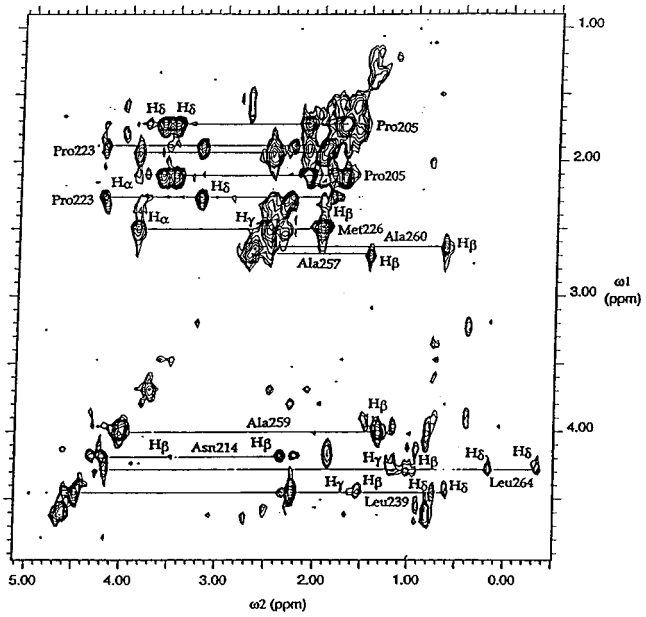

FIGURE 2: 2D slice of ^{13}C-separated 3D HCCH-TOCSY at ^{13}C = 74.03, 54.03, and 34.03 ppm. Three ^{13}C frequencies are represented due to aliasing in ^{13}C dimension. All resonances are labeled with their assignments. Lines represent individual amino acid spin systems.

but the quantitation of NOE intensities is complicated by unpredictable relaxation properties that could make interpretation unreliable. To obtain accurate structures, the NOE intensities were therefore not classified as strong, medium, or weak.

Tertiary Structure Calculation. All structures were calculated with DISTGEOM, a component of the TINKER molecular modeling package, using a protocol similar to that developed by Hodsdon et al. (1996). DISTGEOM differs from other protocols in three critical ways. First, it uses an iteratively optimized Gaussian distribution to select trial distances during metrization, rather than the usual uniform distribution. As described by Oshiro et al. (1991), a Gaussian distribution centered around a fractional distance of 0.6−0.7 is appropriate for larger globular proteins; DISTGEOM further expands this idea to iteratively optimize the distribution (J. W. Ponder, personal communication). Second, the metrization of DISTGEOM uses pairwise elements of the distance matrix, not the atoms in the molecules, resulting in more extensive sampling of conformational space. Third, the metrization step is more computationally efficient than others, permitting its implementation for larger proteins. To summarize, this protocol consists of a distance geometry calculation with random pairwise 5% metrization and full structure embedding, followed by a brief minimization of the penalty function.

The structures were calculated in an iterative fashion, first with 600 unambiguous NOE distance constraints. Dihedral angle constraints for both α-helices were also included, determined from CSI, backbone NOEs, and $^3J_{HN\alpha}$ constants. Ambiguous NOE distance constraints were then carefully examined, and 1300 additional NOE distance constraints were assigned on the basis of secondary structure and the initial global fold. Those NOE constraints that were still ambiguous at this stage were further assigned if they were consistent with at least 18 of 20 structures calculated.

Table 1: Final Restraints and Restraints Violations

Distance Restraints	
total	2049 (23.3/residue)
intraresidue	276
interresidue	614
long range	1127
H-bonds	32
Torsional Restraints	
helical (Φ/Ψ)	40
β-strands (Φ/Ψ)	42
Restraints Violations[a]	
upper bounds	
average violation (Å)	0.04 ± 0.03
largest violation (Å)	0.11
no. of violations	36 of 40980[b]
lower bounds	
average violation (Å)	0.05 ± 0.06
largest violation (Å)	−0.18
no. of violations	94 of 40980[b]

[a] Analyzed using AQUA version 0.40 and PROCHECK-NMR version 3.4.2 (Rullman, 1996; MacArthur et al., 1994). [b] The product of the number of restraints and the number of structures in the final ensemble.

The final round of structure calculation included several additional constraints. Now the dihedral angle constraints for β-strands were included (obtained from CSI, backbone NOEs, and $^3J_{HN\alpha}$), and the result was to improve pairwise RMSD of C_α by 0.1 Å, while having no effect on the twist of the β-sheet. H-bonds were included for nonexchangeable amides in stretches of regular secondary structure which have been identified by the backbone NOEs, CSI, and H-exchange data (Lu & Hall, 1995). Hydrogen bond constraints were implemented as 32 NH−O (1.8−2.0 Å) and 32 N−O (2.7−3.0 Å) NOE-type distance constraints (Kuntz et al., 1989) for a total of 32 H-bonds. The penalty function values of the final ensemble of 20 structures were 1.6 ± 0.1; values less than 10 indicate good agreement with the experimental restraints (Hodsdon et al., 1996). The statistics for the final set of constraints and the final restraint violations are listed in Table 1.

Structure Analysis. For each residue, the number of NOE constraints varies and reflects the secondary structure of the protein (Figure 3A). In particular, residues 195−207 at the N-terminus which are not included in the figure have very few NOEs; loop regions also have fewer NOEs. From the 40 structures calculated using DISTGEOM (see Materials and Methods), the 20 structures with the lowest penalty function were selected. The average pairwise C_α RMSD of the calculated 20 structures is 1.11 Å (0.73 Å RMSD to the average coordinates) for residues 208−282 and 0.54 Å (0.4 Å RMSD to the average coordinates) for residues involved in secondary structure regions of the protein. The actual values for each residue-based pairwise RMSD are shown in Figure 3B. As expected, those regions that are involved in secondary structures are better defined than those involved in loop regions. Residues 195−207 at the N-terminus of the protein are largely disordered in the family of the calculated structures.

The quality of the generated structures was analyzed by PROCHECK version 3.0 (Laskowski et al., 1993), and parameters are summarized in Table 2. The first 12 residues in the N-terminal region were excluded from this analysis; 98.5% of the residues were found in allowed region of the Ramachandran plot with 63.3% in the most favored regions.

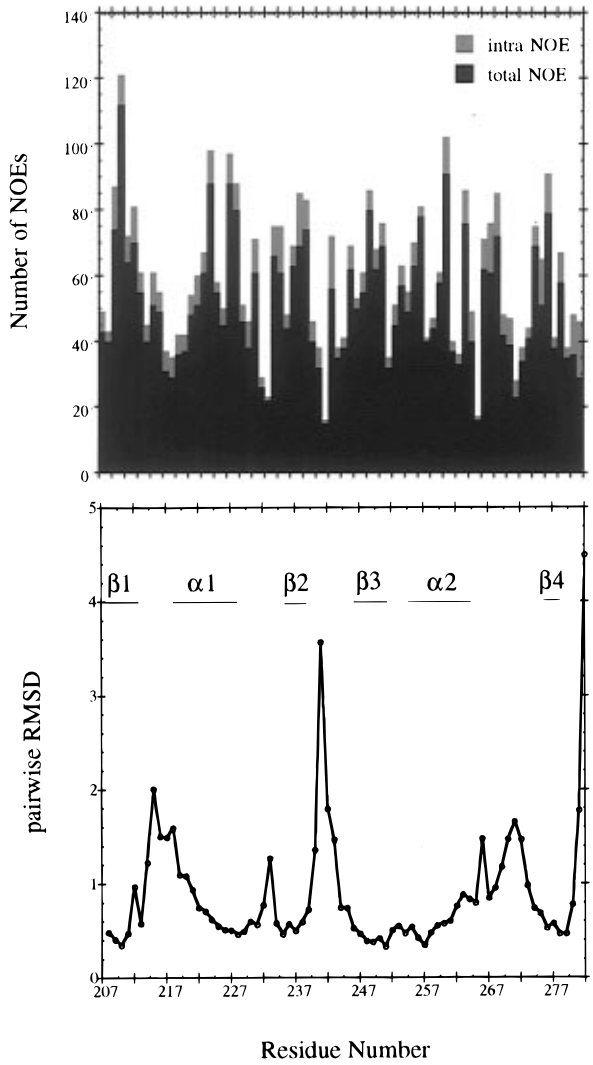

FIGURE 3: (A, top) Distribution of NOEs used for structure calculation. Secondary structure is indicated. Intra NOE and total NOE represent the number of intraresidue NOEs and the total number of NOEs from the residue, respectively. (B, bottom) Residue-based pairwise root mean square deviation of backbone C_α for 20 structures.

The inclusion of dihedral angle constraints in β-strands in the last round of structure calculations increased the percentage of the residues in the most favored region to 66.9%. The apparently low percentage of the residues in the most favored region of Ramachandran plot is due to the Tinker algorithm which utilizes a limited penalty function that omits a general torsional energy term. The inclusion of such a term would be expected to localize backbone torsion angles to more favorable values. Main-chain and side-chain statistics determined from the PROCHECK analysis are shown in Table 2. On the basis of these parameters, stereochemical quality of the family of the structures is comparable to those of 2.5 Å resolution crystal structures.

Description of the Structures. RBD2 of the human U1A protein folds to a $\beta\alpha\beta-\beta\alpha\beta$ secondary structure, similar to that found in other members of this family. Four β-strands form an antiparallel β-sheet on one side of protein. The two RNP consensus sequences, RNP-1 (244−251) and RNP-2 (210−215), are located in two adjacent β-strands, corresponding to β3 and β1, respectively. The stereo diagram of the C_α superposition of 20 structures and the ribbon diagram of 1 of the representative structures are shown in the Figure

Table 2: Stereochemical Statistics for RBD2 of Human U1A Protein

statistics	ensemble average
Ramachandran Plot Statistics[a]	
residues in allowed regions	98.5% (98.0%)
most favored regions	63.3% (66.9%)
additionally allowed regions	29.1% (27.3%)
generously allowed regions	6.1% (3.7%)
residues in disallowed regions	1.5% (2.0%)
Main-Chain Statistics	
standard deviations of ω (deg)	2.1
no. of bad contacts/100 residues	9.6
standard deviation of "zeta-angle" (deg)	3.0
standard deviation of H-bond energy (kcal/mol)	0.8
overall *G*-factor	−0.8
Side-Chain Statistics	
χ_1 gauche minus	25.6
χ_1 trans	25.3
χ_1 gauche plus	24.8
χ_1 "pooled"	24.8
χ_2 trans	29.7

[a] Numbers in parentheses are the percentage of the residues after inclusion of dihedral angle constraints for β-strands.

4. Two aromatic residues from β1 and β3 (Phe211 and Phe248 in this protein) have been implicated in RNA binding by RBD proteins; their side chains are exposed on the surface of the β-sheet. In RBD2, NOEs between Ile209 on β1 and Phe248 on β3 and between Phe211 and Ile246 are clearly observed, resulting in the juxtapositions as shown in Figure 5.

β-Bulges have been observed in RBD2 of Sxl protein (Ile40 and Val41) (Lee et al., 1994) and RBD1 of human U1A protein (Ile40 and Leu41) (Nagai et al., 1990), as well as in hnRNP C (Wittekind et al., 1992). A β-bulge in U1A RBD2 β2 involving residues Phe234 and Lys235 is observed in all 20 structures, consistent with NOE data. There is no NOE observed between the H_α of Phe251 and H_α of Lys235 as expected for a normal β-sheet, but there are NOEs between the H_α of Phe234 and both the H_α and NH of Phe251, consistent with Lys235 forming a bulge.

Two α-helices composed of residues Asn220 to Asn229 (α1) and Asn253 to Leu264 (α2) pack against one side of the β-sheet. In RBD2, α1 is one turn shorter than in the U1A RBD1 (Nagai et al., 1990) but has the same length as α1 in hnRNP A1 (Garret et al., 1994), hnRNP C (Wittekind et al., 1992), and RBD2 of Sxl protein (Lee et al., 1994). The relative orientation of the two α-helices and their relationship to the β-sheet were determined by a network of helix−helix and helix−sheet NOEs. The helix−sheet orientation is in part defined by a cluster of NOEs involving aromatic side chains of Phe228, Phe231, and Phe251 with the methyl group of Ala260, which results in the very large upfield shifts of the Ala260 H_α proton (2.7 ppm) and of its methyl protons (0.65 ppm) and also defines part of the hydrophobic core of this protein (including Leu210, Leu212, Leu215, Leu224, Leu227, Phe228, Phe231, Val237, Ala247, Val249, Phe251, Ala257, Ala260, Leu264, Phe267). Those hydrophobic residues that are conserved in the RBD family and are thought to be important determinants of the structure (Birney et al., 1993) correspond to Leu215, Leu224, Phe228, Val249, Phe251, and Ala260 in this protein. All are found in this hydrophobic core region.

(i) Helix Capping Box. The N-terminal capping box for an α-helix has been found to be an important structural motif

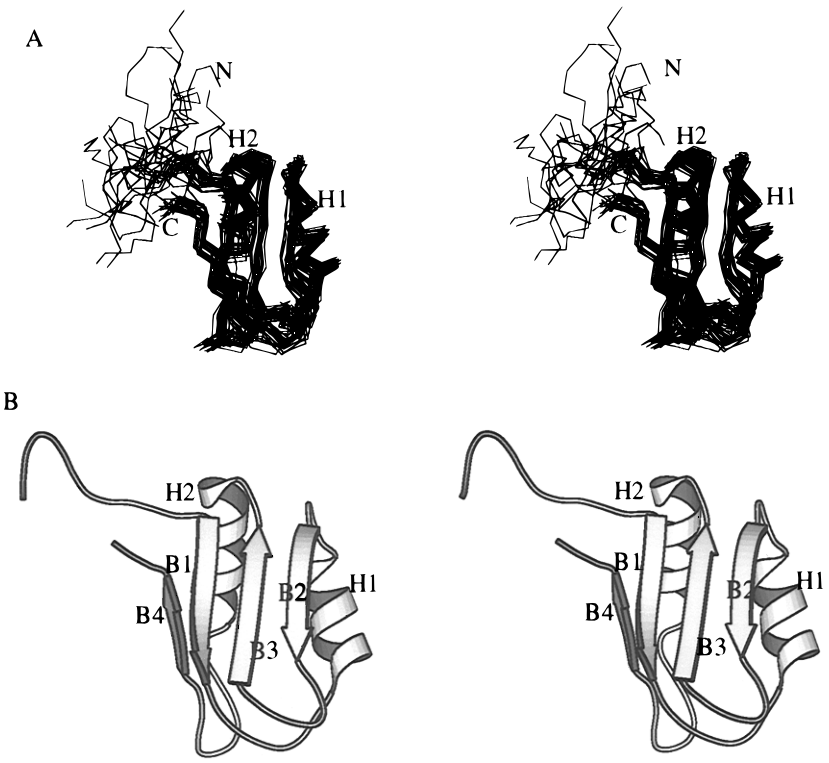

FIGURE 4: (A) Stereo diagram of the C_α trace of a family of 20 structures superimposed to backbone C_α atoms from α-helices and β-strands. (B) Ribbon diagram of RBD2 of the human U1A protein shown in stereo. The secondary structure elements are labeled and shown as ribbons. All other regions are shown as tubes. The figure was created with MOLSCRIPT software (Kraulis, 1991).

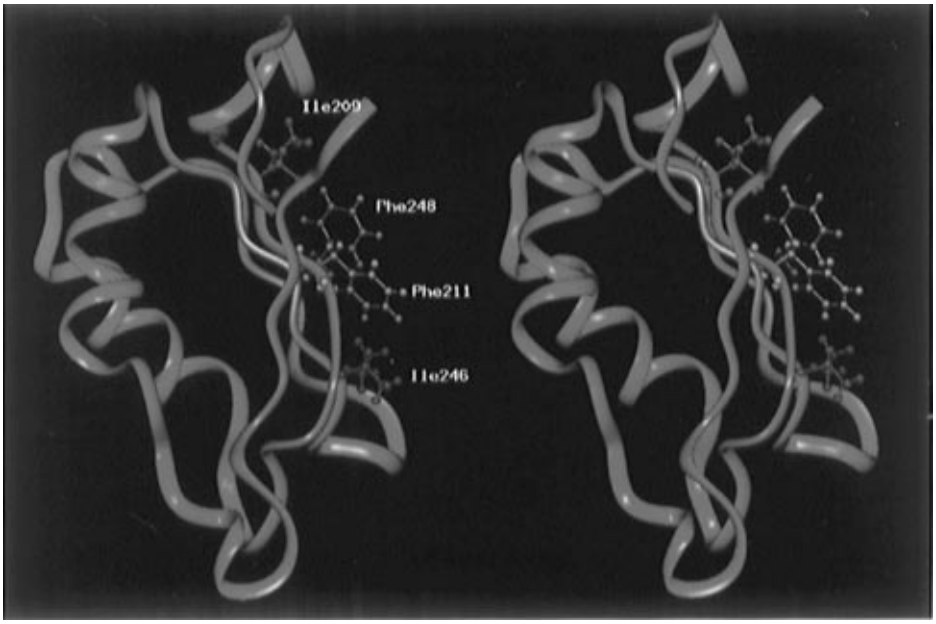

FIGURE 5: Stereo ribbon diagram of RBD2. The side-chain positions of Ile209, Phe211, Ileu246, and Phe248 are shown in ball-and-stick representation. The figure was generated with INSIGHT II (Biosym Technologies Inc.).

in stabilizing and initiating α-helices (Harper & Rose, 1993). An N-capping box has been proposed for the α1 helix of hnRNP-A1 (Garrett et al., 1994) and of hnRNP C (Wittekind et al., 1992). In the tertiary structure of Sxl RBD2, an N-capping box has been observed for the α1 helix (Lee et al., 1994). N-Capping boxes are present in both α-helices in RBD2. An unusual feature of the α1 N-cap arises from Met223 in the N3 position. Since the side chain of N3 (Met223) is not capable of making the normal reciprocal hydrogen bond with the backbone amide proton of the N-cap (Asn220), the amide proton of Asn220 was not protected from exchange, as it would be in a normal N-cap structure. For α2, Asn253 appears to act as an N-cap, based on the following evidence: (1) slow amide hydrogen exchange of Asn253 and Gln256 (N3) and relatively faster exchange for Glu254 and Val255; (2) NOEs from Asn253 NH to Gln256 H_β, Asn253 side-chain NH_2 to Gln256 H_β, Glu254 NH to Asn253 H_α and H_β, and Glu254 NH to Gln256 H_β; (3) 1.2 ppm upfield shift of Asn253 C_α and 1.3 ppm downfield shift of C_β characteristic of N-cap residues

(Gronenborn & Clore, 1994). On the basis of these data, three additional hydrogen bond constraints were included in the final structure calculation.

(ii) Structure of Loop 5. Loop 5 connecting the α2 helix and β4 contains 12 residues. Many NOEs between loop 5 residues are observed, which are consistent with formation of an antiparallel structure resembling a β-sheet. The carbonyl oxygen of Phe267 forms a H-bond with the amide proton of Met275, and the amide proton of Thr270 is H-bonded to the carbonyl oxygen of Asn273. Amino acids Thr270, Gln271, Asn272, and Asn273 form a turn. On the basis of the consensus Chemical Shift Index (CSI) (Lu & Hall, 1995), this region is random coil, and backbone dihedral angles also deviate from values typical of β-strands, implying that the structure suggested from NOE data may represent a conformationally averaged structure, consistent with the dynamics data described in the following section.

Backbone Dynamics of RBD1 and RBD2. To further characterize the physical properties of these two structurally similar domains, ^{15}N relaxation measurements were performed to determine backbone dynamics of RBD1 and RBD2. The RBD1 construct used in this study contains the N-terminal 1−95 amino acids of the U1A protein. This protein has the same RNA binding specificity as the 1−101 construct, although the affinity is 30-fold weaker. Amide ^{15}N and ^{1}H assignments for U1A RBD1 from amino acids 1−117 were generously provided by G. Varani and colleagues. Most of the resonance assignments were readily identified in the 1−95 RBD1 construct, although the chemical shifts of several resonances near the end of the domain changed significantly (90−95). Their assignments were confirmed by 3D ^{15}N-edited TOCSY-HMQC experiments.

Both RBD1 and RBD2 have a tendency to aggregate at high protein concentration (above 1.0 mM). An increase in line width with increasing protein concentration was observed when the upfield-shifted methyl proton line width was compared, varying from 15 Hz at 0.2 mM to 20 Hz at 2.0 mM protein concentration. Although the proteins are predominantly monomeric at 0.1 mM concentration as shown by equilibrium sedimentation, as the protein concentration increases above 1.0 mM, the protein samples show some degree of polydispersity as measured by inelastic light scattering (data not shown). Initial dynamics measurements were carried out at 2.0 mM RBD2, and the overall tumbling time of about 11.0 ns was calculated from relaxation measurements (data not shown), which is significantly longer than expected for a molecule of this size (9880 Da). Therefore, dynamics experiments were carried out in 0.5 mM RBD2 and RBD1, where the methyl proton line width is about 15 Hz. No further decrease in line width was observed when the protein concentration was lowered to 0.1 mM, indicating that the dominant species is a monomer at 0.5 mM concentration. The calculated τ_m of 6.6 ± 0.4 ns is consistent with the protein molecular weight, but we cannot rule out a contribution from larger species that increases the overall tumbling time.

(i) T_1, T_2, and NOE Values of RBD1 and RBD2. Intensities of resonances as a function of delay time were fit to a single exponential two parameter decay function. T_1, T_2, and NOE values as a function of residue number for RBD1 and RBD2 are shown in Figure 6. For RBD1, eight amide protons, Thr11, Ile12, Asn16, Ser29, Ile33, Gln39, Ser48, and Glu61, are overlapped and so are excluded from these analyses. The four proline residues (Pro4, Pro8, Pro76, and Pro81) are also absent. The average T_1 is 0.53 s with average standard deviation of 4%, and the average T_2 is 0.12 s with average standard deviation of 3%. Ten residues at the N-terminus and the C-terminus (Val3, Glu5, Thr6, Arg7, Asp90, Ser91, Asp92, Ile93, Ile94, Ala95) show significantly increased T_1 and T_2 values. The T_2 values for Ile14, Leu17, Leu26, Leu30, Tyr31, Ala32, Phe34, Ser35, Gln36, Phe37, Val45, Lys50, Arg52, Ala55, Phe56, Met72, Phe75, Phe77, and Asp79 are more than one standard deviation below the average T_2. The average NOE is 0.62. The NOE values for Val3, Glu5, Thr6, Arg7, Asp90, Ser91, Asp92, Ile93, Ile94, and Ala95 are very small, in agreement with their slower T_1 and T_2 relaxation. For RBD2, 78 amide NH resonances were observed; Leu201 is not observed, and neither are the seven prolines (Pro197, Pro200, Pro205, and Pro206 in the N-terminal tail; Pro216, Pro232, and Pro241). The average T_1 for residues 208−282 is about 0.5 s with an average standard deviation of 3%, and the average T_2 is 0.13 s with an average standard deviation of 2.6%. The N-terminal six residues (Ala196, Ala198, Ser202, Glu203, and Asn204) and the last residue at the C-terminus (Lys282) show significantly increased T_1 and T_2 values. The T_2 values for residues Gly266, Lys268, Ile269, Thr270, Gln271, and Met275 in loop 5 are also noticeably smaller than the average. The average NOE is 0.68; those residues which have NOE values more than one standard deviation below the mean (calculated for residues 208−281) are Thr219, Asn220, Leu227, Gln230, Val237, Leu239, Gly242, Lys268, Asn272, Asn273, Ala274, and Lys281.

(ii) Overall Correlation Time (τ_m). Overall correlation times (τ_m) were calculated as described (Materials and Methods). In order to use the simple spectral density function (2) to calculate τ_m, the assumption was made that the proteins tumble isotropically in solution. The relative values of the inertial tensor derived from coordinates of the crystal structure of RBD1 are 0.67:1:1, validating the assumption of isotropic tumbling. The calculated τ_m was 6.6 ± 0.4 ns for RBD2 (9880 Da) and 7.6 ± 0.5 ns for RBD1 (10924 Da), approximately correct for globular proteins of these sizes.

(iii) General Order Parameters. The model-free approach proposed by Lipari and Szabo (1982a,b) to describe the spectral density and an extension of this formalism to include two time scales for the motion of the N−H vector (Clore et al., 1990b) were then used to analyze the data (Dellwo & Wand, 1989) to obtain the overall correlation time (τ_m), effective correlation time (τ_e), order parameter (S^2), and, in some cases, the exchange contribution to T_2 (R_{ex}). The analysis of the relaxation data in terms of the possible models of motion was essentially performed as described by Farrow et al. (1994) (see Materials and Methods). The model with fewest parameters was selected if more than one model was found to fit the relaxation data. The two time-scale model was considered only if none of the other models fits the data.

The average value of the S^2 order parameter is 0.74 for both RBDs. For RBD1, four residues in the N-terminal region (Val3, Glu5, Thr6, and Arg7) that can be observed in 2D ^{15}N−^{1}H HSQC spectra have very small order parameters, suggesting this region is largely disordered on the picosecond to nanosecond time scale. The C-terminal tail region of RBD1 is six residues longer than in RBD2; its order parameters decrease from 0.53 for Asp89 to 0.17 for

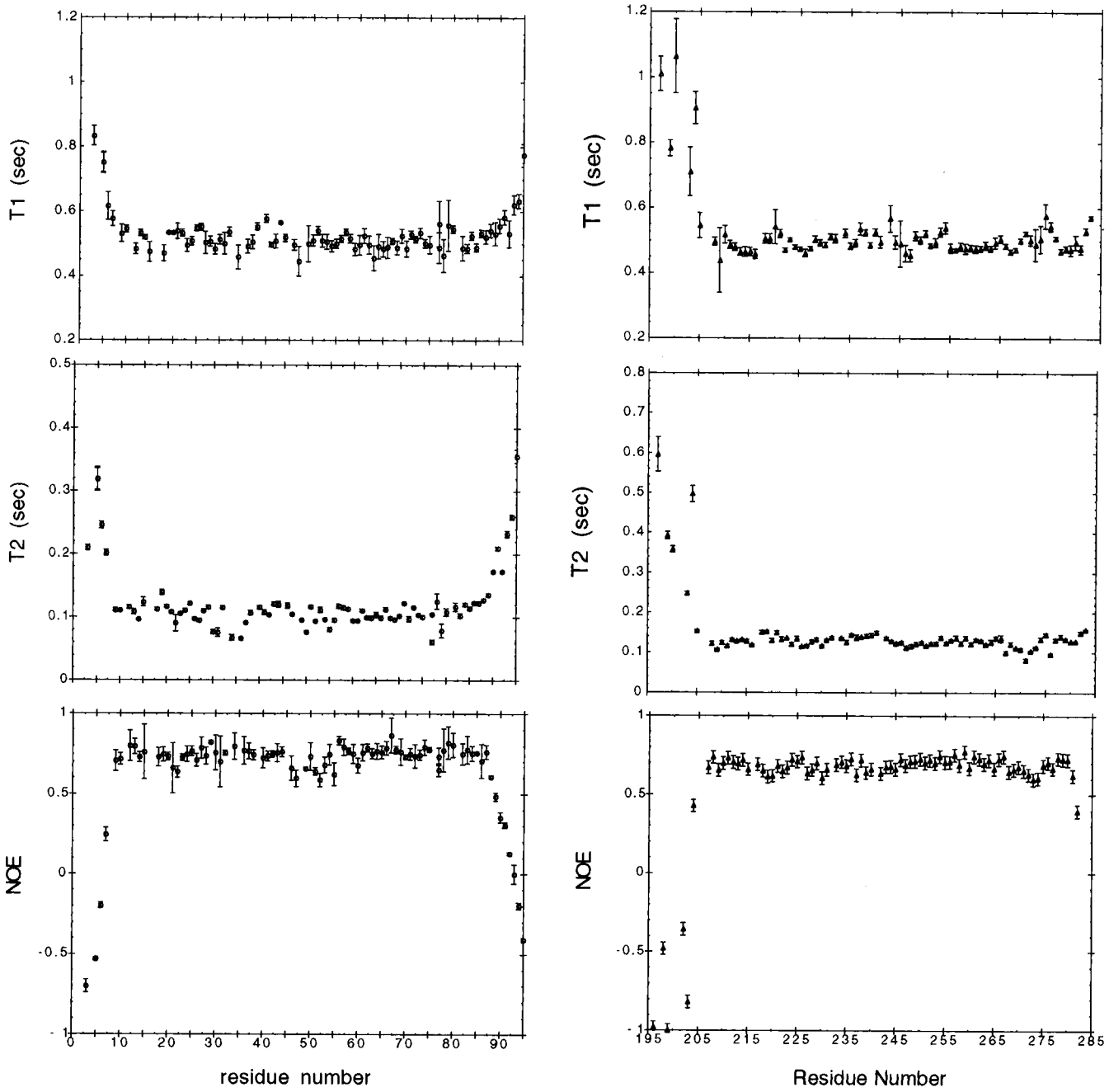

FIGURE 6: Relaxation data for RBD1 (left) and RBD2 (right) as a function of residue number. ^{15}N T_1, T_2, and NOE values were measured at 500 MHz.

Ala95. Forty-three of 82 residues observed for RBD1 can be fit with the simple spectral density function with either τ_e fixed to zero or as a fitting parameter ranging from 20 to 100 ps. Three exceptions are Arg47, Leu49, and Lys60, where the simple spectral density function gives the best fit to the relaxation data, but τ_e is between 200 and 400 ps. Both Arg47 and Leu49 are located in loop 3. In contrast, most residues (59 of 78) observed for RBD2 require only the simple spectral density function (2) with τ_e fixed to zero or τ_e as a fitting parameter ranging from 10 to 100 ps. Order parameters for the N-terminal six residues are very small, increasing from 0.1 for Ala196 to 0.66 for Asn204 (see Figure 7). The following Pro205 and Pro206 are absent in these experiments, while Asn207 has order parameters close to the average (0.81). These data suggest that residues 196−204 are largely disordered on the picosecond to nanosecond time scale, which is consistent with the NMR tertiary structure of this domain. The C-terminal end of RBD2 is very short compared to other RBDs, containing only two residues after β4. The last residue, Lys282, is also more flexible and has an order parameter of 0.65. Glu217, Glu218, and Asn220 in loop 1, Lys235 (the β-bulge) and Arg238 in β2, Val240 in loop 3, and Lys276 and Ile277 at the beginning of β4 also have order parameters noticeably smaller than average.

(iv) R_{ex} and the Two Time-Scale Spectral Density Function. Nineteen residues of RBD1 require an exchange term for T_2 relaxation. R_{ex} as a function of residue number is shown in Figure 7. The location of these residues in the 3D structure of RBD1 is shown in Figure 8; most of these residues are located in α1, loop 3, and loop 5. Twenty residues (Val3, Glu5, Thr6, Arg7 at the N-terminus; Asn15, Asn18, Glu19, Lys27, Lys28, Leu44, Ser71, Ile84, Gln85, Ala87, residues 89−95) require a two time-scale spectral density function

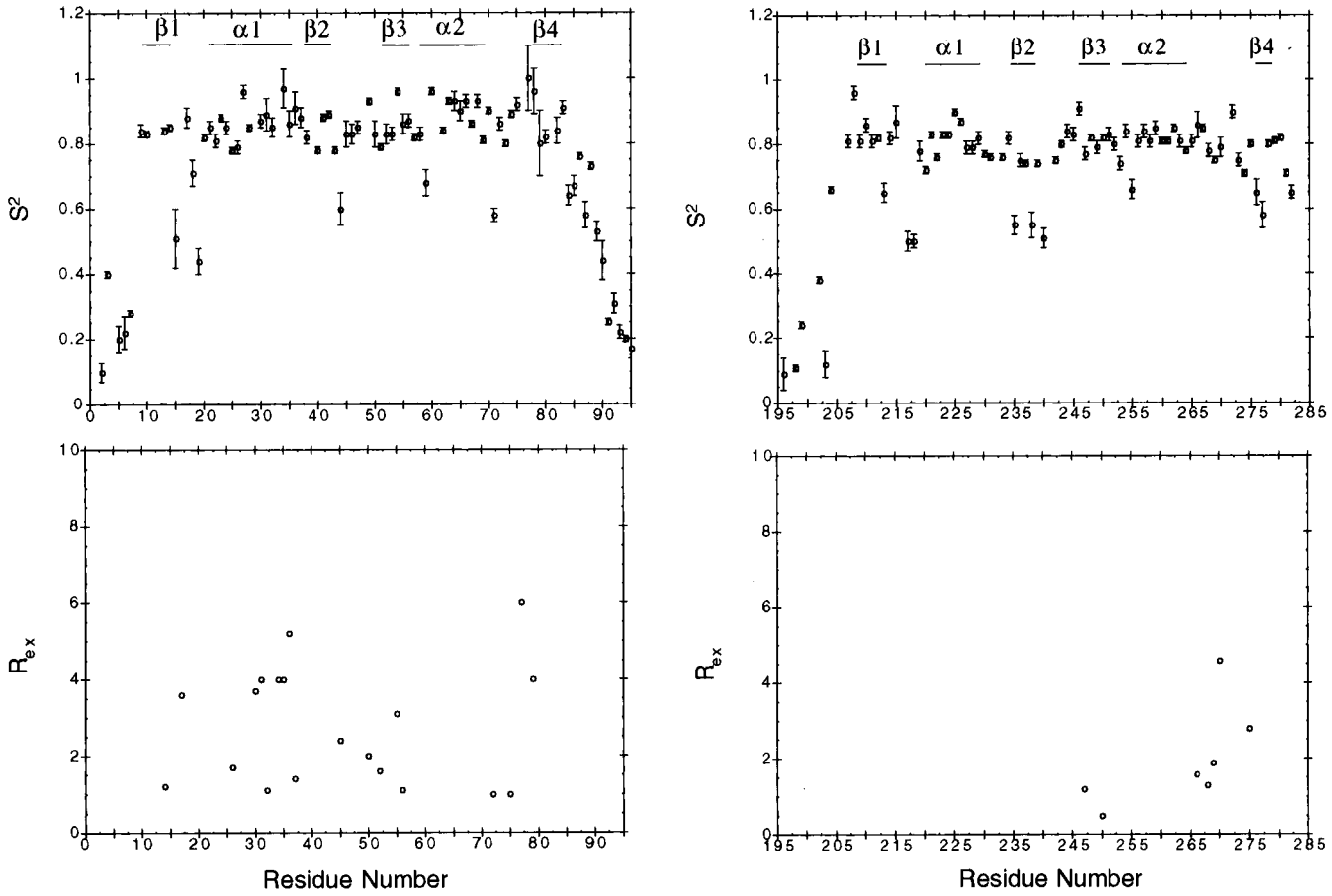

FIGURE 7: Plot of S^2 and R_{ex} as a function of residue number for RBD1 (left) and RBD2 (right) calculated from the data. Relaxation data were fit to five different models as described (Materials and Methods).

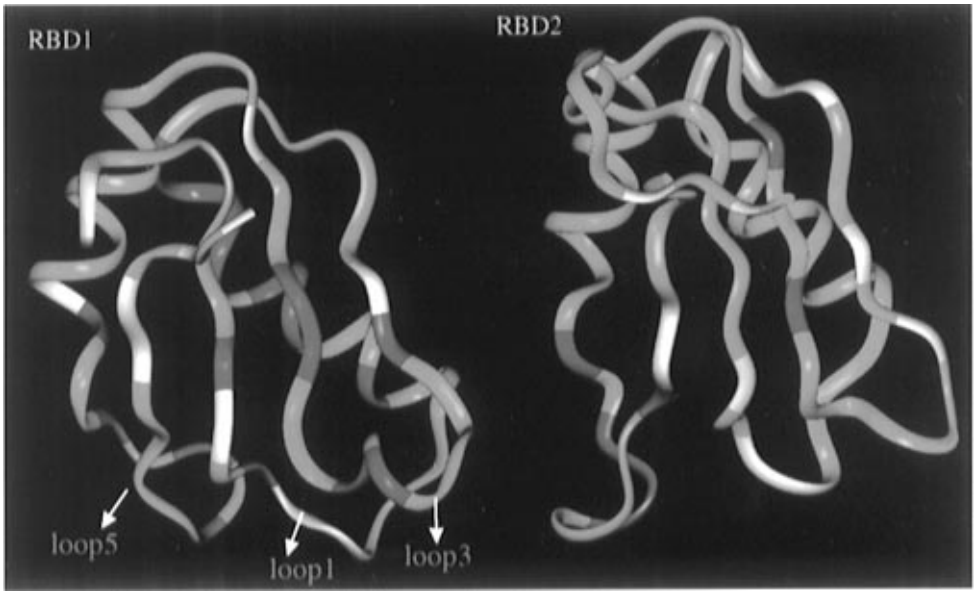

FIGURE 8: Tertiary structures of RBD1 and RBD2 of the human U1A protein. Residues requiring an additional exchange term (R_{ex}) are shown in pink, residues requiring a two time-scale spectral density function are shown in yellow, and those that require only the simple spectral density function are shown in cyan. The structure of RBD1 is from a reconstruction of the crystallographic data (courtesy of Dr. C. van Gelder; van Gelder et al., 1994): the C-terminal five residues (91−95 amino acids) are not shown in this figure, but residues in this region also require a two time-scale spectral density function. The structure of RBD2 is one representative of a family of structures determined from NMR spectroscopy. The N-terminal region of RBD2 shown as extended is largely disordered. Also, for simplicity, those residues for which there are no dynamics data are also colored cyan (see text).

(including τ_s). Most are located in the N-terminal and C-terminal regions; the entire C-terminal region (amino acids 89−95) requires two correlation times to fit the decay. Seven residues of RBD2 (Ala247, Glu250, Gly266, Lys268, Ile269, Thr270, Met275) require an exchange term (R_{ex}) to fit the T_2 data (see Figure 7). Two of these residues (Ala247, Glu250) are located in the β3 strand, while the others are located in the long loop 5 (residues 264−276), where R_{ex}

ranges from 1.3 to 4.6 s^{-1}. Twelve residues of RBD2 (Ala196, Ala198, Glu203, Thr213, Glu217, Glu218, Lys235, Arg238, Val240, Gln256, Lys276, and Ile277) require a two time-scale spectral density function (4) including a correlation time for slow motion (τ_s). These residues are located in the N-terminal region, loop 1, β2 including the β-bulge at Lys235, and the beginning of β4.

The dynamics data identify loop 5 of both proteins as exhibiting some conformational exchange motion. The residues in loop 5 in both RBD1 and RBD2 are particularly sensitive to protein concentration, as most clearly noted in the T_2 relaxation times. For example, in RBD1, the amide proton resonances of F77 and D79 were very weak in ^{15}N/^{1}H HSQC spectra at 2.0 mM and became much stronger at 0.5 mM concentration. When relaxation parameters were analyzed using the Lipari−Szabo model-free formalism, several residues in this region of RBD1 (Phe75, Phe77, Asp79) were found to require an R_{ex} term to account for the measured T_2 relaxation time. It is likely that this whole region undergoes some conformational exchange, although not every residue in this region requires an R_{ex} term. Other data support this conclusion: the ^{19}F Tyr78 resonance of RBD1 is very broad at 2 mM protein concentration but sharpens at 0.5 mM protein concentration or in 0.2 M guanidine hydrochloride (Kranz et al., 1996). These data, combined with the concentration dependence of the resonance intensities, suggest that this region may be involved in nonspecific protein−protein association.

DISCUSSION

RBD2 of the human U1A protein was identified as a member of this family of domains by its two regions of conserved amino acid sequences (Sillikens et al., 1987). However, those sequences, RNP-1 and RNP-2, are rather unusual. The U1A RBD2 was recognized as an odd example of these RBDs when compared with other known sequences (Birney et al., 1993), although its very high homology to RBD2 of the human U2 B″ snRNP protein implied a conserved function of these two domains, perhaps related to pre-mRNA splicing. However, experiments with the isolated RBD2 domain (Lu & Hall, 1995) and experiments with the entire U1A protein in which the RNA binding of RBD1 was blocked by mutation (Scherly et al., 1989) failed to find any evidence of association of an RNA with RBD2. In these studies, we have compared the sequence, structure, and backbone dynamics of RBD2 to other RBDs, particularly U1A RBD1, to look for possible explanations for their functional differences.

Comparison of Other RBD Structures with U1A RBD2. At the level of sequence and structural features, RBD2 can be readily compared with the available structures of RBDs that have been shown to bind RNA. Like other RBDs, this domain adopts a $\beta\alpha\beta-\beta\alpha\beta$ secondary structure and an α/β sandwich tertiary fold. The individual α-helices of RBD2 have lengths equal to those in hnRNP C (Wittekind et al., 1992) and in Sxl RBD2 (Lee et al., 1994); α1 in the U1A RBD1 is one turn longer at the N-terminus (Nagai et al., 1990). Taking these lengths into account, the α1 helix of RBD2 contains two absolutely conserved residues, Leu224 and Phe228, which are in analogous positions in other RBDs. These two residues are part of the hydrophobic core of the protein. An N-terminal capping box (Harper & Rose, 1993) has been identified for α1 of Sxl RBD2 (Lee et al., 1994) and for α2 of the hnRNP C protein (Wittekind et al., 1992). Both capping boxes are present in RBD2. The relative orientation of α1 and α2 in RBD2 is similar to that of U1A RBD1, for a superposition (using Biosym Insight II superpose) by aligning 42 residues in the two helices, gives a C_α RMSD of 1.0 Å (data not shown), taking into account the greater length of RBD1 α1. In this regard, these two U1A domains differ from the Sxl and hnRNP C domains, for in those structures, the two helices are nearly perpendicular to each other.

Comparing other RBD structures with that of RBD2 could possibly show obvious features of RBD2 that might interfere with its ability to bind RNA. Although many interactions between RNA and protein occur through amino acid side chains, many also come from backbone carbonyl oxygen atoms or amide nitrogen atoms [e.g., Oubridge et al. (1994)]. Thus a comparison of the general topological features between domains may illustrate how the protein backbone can differ among RBDs in general. In particular, regions of RBDs that are known to participate in RNA binding are the β-sheet, loop 3, and the carboxyl-terminal tail of the protein.

The topology of the β-sheet is of interest, since this is presumably the RNA binding surface in all RBDs. The β-sheets of Sxl RBD2, hnRNP C, and U1A RBD1 all show an obvious right-handed twist of varying magnitude. In addition, there is a β-bulge near the N-terminus of β2 in Sxl RBD2 and U1A RBD1 and in both β2 and β4 of hnRNP C. As Lee et al. (1994) suggest, the location of the defect in the sheet surface introduced by the bulge may contribute to recognition of specific RNA targets. U1A RBD2 also shares these structural features, with a right-handed twist and a β-bulge at the beginning of β2; in addition, β4 is tilted away from the surface of the β-sheet, a feature it shares with Sxl RBD2. However, in RBD2, the twist of the β-sheet is very shallow, for β2, β3, and β1 are nearly planar. It is possible that this altered β-sheet surface makes a poor binding site for RNA.

The display of amino acids on the surface of the RBD2 β-sheet has some unusual features. First, the relative orientation of the aromatic rings of Phe211 and Phe248 may be influenced by interactions with other nonpolar side chains. In RBD2, the side chains of Ile209, Phe248, Phe211, and Ile246 form a hydrophobic patch on the surface (Figure 5). Although the surface of the β-sheet is nonpolar in these RBDs, RBD2 is one of only two RBDs with a hydrophobic amino acid at the position equivalent to Ile209; the interaction with one of the conserved aromatic residues may have functional consequences, based on the results of experiments with RBD1 (Zeng & Hall, 1997). In the structure of U1A RBD1 (Nagai et al., 1990; Avis et al., 1996), Tyr13 and Phe56 appear to stack on each other; in Sxl RBD2, Tyr17 and Phe59 also appear to stack (Lee et al., 1994). Both U1A RBD1 and Sxl RBD2 bind RNA; if the orientations of the conserved aromatic residues are functionally important, then U1A RBD2 is clearly different. However, the interaction with other amino acid can modulate the presentation of these residues, and these interactions must be considered in descriptions of RNA recognition.

RBD2 is truncated two residues past the end of β4. In other RBDs, the C-terminal tail is much longer, and experiments with hnRNP C (Gorlach et al., 1994) and hnRNP

A1 (Burd & Dreyfuss, 1994b), as well as with U1A RBD1 (Scherly et al., 1989; Lutz-Freyermuth et al., 1990; Boelens et al., 1991; Jessen et al., 1991; Hall, 1994; Zeng & Hall, 1997), show that these sequences can participate in binding the RNA target, which is probably a general feature of these domains. The absence of a longer carboxyl-terminal tail in RBD2 may also remove possible contacts that induce the RNA to adhere to the protein surface. Simply appending the tail of RBD1 to RBD2 does not rescue its RNA binding capability, however (Lu & Hall, 1995).

Finally, loop 3 of U1A RBD1 has been shown to provide critical contacts for association with RNA, thus identifying this element of RBDs as a potential source of interactions. The loops between structural elements are the most variable parts of RBDs, and loop 3 is no exception. For example, a flexible loop 3 is not necessary for RNA binding, as shown by hnRNP C, which instead has a tight β-turn (Wittekind et al., 1992). Sxl RBD2 has a much longer loop 3 containing four positively charged residues, which shows few NOEs and appears to be quite flexible (Lee et al., 1994). U1A RBD2 has a seven amino acid loop 3, containing a ProGly sequence, although it also contains an Arg and His which are common sources of RNA/protein contact. If potential for RNA binding were based on the sequence complexity and flexibility of loop3, hnRNP C would be predicted to be unlikely to bind RNA. Since the isolated RBD of hnRNP C can bind $r(U)_8$ with micromolar affinity (Gorlach et al., 1992), while U1A RBD2 cannot bind any RNA, this hypothesis is clearly too simple.

Backbone Dynamics. One possible use of amide ^{15}N backbone dynamics is to identify or predict those regions of a protein that have the potential flexibility to respond to a functional event, such as binding of a ligand, a conformational change during macromolecular assembly, or an enzymatic activity. Experiments with ubiquitin (Schneider et al., 1992), interleukin-1β (Clore et al., 1990a), and glucose permease IIA domain (Stone et al., 1992) have shown that there is no consistent correspondence between the order parameter (S^2) and secondary structural elements, and therefore this most simple correlation lacks predictive accuracy. The comprehensive discussion of Farrow et al. (1994) concludes that there are no simple rules for the interpretation of the relaxation parameters such as exchange rates (R_{ex}) or correlation times τ_e, τ_s, or τ_f, nor for their functional significance. However, from their experiments with the PLCγ1C SH2 domain, comparing the dynamics of the free and complexed domain, they were able to conclude that the presence of a R_{ex} term to describe T_2 relaxation was consistent with motion on a slow time scale typical of a conformational change (Farrow et al., 1994).

With the U1A RBD1 and RBD2, one has the opportunity to compare the backbone dynamics of two proteins with identical secondary structures ($\beta\alpha\beta-\beta\alpha\beta$) and global topologies (α/β sandwich), which have different functions. Clearly the existing dynamics data indicate that proteins have as yet unpredictable dynamic properties; with RBD1 and RBD2, it is now possible to determine how different those properties can be in the context of structurally homologous proteins.

Similar to the analysis of Farrow et al. (1994), the average order parameters of the discrete structural elements of the two domains are given in Table 3. As other studies have previously noted, there are no apparent patterns to the relation of S^2 to structural elements. For example, loop 5 of both proteins has order parameters similar to those of secondary structure regions. Among the secondary structure regions, $\alpha1$, $\beta3$, and $\alpha2$ of both domains have order parameters higher than average, while $\beta1$, $\beta2$, and $\beta4$ of both proteins have lower order parameters. The lower than average order parameters for strands on the edge of the β-sheet ($\beta2$ and $\beta4$) and their relatively faster amide proton exchange rates (data not shown) suggest possible correlations between order parameters and hydrogen exchange. The correlations are similar to those observed in comparison of amide exchange rates and order parameters in calbindin D_{9k} (Kördel et al., 1992).

Table 3: Order Parameters of RBD Structural Elements[a]

structure		S^2 for RBD2		S^2 for RBD1
$\beta1$	[5]	0.79 (0.65/0.86)	[3]	0.73 (0.52/0.85)
loop 1	[5]	0.69 (0.50/0.87)	[5]	0.74 (0.44/0.88)
$\alpha1$	[10]	0.81 (0.72/0.90)	[14]	0.87 (0.78/0.96)
loop 2	[3]	0.76 (0.76/0.77)	[1]	0.82
$\beta2$	[5]	0.68 (0.55/0.82)	[5]	0.79 (0.60/0.89)
loop 3	[6]	0.77 (0.76/0.85)	[8]	0.84 (0.79/0.93)
$\beta3$	[6]	0.82 (0.77/0.91)	[6]	0.84 (0.69/0.96)
loop 4	[1]	0.80	[1]	0.96
$\alpha2$	[12]	0.80 (0.72/0.85)	[11]	0.86 (0.58/0.93)
loop 5	[10]	0.80 (0.71/0.90)	[6]	0.90 (0.80/1.0)
$\beta4$	[4]	0.71 (0.58/0.81)	[5]	0.77 (0.66/0.91)

[a] Numbers in brackets are the number of residues in the element. Numbers in parentheses are the minimum and maximum S^2.

Of the other two parameters describing backbone dynamics, the exchange term R_{ex} added to account for T_2 relaxation and the slower motion (τ_s) associated with the extended form of the model-independent spectral density (Clore et al., 1990b), the exchange term reports on slower motion (microsecond to millisecond) that can most reasonably be ascribed to conformational exchange. It should be noted that there are examples in these RBD data where the order parameter is high (e.g., Asp79 in RBD1 with $S^2 = 0.80 \pm 0.1$), yet a substantial exchange term is required to fit the T_2 relaxation data ($R_{ex} = 4 \pm 2\ s^{-1}$); such data are puzzling. In addition, as Mandel et al. (1997) have concluded, values of $R_{ex} < 0.5\ s^{-1}$ are not robustly identified by the protocol and should be interpreted with caution. Therefore, in this comparison of RBD1 and RBD2, we are particularly looking for stretches of residues that require R_{ex} terms to fit their T_2 data. It is unlikely that a lone residue that requires a T_2 R_{ex} term is indicative of conformational exchange; such an occurrence cannot be ascribed to a physical model and may simply reflect the inadequacies of the fitting.

Three residues at analogous positions in both RBDs show similar dynamic properties: two residues in loop 1 require a two timescale spectral density function, and the alanines at the fourth position of the RNP-1 octamer $\beta3$ require an R_{ex} term. Residues in loop 5 of each protein also show similar backbone dynamics. This long loop appears to contribute to intermolecular protein:protein association, especially at high protein concentrations. There are several hydrophobic residues in this 12-residue loop, particularly in RBD1. In the crystal structure of RBD1, hydrophobic residues in this region (Phe75, Phe77, Tyr78) pack against Leu26, Leu30, and Phe34 in $\alpha1$ and form part of the hydrophobic core of the protein. In solution, our dynamics data suggest that this region is undergoing some conformational exchange. It is possible that the hydrophobic residues in loop 5 self-associate in intermolecular contacts or interact

with other regions of the protein such as the nonpolar β-sheet surface. An increase in protein concentration may promote intermolecular association of this loop; the line width of several residues in this region becomes much broader with increasing protein concentration (data not shown). When RNA is bound to RBD1, this region becomes stabilized and normal line width is observed (data not shown).

Residues in the two RBDs that show different dynamic behavior include two regions that in RBD1 are critical to interaction with RNA and undergo conformational changes in order to make contact with specific nucleotides. These regions are the C-terminal tail and loop 3 of RBD1 (Oubridge et al., 1994), which in these dynamics experiments require the two time-scale spectral density function or T_2 exchange terms, respectively, to describe their dynamic behavior. In the crystal structure of RBD1 (residues 1−95), the C-terminal 91−95 amino acids were mobile, making them invisible in the electron density map (Nagai et al., 1990). This C-terminal tail becomes stabilized and forms an α-helix centered around Ala95 upon binding to RNA, as shown in the structure of the RBD1(1−102)−RNA complex (Oubridge et al., 1994; Allain et al., 1996). In the 1−95 RBD1 protein, this region is largely disordered on a picosecond to nanosecond time scale, strongly suggesting that this region does not adopt a stable helical conformation. Loop 3 of RBD1 contains seven residues; three residues require R_{ex} terms in their T_2 fits. These dynamic data suggest that it exhibits some conformational flexibility in solution, which may be critical for its ability to interact with RNA. In contrast, RBD2 lacks a C-terminal tail, and the dynamics of its loop 3 can be described by the simple spectral density function without an exchange term (R_{ex}).

We note here that many of the backbone amides in α1 of RBD1 also require additional exchange terms to describe T_2 relaxation. However, the amide protons in this region are highly protected, similar to those amide protons in α1 of RBD2 (data not shown). The lack of correlation between the motions measured by R_{ex} and motions probed by hydrogen exchange suggest that the mechanism of chemical exchange of α1 may involve simple rigid body displacement of the helix. The functional significance of this motion is not clear, but our speculations include propagation of motion through the tertiary structure, linking α1 motion to movements of β3 and loop3.

This comparative study of backbone dynamics suggests some possible correlation between backbone motion and RNA binding. Overall, fewer RBD2 amides require exchange terms (R_{ex}) or application of the two timescale extended spectral density to describe their relaxation. If the relatively slow motions measured in these dynamics experiments can be related to larger amplitude motions, then these data suggest how these two proteins may differ in their RNA binding properties, particularly if there is a correlation between the T_2 exchange broadening (R_{ex}) and the potential for conformational flexibility [e.g., Stone et al. (1992), Farrow et al. (1994), Mandel et al. (1996), and Kay et al. (1996)].

RNA Binding and the β-Sheet. Finally, the dynamics and structural data for RBD2 can be combined in a discussion of the two solvent-exposed aromatic residues on the surface of the β-sheet. These two residues, usually phenylalanine or tyrosine, have been shown to be in direct contact with RNA through UV cross-linking studies of hnRNP A1 (Merrill et al., 1988) and U1A RBD1 (Stump & Hall, 1995) and crystallographic studies of the U1A RBD1−RNA complex (Oubridge et al., 1994). In RBD2, these two aromatic residues correspond to Phe211 and Phe248, which have side chains that are solvent accessible on the surface of the β-sheet. The stacking of these aromatic amino acid side chains with the aromatic nucleic acid bases will contribute to RNA binding through the interaction of the conjugated π systems. However, this stacking can neither provide sequence-specific recognition nor does it make a sufficient energetic contribution to allow detectable RNA binding for RBD2 in the absence of other stabilizing interactions. In many RBDs, these other interactions could come from a C-terminal or N-terminal arm that reaches over the β-sheet to anchor an RNA. In addition, a long and flexible loop 3 could provide additional energetically favorable side-chain and backbone contacts to RNA. The coupling among these regions is critical for RNA recognition. It is important to study this interplay between sequence, structure, and dynamics to understand the mechanism of RNA recognition.

In summary, the overall topology of the U1A RBD2 is very similar to those of three other RBDs which are known to bind RNA. However, the β-sheet surface and relative orientation of two aromatic rings appear to be quite different. There are also differences in the backbone dynamics of U1A RBD1 and RBD2, but without other examples for comparison, we can only speculate that these different motions are functionally significant. Since an intricate network of amino acid interactions within the protein as well as with their RNA substrates define the mechanism of recognition in these complexes, it is hardly surprising that the identification of a binding site cannot be reduced to an examination of protein topology. Now with the confidence that the structure of RBD2 is typical of this family of RNA binding proteins, there is the opportunity to rescue its RNA binding function through substitution of each structural element or specific amino acid substitutions that can directly measure the effect of amino acid side chains on RNA binding.

ACKNOWLEDGMENT

DISTGEOM is available by anonymous FTP at ftp://dasher.wustl.edu or on the Web at http://dasher.wustl.edu/tinker/; we thank Professor Jay Ponder for discussions of its implementation and advantages, as well as the assistance of Dr. Michael Hodsdon and Professor David Cistola in utilizing DISTGEOM. We thank Dr. Changguo Tang for help in implementation of pulse sequences and calculation of the RBD1 inertial tensor and Professor Lewis Kay (University of Toronto) for providing the pulse sequences for T_1, T_2, and NOE measurements and the relaxation data analysis software. The clone of the U1A protein was obtained from Professor Walther van Venrooij (Nijmegen). Professor Gabriele Varani (MRC) kindly provided assignments of RBD1.

SUPPORTING INFORMATION AVAILABLE

A figure of longitudinal and transverse relaxation curves for representative residues and two tables containing current ^{1}H, ^{13}C, and ^{15}N side-chain resonance assignments for RBD2 (12 pages). Ordering information is given on any current masthead page.

REFERENCES

Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D., & Varani, G. (1996) *Nature 380*, 646−650.

Avis, J. M., Allain, F. H.-T., Howe, P. W. A., Varani, G., Nagai, K., & Neuhaus, D. (1996) *J. Mol. Biol. 257*, 398−411.

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

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990) *J. Magn. Reson. 88*, 425−431.

Birney, E., Kumar, S., & Krainer, A. R. (1993) *Nucleic Acids Res. 21*, 5803−5816.

Boelens, W., Scherly, D., Kambach, C., Kolen, K., Mattaj, I. W., & van Venrooij, W. J. (1991) *Nucleic Acids Res. 19*, 4611−4618.

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

Burd, C. G., & Dreyfuss, G. (1994a) *Science 265*, 615−621.

Burd, C. G., & Dreyfuss, G. (1994b) *EMBO J. 13*, 1197−1204.

Clore, G. M., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990a) *Biochemistry 29*, 7387−7401.

Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., & Gronenborn, A. M. (1990b) *J. Am. Chem. Soc. 112*, 4989−4991.

Dellwo, M. J., & Wand, J. A. (1989) *J. Am. Chem. Soc. 111*, 4511−4518.

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

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

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.

Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M., & Gronenborn, A. M. (1994) *Biochemistry 33*, 2852−2858.

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

Gorlach, M., Burd, C. G., & Dreyfuss, G. (1994) *J. Biol. Chem. 269*, 23074−23078.

Gronenborn, A. M., & Clore, G. M. (1994) *J. Biomol. NMR 4*, 455−458.

Hall, K. B. (1994) *Biochemistry 33*, 10076−10088.

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

Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605−7609.

Hodsdon, M., & Cistola, D. (1997) *Biochemistry 36*, 1450−1460.

Hodsdon, M., Ponder, J. W., & Cistola, D. (1996) *J. Mol. Biol. 264*, 585−602.

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

Howe, P. W., Nagai, K., Neuhaus, D., & Varani, G. (1994) *EMBO J. 13*, 3873−3881.

Ikura, M., Kay, L. E., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 86*, 204−209.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.

Kay, L. E., Muhandiram, D. R., Farrow, N. A., Aubin, Y., & Forman-Kay, J. D. (1996) *Biochemistry 35*, 361−368.

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

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

Kördel, J., Skelton, N. J., Akke, M., Palmer, A. G., & Chazin, W. J. (1992) *Biochemistry 31*, 4856−4866.

Kranz, J. K., Lu, J., & Hall, K. B. (1996) *Protein Sci. 5*, 1567−1583.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kuntz, I. D., Thomason, J. F., & Oshiro, C. M. (1989) *Methods Enzymol. 177*, 159−204.

Laskowski, R., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 83−291.

Lee, A. L., Kanaar, R., Rio, D. C., & Wemmer, D. E. (1994) *Biochemistry 33*, 13775−13786.

Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546−4559.

Lipari, G., & Szabo, A. (1982b) *J. Am. Chem. Soc. 104*, 4559−4570.

Lu, J., & Hall, K. B. (1995) *J. Mol. Biol. 247*, 739−752.

Lutz-Freyermuth, C., Query, C. C., & Keene, J. D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6393−6397.

MacArthur, M. W., Laskowski, R. A., & Thornton, J. M. (1994) *Curr. Opin. Struct. Biol. 4*, 731−737.

Mandel, A. M., Akke, M., & Palmer, A. G. (1996) *Biochemistry 35*, 16009−16023.

Mattaj, I. W. (1989) *Cell 57*, 1−3.

Mattaj, I. W. (1993) *Cell 73*, 837−840.

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

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

Oshiro, C. M., Thomason, J., & Kuntz, I. D. (1991) *Biopolymers 31*, 1049−1064.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., & Nagai, K. (1994) *Nature 372*, 432−438.

Rullman, J. A. C. (1996) AQUA, Computer Program, Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.

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

Schneider, D. M., Dellwo, M. J., & Wand, A. J. (1992) *Biochemistry 31*, 3645−3652.

Sillikens, P. T. G., Habets, W. J., Beijer, R. P., & van Venrooij, W. J. (1987) *EMBO J. 6*, 3841−3848.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.

Stone, M. J., Fairbrother, W. J., Palmer, A. G., III, Reizer, J., Saier, M. H., & Wright, P. E. (1992) *Biochemistry 31*, 4394−4406.

Stump, W. T., & Hall, K. B. (1995) *RNA 1*, 55−63.

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

van Gelder, C. W., Leusen, F. J., Leunissen, J. A., & Noordik, J. H. (1994) *Proteins: Struct., Funct,. Genet. 18*, 174−185.

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

Zeng, Q., & Hall, K. B. (1997) *RNA 3*, 303−314.

BI9709811