The EMBO Journal Vol.18 No.16 pp.4523–4534, 1999

Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF^{65})

Takuhiro Ito, Yutaka Muto, Michael R.Green[1] and Shigeyuki Yokoyama[2]

Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [1]Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, MA 01605, USA

[2]Corresponding author
e-mail: yokoyama@y-sun.biochem.s.u-tokyo.ac.jp

The large subunit of the human U2 small nuclear ribonucleoprotein particle auxiliary factor (hU2AF^{65}) is an essential RNA-splicing factor required for the recognition of the polypyrimidine tract immediately upstream of the 3′ splice site. In the present study, we determined the solution structures of two hU2AF^{65} fragments, corresponding to the first and second RNA-binding domains (RBD1 and RBD2, respectively), by nuclear magnetic resonance spectroscopy. The tertiary structure of RBD2 is similar to that of typical RNA-binding domains with the β1–α1–β2–β3–α2–β4 topology. In contrast, the hU2AF^{65} RBD1 structure has unique features: (i) the α1 helix is elongated by one turn toward the C-terminus; (ii) the loop between α1 and β2 (the α1/β2 loop) is much longer and has a defined conformation; (iii) the β2 strand is ^{188}AVQIN^{192}, which was not predicted by sequence alignments; and (iv) the β2/β3 loop is much shorter. Chemical shift perturbation experiments showed that the U2AF-binding RNA fragments interact with the four β-strands of RBD2 whereas, in contrast, they interact with β1, β3 and β4, but not with β2 or the α1/β2 loop, of RBD1. The characteristic α1–β2 structure of the hU2AF^{65} RBD1 may interact with other proteins, such as UAP56.

Keywords: nuclear magnetic resonance/RNA-binding protein/splicing/three-dimensional structure/U2AF

Introduction

Splicing of pre-mRNA is a multistep process mediated by a large RNA–protein complex termed the spliceosome (Moore *et al*., 1993). The U1, U2, U5 and U4/U6 small nuclear ribonucleoprotein particles (snRNPs) are the major components of the spliceosome, and consist of small nuclear RNAs and several polypeptides. In addition, several non-snRNP proteins are essential for the initiation of spliceosome assembly. One of the essential splicing factors is the human U2 snRNP auxiliary factor (hU2AF), which binds to the polypyrimidine tract immediately upstream of the 3′ splice site and is required for the subsequent interaction between the U2 snRNP and the branch point (Ruskin *et al*., 1988; Zamore *et al*., 1989). Two subunits, hU2AF^{65} and hU2AF^{35}, constitute hU2AF (Zamore *et al*., 1992; Zhang *et al*., 1992), but only the large subunit, hU2AF^{65}, contacts the polypyrimidine tract directly (Zamore *et al*., 1991, 1992).

The RNA-binding domain (RBD) consists of 80–90 amino acid residues with two well conserved motifs (RNP2 and RNP1) (Burd and Dreyfuss, 1994), and has been identified in a variety of RNA-binding proteins (Birney *et al*., 1993). The hU2AF^{65} sequence contains three tandem RBDs following an arginine/serine-rich (RS) domain (Figure 1). All three RBDs are required for high affinity binding of hU2AF^{65} to the polypyrimidine tract (Zamore *et al*., 1992). The tertiary structures of the RBDs from several different proteins have been determined (Nagai *et al*., 1990; Wittekind *et al*., 1992; Lee *et al*., 1994; Inoue *et al*., 1997; Shamoo *et al*., 1997; Xu *et al*., 1997; Chi *et al*., 1999; Crowder *et al*., 1999; Nagata *et al*., 1999a,b). These RDBs have a common βαββαβ-type folding topology. The two conserved motifs, RNP2 and RNP1, are located on the central β1 and β3 strands, respectively. The loop between β2 and β3 (the β2/β3 loop) varies significantly in sequence and length (Birney *et al*., 1993).

The RBD·RNA complex structures have been determined for three proteins. First, the structures of complexes between the N-terminal RBD of the U1 snRNP A (U1A) and stem–loop RNAs were determined by X-ray crystallography and by nuclear magnetic resonance (NMR) spectroscopy (Oubridge *et al*., 1994; Allain *et al*., 1996). Then, the crystal structure of the U2B″ RBD·U2A′·RNA ternary complex was determined, where the U2B″ RBD interacts with the stem–loop RNA in a manner similar to the U1A RBD (Price *et al*., 1998). Recently, the crystal structure of the complex between the two tandemly arranged RBDs of the *Drosophila* Sex-lethal (Sxl) protein and the target RNA, which is single stranded and lacks base pairs, was determined (Handa *et al*., 1999). The Sxl protein binds to the uridine-rich polypyrimidine tract of the *transformer* mRNA precursor in competition with U2AF, and regulates the sex-specific alternative splicing (Inoue *et al*., 1990; Valcárcel *et al*., 1993).

In the present study, we determined the solution structures of the first and second RBDs (RBD1 and RBD2, respectively) of hU2AF^{65} by NMR spectroscopy. As RBD1 is longer than RBD2 and other RBDs, it had been considered that the β2/β3 loop of RBD1 is unusually long. However, we found in the present study that the previous sequence alignment of hU2AF^{65} onto the βαββαβ fold was incorrect; the β2 strand of RBD1 was identified as ^{188}AVQIN^{192}. Correspondingly, the β2/β3 loop of RBD1 is unusually short, with only four residues. In contrast, the α1 helix of the hU2AF^{65} RBD1 is elongated by one turn as compared with other RBDs, and the loop between the α1 helix and the β2 strand (the α1/β2 loop) consists

of as many as 12 residues. This characteristic α1–β2 structure may be involved in the interaction with the 56 kDa U2AF-associated protein (UAP56), a DEAD-box splicing factor (Fleckner *et al*., 1997).

Results and discussion

Structure determination of the hU2AF^{65} RBD1 and RBD2

First, we measured the two-dimensional (2D) total correlation spectroscopy (TOCSY), 2D double quantum filtered correlation spectroscopy (DQF-COSY) and 2D nuclear Overhauser effect spectroscopy (NOESY) spectra of the non-labeled RBD1 and RBD2 proteins. We also measured the 2D ^{1}H–^{15}N heteronuclear single quantum coherence (HSQC), three-dimensional (3D) ^{15}N-edited TOCSY-HSQC and 3D ^{15}N-edited NOESY-HSQC spectra of the ^{15}N-labeled proteins. Figure 2A shows the 2D ^{1}H–^{15}N HSQC spectra of RBD1 and RBD2. Using these spectra, we elucidated the amino acid types of the amide protons. Then, we achieved the sequence-specific resonance assignment by tracing the d_{NN} and $d_{N\alpha}$ connectivities. Figure 2B shows the sequential and middle-range nuclear Overhauser effects (NOEs), the {^{1}H}–^{15}N NOE values and the $^{3}J_{HNH\alpha}$ coupling constants. The topology of the secondary structure elements, the four β-strands and the two α-helices, was determined with the middle and long-range NOEs for both RBD1 and RBD2. The inter-residue NOE network that forms the antiparallel β-sheet structure is shown in Figure 2C. The three-dimensional structures of RBD1 and RBD2 were determined from a total of 1039 and 1211 experimental restraints, respectively. A summary of the NMR-derived experimental constraints and the structural statistics of the 20 final simulated annealing structures for RBD1 and RBD2 are shown in Table I. These structures satisfy the distance restraints, and have no violations greater than 0.50 Å and no dihedral violations greater than 5.0°. The 20 RBD1 structures were superimposed with respect to the backbone atoms (N, C^{α} and C′) of residues 150–175, 186–192, 197–203 and 207–227, and the 20 RBD2 structures were superimposed with respect to those of residues 260–266, 272–291, 302–320 and 331–334 (Figure 3); the 20 structures were averaged and restrained-minimized to a mean structure. The average root mean square deviations of the backbone atoms of these elements are 0.36 ± 0.08 Å for RBD1 and 0.33 ± 0.06 Å for RBD2. The mean structures have been submitted to the Brookhaven Protein Data Bank (accession codes 1u2f and 2u2f for RBD1 and RBD2, respectively).

The secondary structure elements of the hU2AF^{65} RBD1 and RBD2

The hU2AF^{65} RBD1 and RBD2 have the βαββαβ fold, similar to other characterized RBD structures (Nagai *et al*., 1990; Wittekind *et al*., 1992; Lee *et al*., 1994; Inoue *et al*., 1997; Shamoo *et al*., 1997; Xu *et al*., 1997; Chi *et al*., 1999; Crowder *et al*., 1999; Nagata *et al*., 1999a,b). The arrangement of the secondary structure elements in the hU2AF^{65} RBD2 is similar to that in canonical RBDs; the present sequence alignment based on the secondary structure elements of RBD2 (Figures 2B and 4A) agrees with that proposed previously (Birney *et al*., 1993). On the other hand, for the hU2AF^{65} RBD1, the β2 strand had been assigned to Gly177–Ala181, and the loop between the β2 and β3 strands (the β2/β3 loop) had been considered to be unusually long (15 residues) (Birney *et al*., 1993). This sequence alignment seemed reasonable, as the β2/β3 loop varies in length and sequence even among typical RBDs, much more significantly than other secondary structure elements (Birney *et al*., 1993). However, the β2/β3 loop of the hU2AF^{65} RBD1 actually comprises only four residues (Figures 2B and 4A). Instead, the α1/β2 loop has as many as 12 residues, while the α1/β2 loop is five residues long in most RBDs (Figure 4A). Thus, the β2 strand was identified to be from Ala188 to Asn192 (Figures 2B and C, and 4A), which are posterior, by 11 amino acid residues, to the positions previously proposed in the hU2AF^{65} RBD1 sequence.

The tertiary structures of the hU2AF^{65} RBD1 and RBD2

Thus, the tertiary structure of the hU2AF^{65} RBD1 was found to be distinct from those of the canonical RBDs. We discuss below the structures of the hU2AF^{65} RBD1 and RBD2 in comparison with the U1A RBD1 (Nagai *et al*., 1990) and the Sxl RBD2 (Lee *et al*., 1994, Chi *et al*., 1999; Crowder *et al*., 1999). Figure 4A shows the amino acid sequences and the secondary structure elements

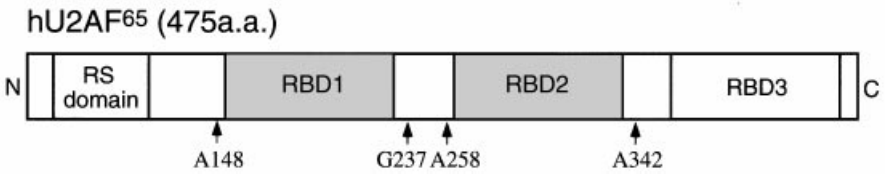


Fig. 1. A schematic diagram of the domain structure of hU2AF^{65}. The RS domain refers to the arginine/serine-rich domain, and RBD stands for an RNA-binding domain. The arrows indicate the N- and C-terminal positions of the polypeptides used in this study.

Fig. 2. (**A**) The ^{1}H–^{15}N HSQC spectra of the hU2AF^{65} RBD1 and RBD2. (**B**) A summary of the observed short- and middle-range NOEs, the {^{1}H}–^{15}N NOE values and the $^{3}J_{HNH\alpha}$ coupling constants. The intensities of the sequential NOEs are represented by the block height. The middle-range NOEs are shown with bars ending with two circles, and the intensity is represented by the thickness of the bar. The amide protons protected from $^{2}H_2O$ exchange are indicated with asterisks. The {^{1}H}–^{15}N NOE values are represented by the block height. 'P' indicates a proline residue, '△' indicates a residue that was not assigned, '#' indicates a residue whose signal was overlapped and whose NOE value was not determined, and '§' indicates a residue whose signal was too broad to determine the NOE value. Downward and upward pointing arrows indicate $^{3}J_{HNH\alpha}$ <5 Hz and $^{3}J_{HNH\alpha}$ >8 Hz, respectively, as measured by 3D HNHA spectra. The identified secondary structure elements are indicated. (**C**) A schematic representation of the antiparallel β-sheet structures of the hU2AF^{65} RBD1 and RBD2. Head-to-head arrows indicate the observed interresidual NOEs. The circles indicate the amide protons protected from $^{2}H_2O$ exchange.

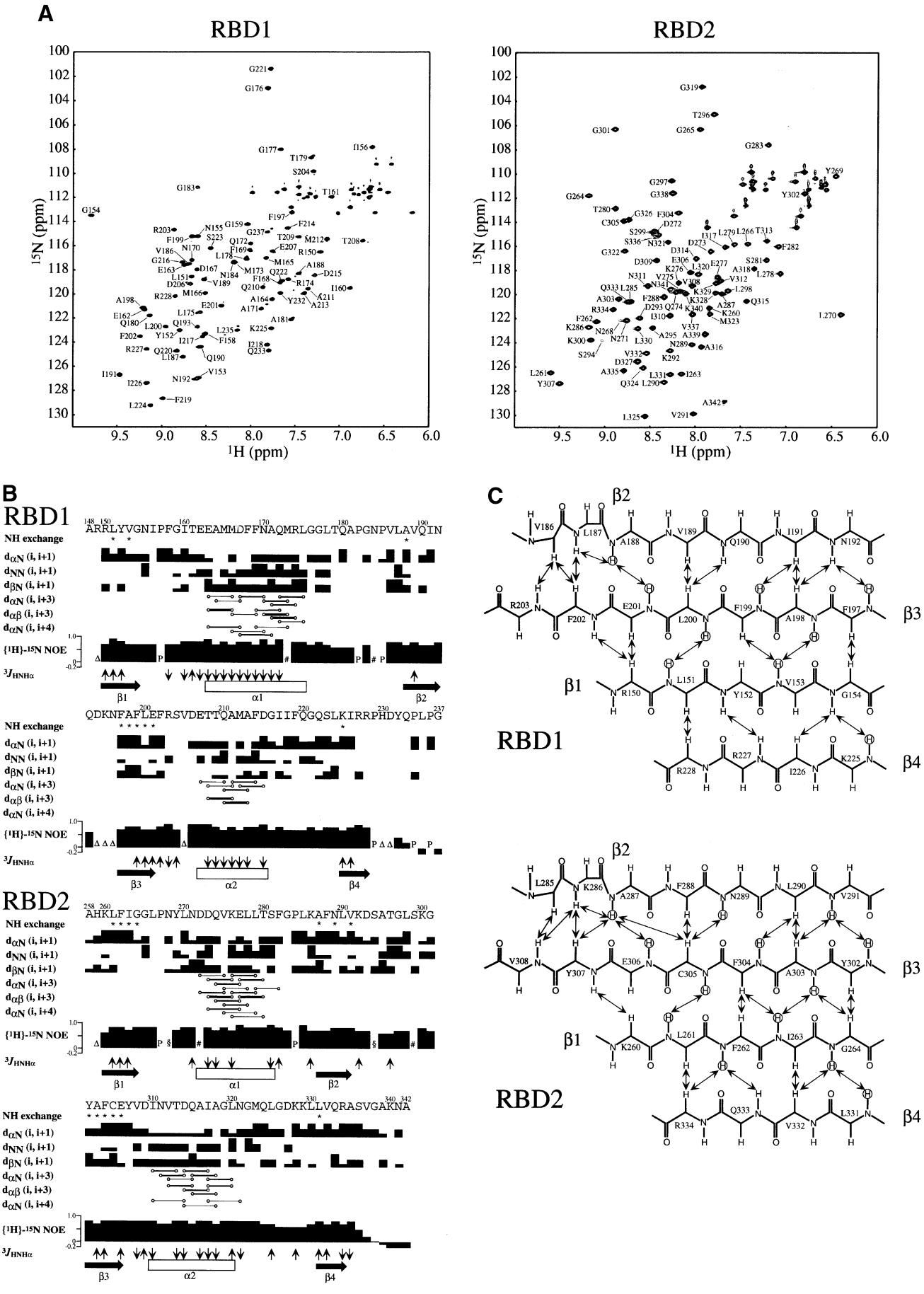

A
RBD1
RBD2
^{15}N (ppm)
^{1}H (ppm)
B
RBD1
148 150 160 170 180 190
ARRLYVGNIPFGITEEAMMDFFNAQMRLGGLTQAPGNPVLAVQIN
NH exchange
dαN (i, i+1)
dNN (i, i+1)
dβN (i, i+1)
dαN (i, i+3)
dαβ (i, i+3)
dαN (i, i+4)
{1H}-15N NOE
3JHNHα
β1
α1
β2
200 210 220 230 237
QDKNFAFLEFRSVDETTQAMAFDGIIFQGQSLKIRRPHDYQPLPG
β3
α2
β4
RBD2
258 260 270 280 290 300
AHKLFIGGLPNYLNDDQVKELLTSFGPLKAFNLVKDSATGLSKG
310 320 330 340 342
YAFCEYVDINVTDQAIAGLNGMQLGDKKLLVQRASVGAKNA
C
β2
β3
β1
β4
RBD1
RBD2

Table I. Summary of NMR-derived experimental constraints and structural statistics

	hU2AF^{65} RBD1	hU2AF^{65} RBD2
No. of distance and dihedral restraints		
NOE distances		
Total	943	1105
Intra-residue	394	426
Sequential ($\|i-j\|=1$)	216	257
Medium-range ($2\leqslant\|i-j\|\leqslant 4$)	100	142
Long-range ($\|i-j\|\geqslant 5$)	233	280
Hydrogen bonds	9	13
Dihedral angles	78	80
Distance restraint violations >0.5 Å	0	0
Dihedral angle restraint violations >5°	0	0

	⟨SA⟩	(SA)r	⟨SA⟩	(SA)r
X-PLOR energies (kcal/mol)				
*E*total	272 ± 18	218	102 ± 8	85
*E*noe	35 ± 7	20	15 ± 3	11
*E*cdih	1.8 ± 0.7	0.8	0.6 ± 0.3	0.5
R.m.s.d. from idealized geometry				
Bonds (Å)	0.003 ± 0.000	0.002	0.002 ± 0.000	0.002
Angles (°)	0.69 ± 0.014	0.65	0.43 ± 0.011	0.41
Impropers (°)	0.40 ± 0.013	0.35	0.28 ± 0.013	0.26
R.m.s.d. from experimental distance restraints				
All (Å)	0.027 ± 0.0027	0.020	0.016 ± 0.0017	0.013
R.m.s.d. from experimental dihedral angle restraints				
All (°)	0.61 ± 0.13	0.40	0.34 ± 0.09	0.32
Cartesian coordinate r.m.s.d. (Å)				
	150–175, 186–192, 197–203, 207–227		260–266, 272–291, 302–320, 331–334	
⟨SA⟩ vs SA	Backbone	All non-H	Backbone	All non-H
	0.36 ± 0.08	0.87 ± 0.14	0.33 ± 0.06	0.68 ± 0.04
Ramachandran statistics of (SA)r structure				
Percentage residues in				
most favored regions	57.3		65.3	
additional allowed regions	36.0		31.9	
generously allowed regions	6.7		2.8	
disallowed regions	0.0		0.0	

⟨SA⟩ refers to the final 20 simulated annealing structures; SA is the mean structure obtained by averaging the coordinates of the 20 individual SA structures; ⟨SA⟩r is the restrained minimized mean structure obtained by restrained regularization of the mean structure SA. Analysis of the Ramachandran plot was performed by PROCHECK-NMR program (Laskowski *et al*., 1996).

of several RBDs, and Figure 4B shows ribbon models of the hU2AF^{65} RBD1 and RBD2, the U1A RBD1 and the Sxl RBD2.

The hU2AF^{65} RBD1. First, the α1 helix of the hU2AF^{65} RBD1 is one turn longer at the C-terminus (Figure 4A), on the basis of the spatial location (Figure 4B), although the α1 helices of canonical RBDs, such as the hU2AF^{65} RBD2 and the Sxl RBD1 and RBD2, consist of 10 residues. The loop between the α1 helix and the β2 loop is the most characteristic structure in the hU2AF^{65} RBD1. In the other RBDs in Figure 4, the α1/β2 loop contains five residues, including a well conserved aromatic–Gly sequence required for the hydrophobic interaction between the α1 and α2 helices. However, in the hU2AF^{65} RBD1, an exceptionally long loop, consisting of 12 residues, exists between the α1 helix and the β2 strand, and lacks an aromatic–Gly sequence. Instead, around the α1/β2 loop of the hU2AF^{65} RBD1, there is a more extensive hydrophobic core (Figure 5). First, the hydrophobic side chains of Met173, Leu178, Thr179 and Pro185 are located on the inner side of the α1/β2 loop. In addition, Val205 from the β3/α2 loop participates in this hydrophobic interaction, which determines the orientation of the α1 and α2 helices. On the other hand, the hydrophilic side chains of Arg174, Gln180 and Asn184 are exposed to the solvent. As a result, the β2 strand in the hU2AF^{65} RBD1 consists of ^{188}AVQIN^{192}, which had not been identified from simple sequence alignment approaches. In the precise alignment determined in this study, the β2 strand structure of the hU2AF^{65} RBD1 shares a common feature with that of many RBDs. The hydrophobic side chains in the second and fourth positions of the β2 strand (Val189 and Ile191 in the hU2AF^{65} RBD1, for example) are buried inside the protein, while the side chains in the first, third and fifth positions of the β2 strand (Ala188, Gln190 and Asn192 in the hU2AF^{65} RBD1) are exposed on the β-sheet surface to the solvent. The β2 strand is followed by the short β2/β3 loop, consisting of only four amino acid residues, ^{193}QDKN^{196}. The β2/β3 loop is usually rich in basic amino acid residues, while there is one Lys residue in the short loop of the hU2AF^{65} RBD1. The β3/α2 loop of the hU2AF^{65} RBD1 has a different conformation from those of the other RBDs. First, the β3/α2 loop of the hU2AF^{65} RBD1 is composed of five amino acid residues (Figure 4A). Secondly, Val205 in the β3/α2 loop participates in the hydrophobic interaction with residues such as Pro185 in the characteristically elongated α1/β2 loop, as discussed above. As Figure 4 shows, the α2 helix in the hU2AF^{65} RBD1 is shorter than those in the other RBDs. The α2 helix in the hU2AF^{65} RBD1 is composed of nine residues, while the other RBDs are composed of 10 or 11 residues. Therefore, the α2 helix in the hU2AF^{65} RBD1 is a half turn shorter than the others at the N-terminus.

The hU2AF^{65} RBD2. The tertiary structure of the hU2AF^{65} RBD2 is almost the same as that of the Sxl RBD2 (Figure 4). The residues that form the hydrophobic core are well conserved in the hU2AF^{65} RBD2. One exceptional feature of the hU2AF^{65} RBD2 is the presence of a Leu residue at position 279 in the α1 helix, corresponding to a highly conserved aromatic residue (e.g. Phe34 in the U1A RBD1). However, Leu279 plays a role similar to that of the aromatic residues in the other RBDs, and participates in the hydrophobic core. The β2/β3 and α2/β4 loops in the hU2AF^{65} RBD2 appear to be more flexible, which is supported by the {^{1}H}–^{15}N NOE values of ~0.6 in these loops (Figure 2B).

In summary, the tertiary structure of the hU2AF^{65} RBD1 differs from those of the hU2AF^{65} RBD2, the U1A RBD1 and the Sxl RBD2, particularly in the region from the α1 helix to the β2/β3 loop, while the β-sheet structure with the RNP1 and RNP2 consensus motifs is well conserved among the four RBDs.

Sequence conservation among U2AFs

Homologs of hU2AF^{65} have already been cloned from several eukaryotes, such as mouse (Sailer *et al*., 1992),

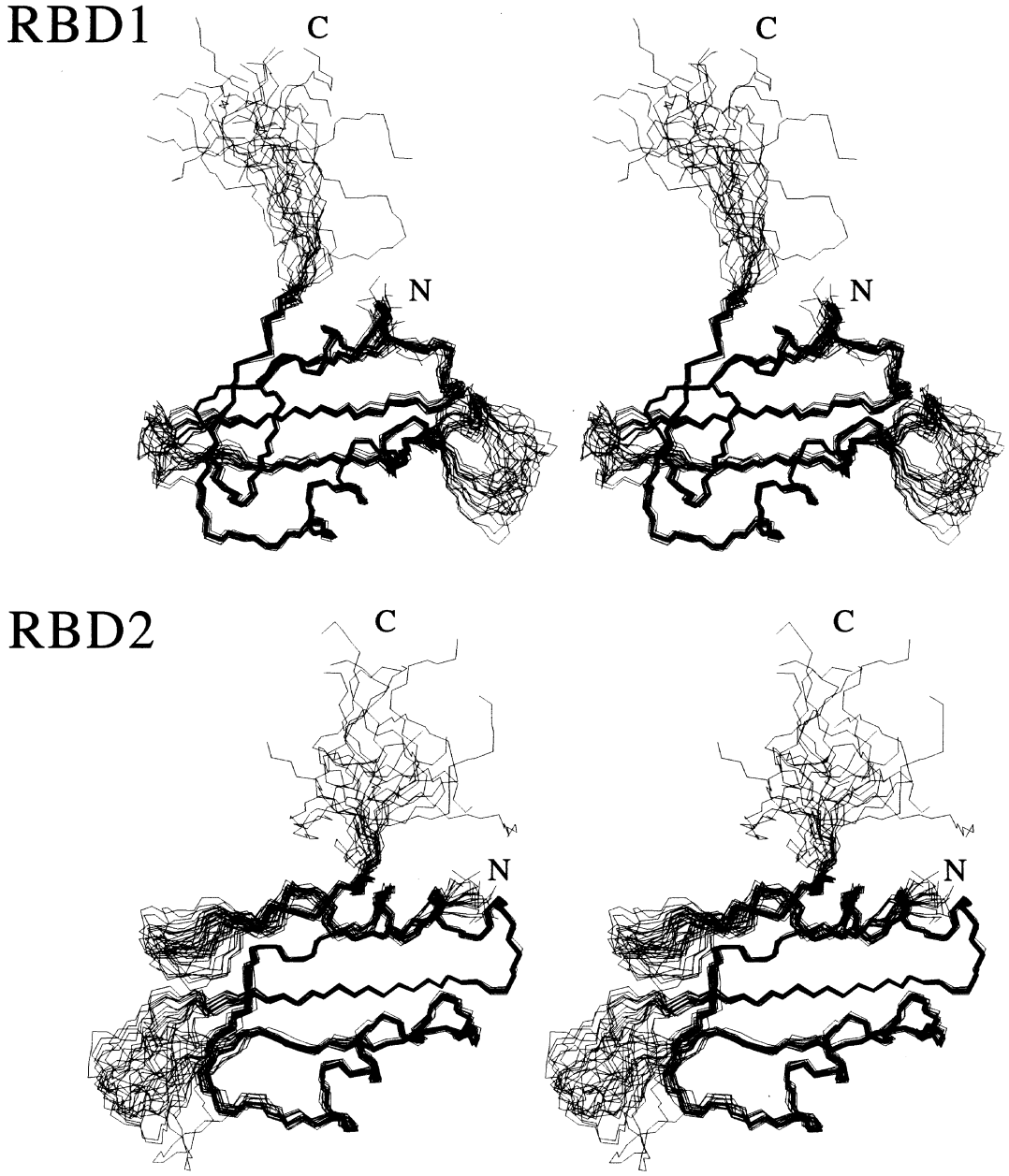


Fig. 3. The solution structures of the hU2AF^{65} RBD1 and RBD2. The 20 backbone structures are shown in stereoviews, which were superimposed for minimal root mean square deviation of the backbone atoms (N, C_α and C′) of residues 150–175, 186–192, 197–203 and 207–227 in RBD1, and residues 260–266, 272–291, 302–320 and 331–334 in RBD2. The figures were generated with MidasPlus (Ferrin *et al*., 1988).

Drosophila (Kanaar *et al*., 1993), *Caenorhabditis elegans*, *Caenorhabdits briggsae* (Zorio *et al*., 1997) and *Schizosaccharomyces pombe* (Potashkin *et al*., 1993). Except for the *S.pombe* U2AF, the entire amino acid sequences of the RBD1 and RBD2 domains are highly homologous (Figure 6). The identity is particularly high for the amino acid residues forming the hydrophobic cores, which have been found here for the hU2AF^{65} RBDs, shown in red in Figure 6. Furthermore, the amino acid residues of the α1–β2–β3 region of RBD1 are very well conserved among the five U2AFs, and therefore form the characteristic structure, as shown in Figure 5. On the other hand, in the case of the *S.pombe* U2AF, the hydrophobic core residues are well conserved in RBD2, but are appreciably less conserved in RBD1 (Figure 6). Therefore, it would be interesting to test experimentally whether the characteristic α1–β2–β3 structure is also present in the *S.pombe* U2AF RBD1.

Chemical shift perturbation experiments of the hU2AF^{65} RBD1 and RBD2 with target RNAs

We performed chemical shift perturbation experiments for RBD1 and RBD2 with three RNA sequences:$U_5C_3U_5$, $ACUCU_4CACAUAG$ and A_{15}. The $U_5C_3U_5$ sequence had been found by SELEX searching for RNAs that bind to the full-length hU2AF^{65} (Singh *et al*., 1995). On the other hand, the $ACUCU_4CACAUAG$ sequence is the female-specific 3′ splice site of the *transformer* pre-mRNA in *Drosophila*, which is known as a typical U2AF-binding site (Valcárcel *et al*., 1993). The A_{15} sequence is a control

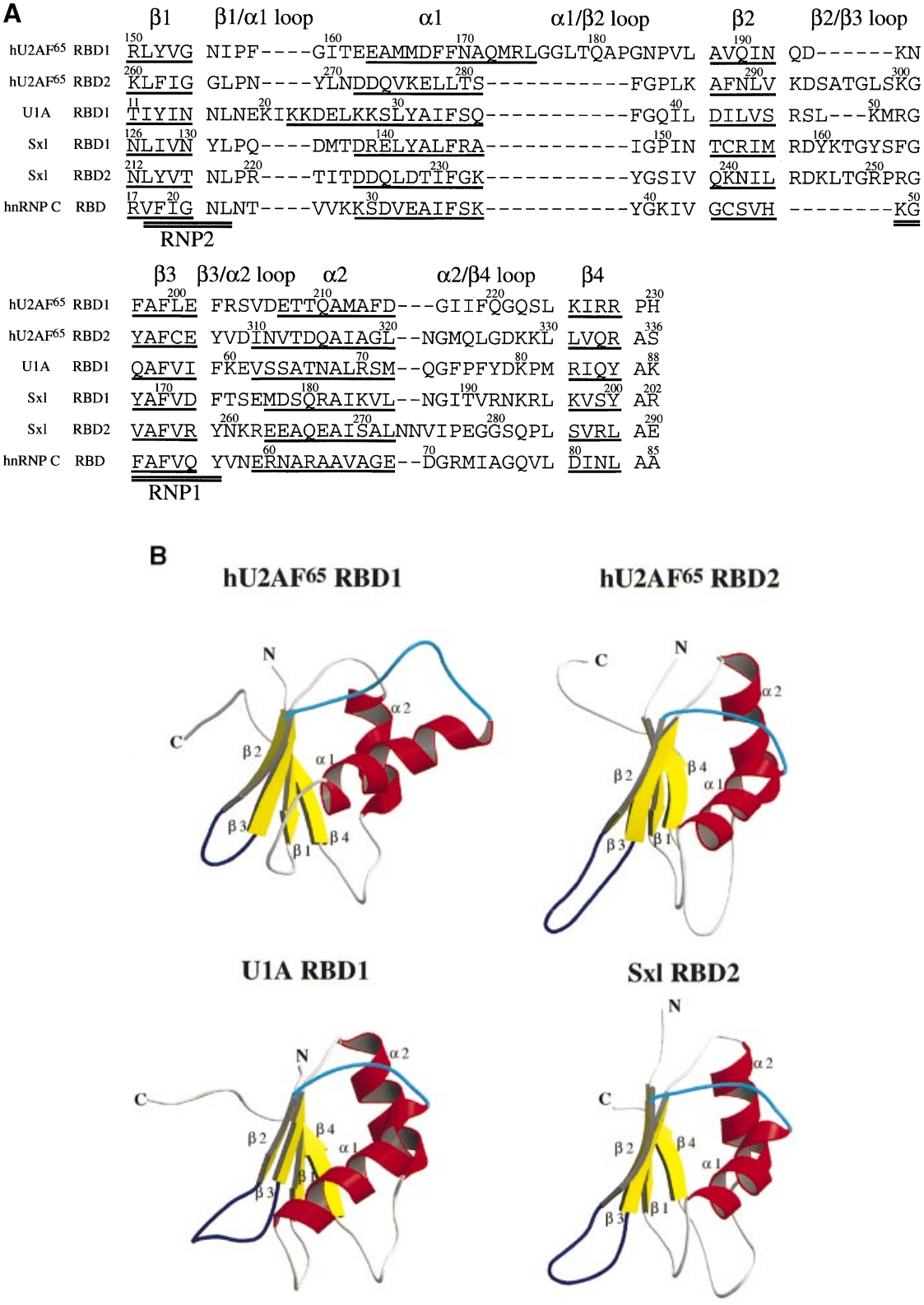


Fig. 4. (**A**) Alignment of the RBD sequences of hU2AF^{65} (Zamore *et al*., 1992), U1A (Sillekens *et al*., 1987), Sxl (Bell *et al*., 1988) and hnRNP C (Burd *et al*., 1989). The secondary structure elements are underlined. The RNP2 and RNP1 sequences are double underlined. (**B**) Ribbon models of the hU2AF^{65} RBD1 and RBD2 in comparison with those of the U1A RBD1 and the Sxl RBD2. On the models, the α1/β2 loops of the proteins are shown in cyan and the β2/β3 loops in blue. This figure and Figures 5, 7G and 8 were generated with MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt *et al*., 1997).

sequence with no pyrimidine residue. The chemical shift differences between the hU2AF^{65} RBD (RBD1 or RBD2) alone and the 1:1 mixture of RBD and RNA are summarized in Figure 7A–F. The A_{15} sequence actually has very small effects on the chemical shifts of RBD1 and RBD2 (Figure 7C and 7F). In contrast, the chemical shift perturb-

ations with the target RNAs, $U_5C_3U_5$ (Figure 7A and D) and ACUCU$_4$CACAUAG (Figure 7B and E), are much larger than those with A_{15}. Therefore, these RBD fragments are capable of sequence-specific RNA binding. Upon addition of the target RNA, a number of peaks on the 2D ^{1}H–^{15}N HSQC spectra shifted as the molar ratio of the RNA increased, indicating that the exchange between the free and RNA-bound states was fast on the NMR time scale. The exceptions are the resonances of RBD2 residues 336, 338 and 339 which broadened in the presence of ACUCU$_4$CACAUAG (intermediate exchange). Relatively fast exchange between RBD and RNA has also been observed for other RBDs (Lee *et al*., 1997; Chi *et al.*, 1999).

Curve-fitting analysis was done for the amide ^{1}H–^{15}N chemical shifts of Val153, Gly154, Leu200, Phe202, Ile226 and Arg228 of RBD1 and Leu261, Phe262, Val291, Gly301, Tyr307 and Arg334 of RBD2, at the RNA/RBD molar ratios of 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. Thus, the dissociation constants of the complexes with $U_5C_3U_5$ and ACUCU$_4$CACAUAG were determined to be 0.7 ± 0.2 mM and 1.1 ± 0.4 mM, respectively, for RBD1 and 0.2 ± 0.05 mM and 0.3 ± 0.08 mM, respectively, for RBD2, as well as the chemical shift values of those amino acid residues in the RNA-bound state. The chemical shift difference data with the non-cognate RNA, A_{15}, were not large enough to determine the dissociation constants quantitatively. Intriguingly, the SELEX-derived $U_5C_3U_5$ sequence binds more tightly than the natural ACUCU$_4$CACAUAG sequence to both RBD1 and RBD2. On the other hand, both of the RNAs bind more tightly to RBD2 than to RBD1. In the condition of the 1:1 mixture of RBD and RNA (as in Figure 7A–F), 55 and 44% of RBD1 and 75 and 68% of RBD2 are bound to the target RNAs, $U_5C_3U_5$ and ACUCU$_4$CACAUAG, respectively. From these RNA-bound fractions and the observed values of the chemical shift difference ($\Delta H + \Delta N$) in the 1:1 mixture conditions, the $\Delta H + \Delta N$ values in the RNA-bound states can be calculated. According to these $(\Delta H + \Delta N)_{bound}$ values, the amino acid residues were divided into three classes: 'large' [$(\Delta H + \Delta N)_{bound} \geqslant 150$ Hz], 'medium' [$80 \text{ Hz} \leqslant (\Delta H + \Delta N)_{bound} < 150$ Hz] and 'small or negligible' [$(\Delta H + \Delta N)_{bound} < 80$ Hz]. These classifications are color-coded on the ribbon structures of the RNA-free RBD1 and RBD2 (Figure 7G). The patterns of the chemical shift perturbations with the two distinct target RNAs, $U_5C_3U_5$ and ACUCU$_4$CACAUAG, are remarkably similar to each other. This result is consistent with the results of a previous gel-shift assay showing that the hU2AF^{65} RBD1 and RBD2–RBD3 fragments, in their isolated states, are capable of binding to three types of polypyrimidine tract RNAs, though less strongly than the three-domain fragment (Zamore *et al.*, 1992).

In all four cases (Figure 7G), the chemical shift perturbations were much larger for the β-sheet than for the α-helices, which was also the case for the other RBDs of the heterogeneous nuclear ribonucleoprotein (hnRNP) C (Görlach *et al*., 1992) and the Sxl protein (Lee *et al*., 1997; Chi *et al.*, 1999). In the crystal structures of the U1A RBD1·RNA complex (Oubridge *et al*., 1994) and the Sxl RBD1–RBD2·RNA complex (Handa *et al*., 1999),

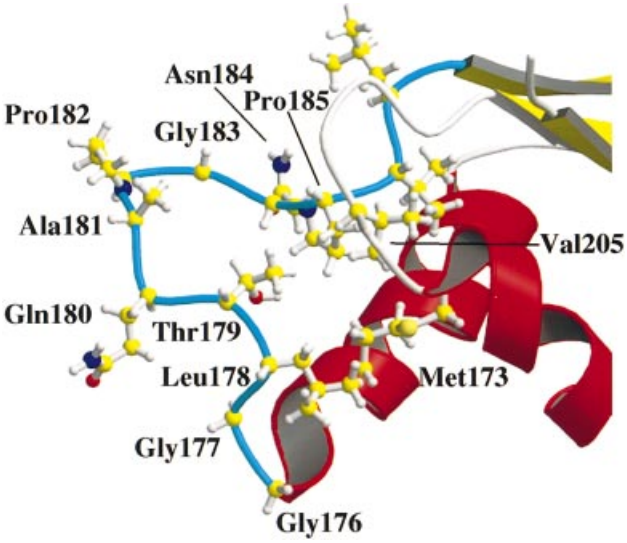


Fig. 5. The structure of the α1/β2 loop of the hU2AF^{65} RBD1. Met173 in the α1 helix and Val205 in the β3/α2 loop are also shown.

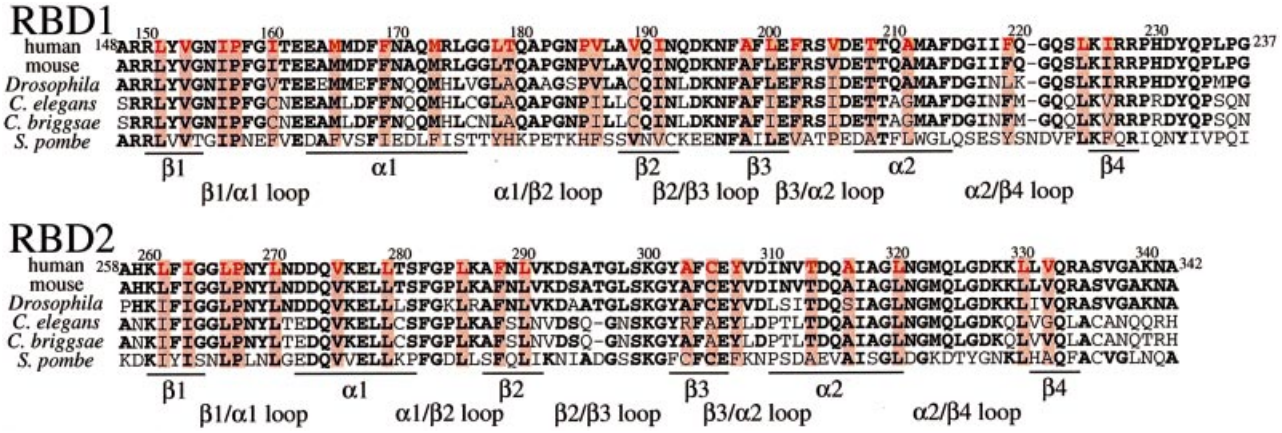
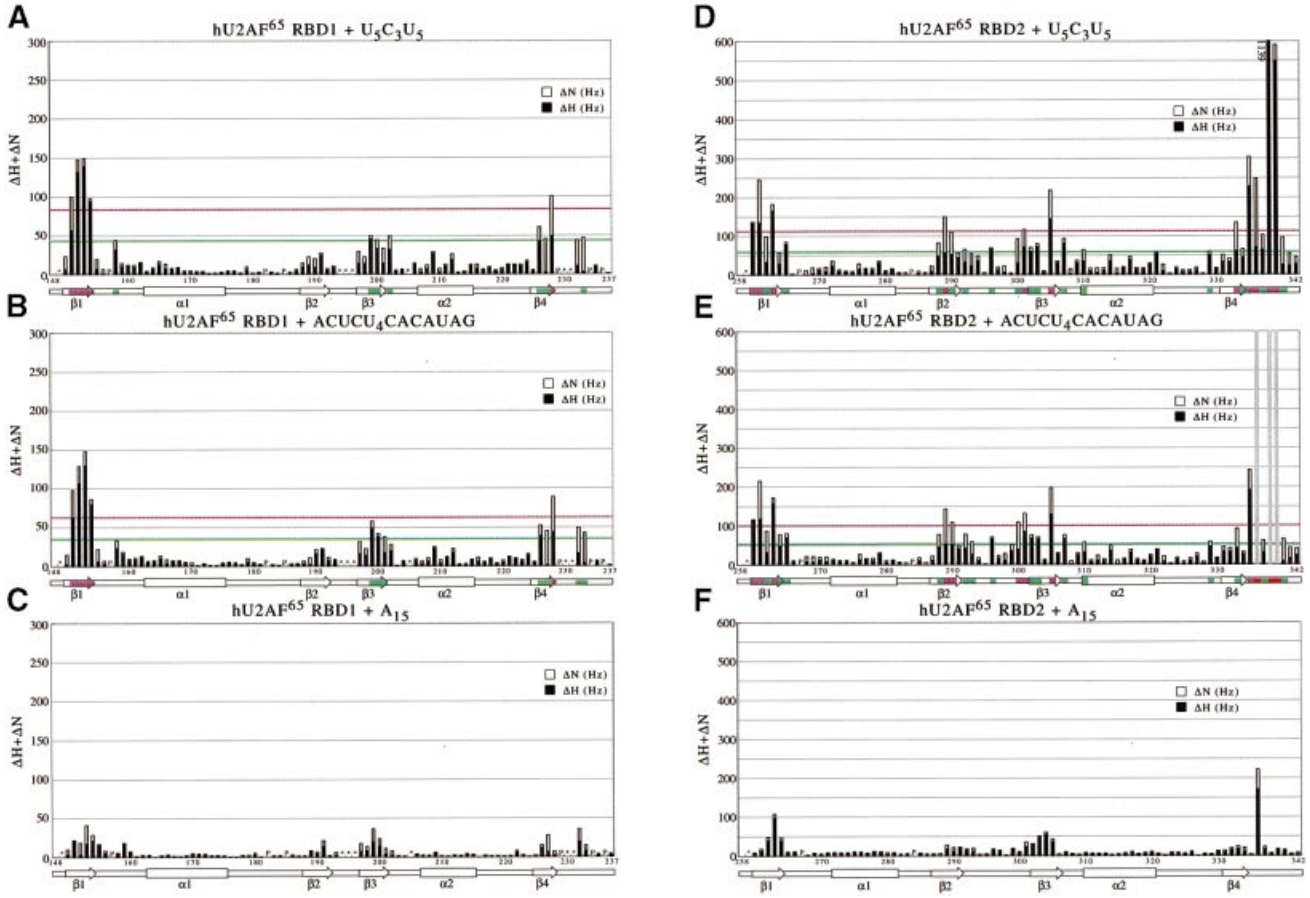
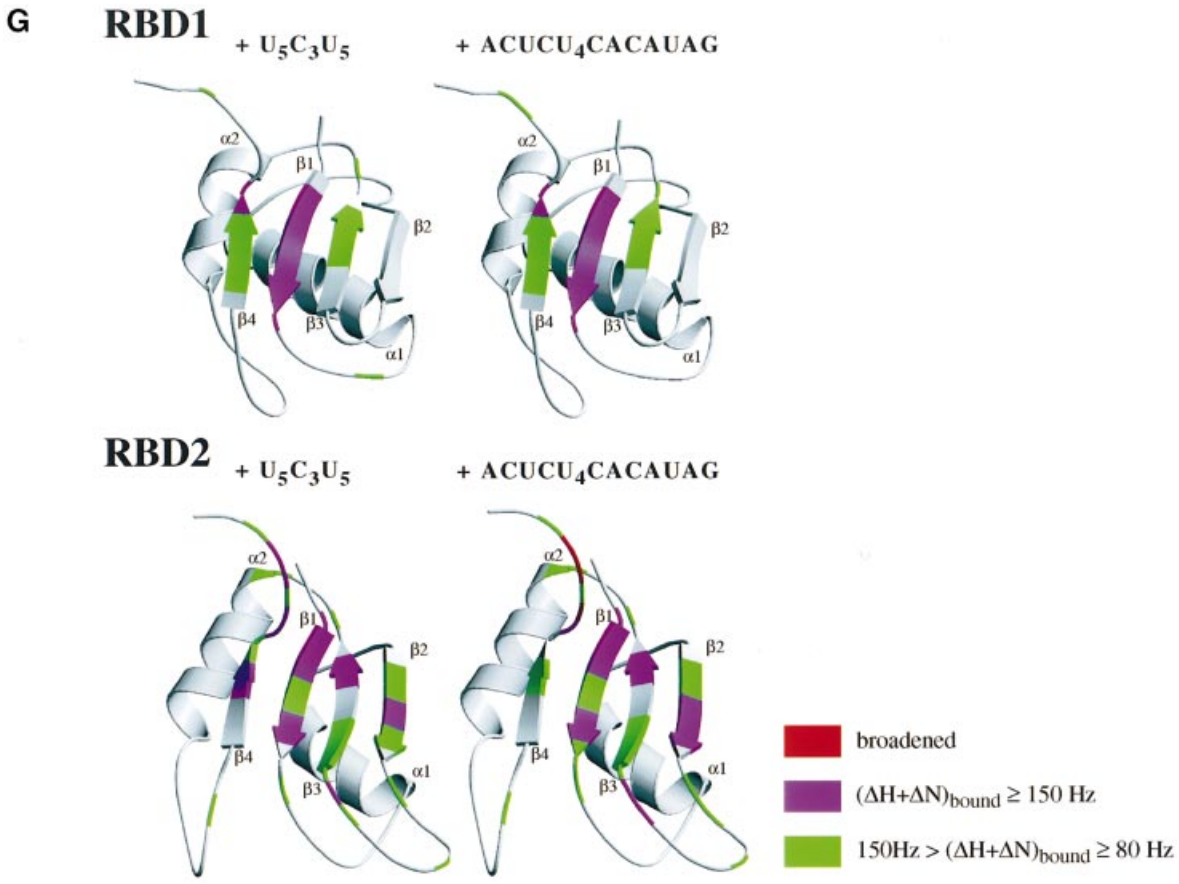


Fig. 6. Sequence alignment of RBD1 and RBD2 of the large subunit U2AFs from human (Zamore *et al*., 1992), mouse (Sailer *et al*., 1992), *Drosophila* (Kanaar *et al*., 1993), *C.elegans* (Zorio *et al*., 1997), *C.briggsae* (Zorio *et al*., 1997) and *S.pombe* (Potashkin *et al.*, 1993). The residues that are identical to those of hU2AF^{65} are indicated in bold. The residues whose side chains are buried inside the protein in the hU2AF^{65} RBD1 and RBD2 are colored in red, and the corresponding residues of the other species are outlined in red. The secondary structure elements are shown at the bottom.


A
hU2AF65 RBD1 + U5C3U5
B
hU2AF65 RBD1 + ACUCU4CACAUAG
C
hU2AF65 RBD1 + A15
D
hU2AF65 RBD2 + U5C3U5
E
hU2AF65 RBD2 + ACUCU4CACAUAG
F
hU2AF65 RBD2 + A15
ΔN (Hz)
ΔH (Hz)
ΔH+ΔN
β1 α1 β2 β3 α2 β4


G
RBD1
+ U5C3U5
+ ACUCU4CACAUAG
RBD2
+ U5C3U5
+ ACUCU4CACAUAG
α2 β1 β2 β4 β3 α1
broadened
(ΔH+ΔN)bound ≥ 150 Hz
150Hz > (ΔH+ΔN)bound ≥ 80 Hz

five and three nucleotide residues are bound on and around the β-sheet of the U1A RBD1 and the Sxl RBD2, respectively. The central three base-binding sites of the U1A RBD1 correspond spatially to the three sites of the Sxl RBD2: the first site between the β1/α1 and β2/β3 loops, the second site on the aromatic ring at the second position of the RNP2 motif located in the β1 strand and the third site on the aromatic ring at the fifth position of the RNP1 motif located in the β3 strand. The amino acid sequences of the hU2AF^{65} RBD1 and RBD2 are consistent with the structural properties of these three base-binding sites. Therefore, it is possible that the hU2AF^{65} RBD1 and RBD2 each possess three base-binding sites (Figure 8), corresponding to those conserved in the U1A RBD1 and the Sxl RBD2. This hypothesis is consistent with the chemical shift perturbations by RNA binding of the hU2AF^{65} RBD1 and RBD2. In addition, the resonances from the C-terminal region following the β4 strand are significantly affected in both RBD1 and RBD2, suggesting that this C-terminal region interacts strongly with the RNA. In fact, the corresponding C-terminal region of the U1A RBD1 interacts with RNA (Oubridge *et al*., 1994).

Further comparisons of the chemical shift perturbation patterns between the hU2AF^{65} RBD1 and RBD2 indicate that the RNA interaction modes of these two RBDs are not the same. As for RBD1, the β1 strand exhibited the largest perturbations upon specific interactions with RNAs (Figure 7A and B). The next largest perturbations were observed for the β3 and β4 strands and the C-terminal region, which follows the β4 strand. The resonances from the β2 strand, the α2 helix and the β1/α1 loop were slightly perturbed, and those of the α1 helix, and the α1/β2, β3/α2 and α2/β4 loops were only negligibly affected. On the other hand, in the specific interactions of RBD2 with RNAs, the resonances from all four β strands were significantly perturbed, and those from the C-terminal region were especially affected. The resonances of the two α-helices and the loops around these helices were only slightly affected. Therefore, the major differences in the perturbation patterns between RBD1 and RBD2 occur in the C-terminal region and the β2 strand. The interaction of the C-terminal region with RNA appears to be much stronger in RBD2 than in RBD1. These remarkably large chemical shift perturbations of the C-terminal region of RBD2 suggest that a main chain amide or carbonyl group forms a hydrogen bond with the RNA, as in the U1A RBD1·RNA complex (Oubridge *et al*., 1994). The interaction of the β2 strand of RBD2 with the RNA is probably as significant as those of the other three β-strands, while the β2 strand of RBD1 does not seem to be very involved in the RNA interaction.

Although the structure around the β2 strand of RBD1, including the unusually long, well structured α1/β2 loop, is characteristically well conserved among U2AFs, the α1/β2 loop of the hU2AF^{65} RBD1 is not involved in RNA binding, as indicated by the present chemical shift perturbation experiments. Therefore, the presence of the long α1/β2 loop around the β2 strand of RBD1 might

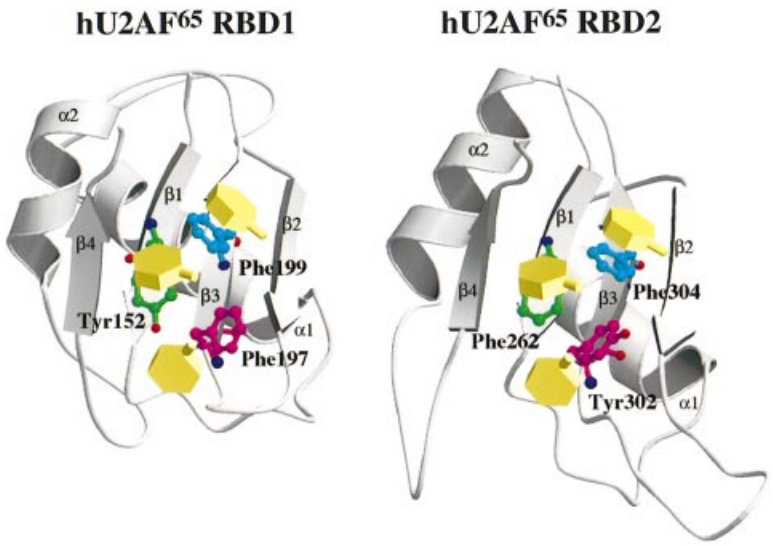


Fig. 8. Possible base-binding positions of the hU2AF^{65} RBD1 and RBD2. The amino acid residues that might stack with bases in the first, second and third canonical base-binding sites are shown, colored in magenta, green and cyan, respectively. On each binding site, a schematic pyrimidine base and a N1–C1′ bond are presented.

Fig. 7. (A–F) Chemical shift perturbations of RBD1 (**A**–**C**) and RBD2 (**D**–**F**). The absolute values of the differences in the chemical shifts of the ^{1}H (solid bars) and ^{15}N (open bars) resonances (in Hz at a ^{1}H frequency of 500 MHz) between the 1:0 and 1:1 molar ratios of hU2AF^{65} RBD1 and RNA ($U_5C_3U_5$, $ACUCU_4CACAUAG$ or A_{15}) are plotted versus the amino acid sequence of RBD1 and RBD2. The gray bar indicates a residue whose signal largely broadened during the titration. 'P' indicates a proline residue; '*' indicates a residue that was not assigned. The green and magenta lines (A, B, D and E) indicate the thresholds for the 'medium' and 'large' perturbations, respectively, corresponding to the $(\Delta H + \Delta N)_{bound}$ values (the $\Delta H + \Delta N$ values calculated for the fully RNA-bound RBDs with the dissociation constants of the complexes) of 80 and 150 Hz, respectively. The secondary structural elements of the hU2AF^{65} RBD1 and RBD2 are shown at the bottom, and are colored in green ('medium' perturbation), magenta ('large' perturbation) and red (the signals broadened during the titration) (A, B, D and E). (**G**) Distribution of chemical shift perturbations of the hU2AF^{65} RBD1 and RBD2 with $U_5C_3U_5$ or $ACUCU_4CACAUAG$. The distribution is displayed on a ribbon model of the solution structure of the hU2AF^{65} RBD1 and RBD2 in the free state. The color coding of the perturbations is the same as in (A), (B), (D) and (E).

affect its interaction with RNA. On the other hand, the conserved α1/β2 loop of the hU2AF^{65} RBD1 might have some role in steps other than target RNA binding, such as protein–protein interactions, in the splicing reaction. Fleckner *et al*. (1997) showed that residues 138–183, including the β1 strand to the α1/β2 loop, of the hU2AF^{65} RBD1 are important for the overall U2AF function, but not for the RNA binding, and are crucial for the interaction with the 56 kDa U2AF associated protein (UAP56), a DEAD-box splicing factor also involved in U2 snRNP recruitment. It seems that the α1/β2 loop, with its strikingly long and defined conformation, and/or the C-terminally extended α1 helix of the hU2AF^{65} RBD1 are good candidates for sites of interaction with UAP56.

Materials and methods

Preparation of RBD1 and RBD2 of hU2AF^{65}

For the structural analyses, we constructed overproduction systems using the cDNAs encoding the residues Ala148–Gly237 for RBD1 and residues Ala258–Ala342 for RBD2. These proteins were expressed in *Escherichia coli* BL21 (DE3) by the protein expression vector pK7 (Kigawa *et al*., 1995). LB broth was used for the non-labeled proteins, while a modified minimal medium (15 g/l Na_2HPO_4, 6 g/l KH_2PO_4 and 5 g/l NaCl) containing 1 g/l $^{15}NH_4Cl$, 4 g/l glucose, 20 mg/l thiamine, 1 g/l $MgSO_4$, 25 mg/l kanamycin and metals [100 mg/l $FeSO_4$, 1 mg/l $CuSO_4$, 0.5 mg/l $MnSO_4$, 2 mg/l $CaCl_2$, 0.22 mg/l $Na_2B_4O_7$, 0.1 mg/l $(NH_4)_6Mo_7O_{24}$, 2.2 mg/l $ZnSO_4$ and 0.6 mg/l $CoCl_2$] was used for the ^{15}N-labeled proteins. In the case of the modified minimal medium, expression was induced with isopropyl-β-D-thiogalactopyranoside when the cells reached an A_{600} of 0.8. After 24 h of cultivation, the cells were harvested. After cell lysis by sonication and centrifugation to remove the cellular debris, the supernatant was applied to a DEAE–Sephacel (Pharmacia) anion exchange column. The proteins were eluted with a concentration gradient of sodium chloride up to 500 mM. Then, in the case of RBD1, we purified the protein with a butyl-Toyopearl (Tosoh, Japan) hydrophobic column with a reverse concentration gradient of ammonium sulfate from 800 to 0 mM. An FPLC Mono-Q anion exchange column (Pharmacia) was used for the final step of purification. In the case of RBD2, after DEAE–Sephacel column chromatography, the protein was purified by CM-Toyopearl (Tosoh, Japan) cation exchange column chromatography with a concentration gradient of ammonium formate up to 500 mM. Finally, FPLC Mono-S cation exchange column (Pharmacia) chromatography was performed. Each preparation yielded 5–20 mg protein/l of culture. These polypeptides were soluble and sufficiently stable for NMR structure determinations.

NMR measurements

NMR spectra were measured with Bruker DMX500 and DRX600 spectrometers at a probe temperature of 298 K. The NMR samples were prepared in 100 mM sodium phosphate buffer (pH 7.5 for RBD1 and pH 6.5 for RBD2) containing either 10 or 100% ^2H_2O. In the homonuclear 2D proton NMR measurements, the water suppression was performed by selective pre-irradiation. In the TOCSY experiments (Bax and Davis, 1985; Davis and Bax, 1985), an MLEV17 pulse train of 45 ms was used for isotropic mixing. In the NOESY experiments (Jeener *et al*., 1979), mixing times were in the range of 30–150 ms. All of the homonuclear 2D proton spectra were measured by the States-TPPI method, with 2048 data points in the t_2 domain. In the t_1 domain, 256 increments in the NOESY and TOCSY experiments and 400 increments in the DQF-COSY experiment (Rance *et al*., 1983) were performed. The 2D ^{1}H–^{15}N HSQC spectra (Bodenhausen *et al*., 1980) were measured by the States-TPPI method, with 1024 data points in the t_2 (^{1}H) domain and 256 increments in the t_1 (^{15}N) domain. The 3D ^{15}N-edited TOCSY-HSQC experiment and the 3D ^{15}N-edited NOESY-HSQC experiment (Marion *et al*., 1989a,b) were conducted by the States-TPPI method, with 512 data points in the t_3 (^{1}H) domain, 64 increments in the t_2 (^{1}H) domain and 32 increments in the t_1 (^{15}N) domain. In the 3D ^{15}N-edited NOESY-HSQC experiment, a 150 ms mixing time was used. In the 3D ^{15}N-edited TOCSY-HSQC experiment, a DIPSI-2rc pulse train of 47 ms was used (Cavanagh and Rance, 1992). The 3D HNHA experiments (Vuister *et al*., 1993) were performed by the States-TPPI method, with 1024 data points in the t_3 (^{1}H) domain, 32 increments in the t_2 (^{1}H) domain and 14 increments in the t_1 (^{15}N) domain. The 2D {^{1}H}–^{15}N NOE spectra (Grzesiek and Bax, 1993) were measured by the States-TPPI method, with 1024 data points in the t_2 (^{1}H) domain and 128 increments in the t_1 (^{15}N) domain. The NMR data were processed using the NMRpipe system (Delaglio *et al*., 1995). The NMRView software package (Johnson *et al*., 1994) was used to analyze the processed spectra.

Distance restraints

Interproton distance restraints were obtained from the 2D NOESY and 3D ^{15}N-edited NOESY-HSQC spectra. The distance restraints were calculated as $kI^{-1/6}$, where I is the peak intensity. The constant k was calibrated so that the peak intensity of the strong NOE between $C^\alpha H$ and $C^\alpha H$ in the antiparallel β-sheet corresponded to a 2.3 Å distance. The assigned NOEs were classified into upper bound distance restraints of 2.5, 3.5 and 5.0 Å. The lower bounds for the interproton distance restraints were set to the sum of the van der Waals radii of two protons, 1.8 Å. Distances involving methyl protons, aromatic ring protons and non-stereospecifically assigned methylene protons were corrected appropriately for center averaging (Wüthrich *et al*., 1983). In addition, 0.5 Å was added to the upper limits for distances involving methyl protons (Clore *et al*., 1987).

Dihedral angle restraints and stereospecific assignments

Stereospecific assignments of the β-methylene protons, the φ angle restraints of Gly159 and the χ_1 and χ_2 angle restraints were obtained from the TOCSY, NOESY and rotating frame Overhauser effect spectroscopy (ROESY) (Bothner-By *et al*., 1984; Griesinger and Ernst, 1987) spectra using short mixing times. The other φ angle restraints were based on $^3J_{HNH\alpha} > 8$ Hz or < 5 Hz, as measured from the 3D HNHA spectra.

Hydrogen bond restraints

The amide protons that exchanged at slow rates with the solvent protons were identified by the NOESY and TOCSY spectra in the ^2H_2O solution. Hydrogen bond restraints within the secondary structure elements were included in the set of distance restraints. The upper and lower limits of the constraints for N–O were 3.5 and 2.5 Å respectively, and those for HN–O were 2.5 and 1.5 Å, respectively (Clore *et al*., 1991).

Structure calculations

All calculations were carried out using the simulated annealing protocol (Nilges *et al*., 1988; Nilges, 1993) with X-PLOR version 3.1 (Brünger, 1993). In the case of RBD1, the final structure calculation was based on 1039 restraints: 943 interproton distance restraints (549 inter-residue and 394 intra-residue restraints), 18 distance restraints for backbone hydrogen bonds and 78 dihedral angle restraints (35 φ, 42 χ_1 and one χ_2 angle). All structure calculations were performed on an O2 workstation (Silicon Graphics). A total of 200 calculations were carried out on the NMR-derived distance information. We selected 20 out of 107 converged structures, based on the criteria of the smallest residual energy values of the distance restraints, the dihedral angle restraints and the van der Waals repulsion. There were no distance restraints violated by > 0.5 Å and no dihedral angle restraints violated by $>5°$ in any structure. In the case of RBD2, the final structure calculation was based on 1211 restraints: 1105 interproton distance restraints (679 inter-residue and 426 intra-residue restraints), 26 distance restraints for backbone hydrogen bonds and 80 dihedral angle restraints (31 φ, 47 χ_1 and two χ_2 angles). Without the restraints, all processes were the same as for the calculation of RBD1, and we selected 20 out of 127 converged structures of RBD2.

Preparation of RNAs

Each oligonucleotide was synthesized on a DNA/RNA synthesizer on a 1 μM scale. The final dimethoxytrityl group was removed. Deprotection was performed with methylamine (MA) and anhydrous triethylamine/hydrogen fluoride in *N*-methylpyrrolidinone (TEA·HF/NMP) as described by Wincott *et al*. (1995). Deprotected samples were purified by 20% polyacrylamide gel electrophoresis (PAGE). After PAGE, the band was located by UV shadowing, excised and eluted with H_2O at 37°C for 2 days. The purified oligonucleotide was desalted by a Sep-Pak cartridge (Waters). The eluted sample was evaporated and then checked by UV spectroscopy.

Chemical shift perturbation experiment

The oligonucleotide was dissolved in 10 μl of H_2O to a concentration of 16 mM. The protein concentration was 0.8 mM in 200 μl of buffer. The oligonucleotide solution (2 μl) was added to the protein sample. For each ratio, the 2D ^{1}H–^{15}N HSQC spectrum was measured. These steps were repeated until an RNA–protein ratio of 1:1 was achieved.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (No. 04272103) from the Ministry of Education, Science and Culture of Japan. T.I. was supported by a fellowship from the Japan Society for the Promotion of Science for Japanese Junior Scientists.

References

Allain,F.H.-T., Gubser,C.C., Howe,P.W.A., Nagai,K., Neuhaus,D. and Varani,G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature*, **380**, 646–650.

Bax,A. and Davis,D.G. (1985) MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.*, **65**, 355–360.

Bell,L.R., Maine,E.M., Schedl,P. and Cline,T.W. (1988) *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell*, **55**, 1037–1046.

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

Bodenhausen,G. and Ruben,D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.*, **69**, 185–189.

Bothner-By,A.A., Stephens,R.L., Lee,J.-m., Warren,C.D. and Jeanloz, R.W. (1984) Structure determination of a tetrasaccharide: transient nuclear overhauser effects in the rotating frame. *J. Am. Chem. Soc.*, **106**, 811–813.

Brünger,A.T. (1993) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Burd,C.G., Swanson,M.S., Görlach,M. and Dreyfuss,G. (1989) Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc. Natl Acad. Sci. USA*, **86**, 9788–9792.

Cavanagh,J. and Rance,M. (1992) Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. *J. Magn. Reson.*, **96**, 670–678.

Chi,S.-W., Muto,Y., Inoue,M., Kim,I., Sakamoto,H., Shimura,Y., Yokoyama,S., Choi,B.-S. and Kim,H. (1999) Chemical shift perturbation studies of the interactions of the second RNA-binding domain of the *Drosophila* sex-lethal protein with the *transformer* pre-mRNA polyuridine tract and 3′ splice-site sequences. *Eur. J. Biochem.*, **260**, 649–660.

Clore,G.M., Gronenborn,A.M., Nilges,M. and Ryan,C.A. (1987) Three-dimensional structures of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry*, **26**, 8012–8023.

Clore,G.M., Wingfield,P.T. and Gronenborn,A.M. (1991) High-resolution three-dimensional structure of interleukin 1β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. *Biochemistry*, **30**, 2315–2323.

Crowder,S.M., Kanaar,R., Rio,D.C. and Alber,T. (1999) Absence of interdomian contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc. Natl Acad. Sci. USA*, **96**, 4892–4897.

Davis,D.G. and Bax,A. (1985) Assignment of complex ^{1}H NMR spectra via two-dimensional homonuclear Hartmann–Hahn spectroscopy. *J. Am. Chem. Soc.*, **107**, 2820–2821.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX PIPES. *J. Biomol. NMR*, **6**, 277–293.

Ferrin,T.E., Huang,C.C., Jarvis,L.E. and Landridge,R. (1988) The MIDAS display system. *J. Mol. Graphics*, **6**, 13–27.

Fleckner,J., Zhang,M., Valcárcel,J., and Green,M.R. (1997) U2AF^{65} recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.*, **11**, 1864–1872.

Görlach,M., Wittekind,M., Beckman,R.A., Mueller,L. and Dreyfuss,G. (1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.*, **11**, 3289–3295.

Grzesiek,S. and Bax,A. (1993) The importance of not saturating H_2O in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.*, **115**, 12593–12594.

Griesinger,C. and Ernst,R.R. (1987) Frequency offset effects and their elimination in NMR rotating-frame cross-relaxation spectroscopy. *J. Magn. Reson.*, **75**, 261–271.

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.

Inoue,K., Hoshijima,K., Sakamoto,H. and Shimura,Y. (1990) Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript. *Nature*, **344**, 461–463.

Inoue,M., Muto,Y., Sakamoto,H., Kigawa,T., Takio,K., Shimura,Y. and Yokoyama,S. (1997) A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino-terminal RNA-binding domain of *Drosophila* Sex-lethal. *J. Mol. Biol.*, **272**, 82–94.

Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) Investigation of exchange process by two-dimensional NMR spectroscopy. *J. Chem. Phys.*, **71**, 4546–4553.

Johnson,B.A. and Blevins,R.A. (1994) NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.

Kanaar,R., Roche,S.E., Beall,E.L., Green,M.R. and Rio,D.C. (1993) The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. *Science*, **262**, 569–573.

Kigawa,T., Muto,Y. and Yokoyama,S. (1995) Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR*, **6**, 129–134.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., Rullmann,J.A.C., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.

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.

Lee,A.L., Volkman,B.F., Robertson,S.A., Rudner,D.Z., Barbash,D.A., Cline,T.W., Kanaar,R., Rio,D.C. and Wemmer,D.E. (1997) Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein Sex-lethal. *Biochemistry*, **36**, 14306–14317.

Marion,D., Driscoll,P.C., Kay,L.E., Wingfield,P.T., Bax,A., Gronenborn, A.M. and Clore,G.M. (1989a) Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H–^{15}N Hartmann–Hahn–multiple quantum coherence and nuclear overhauser–multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry*, **28**, 6150–6156.

Marion,D., Kay,L.E., Sparks,S.W., Torchia,D.A. and Bax,A. (1989b) Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. Chem. Soc.*, **111**, 1515–1517.

Merritt,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

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.

Nagata,T., Kurihara,Y., Matsuda,G., Saeki,J.-i., Kohno,T., Yanagida,Y., Ishikawa,F., Uesugi,S. and Katahira,M. (1999a) Structure and interactions with RNA of the N-terminal UUAG-specific RNA-binding domain of hnRNP D0. *J. Mol. Biol.*, **287**, 221–237.

Nagata,T., Kahno,R., Kurihara,Y., Uesugi,S., Imai,T., Sakakibara,S.-i., Okano,H. and Katahira,M. (1999b) Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1. *J. Mol. Biol.*, **287**, 315–330.

Nilges,M. (1993) A calculation strategy for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins: Struct. Funct. Genet.*, **17**, 297–309.

Nilges,M., Clore,M. and Gronenborn,A.M. (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS Lett.*, **239**, 129–136.

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

Potashkin,J., Naik,K. and Wentz-Hunter,K. (1993) U2AF homolog required for splicing *in vivo*. *Nature*, **262**, 573–575.

Price,S.R., Evans,P.R. and Nagai,K. (1998) Crystal structure of the spliceosomal U2B″–U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645–650.

Rance,M., Sørensen,O.W., Bodenhausen,G., Wagner,G., Ernst,R.R. and Wüthrich,K. (1983) Improved spectral resolution in COSY ^{1}H NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Ruskin,B., Zamore,P.D. and Green,M.R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*, **52**, 207–219.

Sailer,A., MacDonald,N.J. and Weissmann,C. (1992) Cloning and sequencing of the murine homologue of the human splicing factor U2AF^{65}. *Nucleic Acids Res.*, **20**, 2374.

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. *Nature Struct. Biol.*, **4**, 215–222.

Sillekens,P.T., Habets,W.J., Beijer,R.P. and van Venrooij,W.J. (1987) cDNA cloning of the human U1 snRNA-associated protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.*, **6**, 3841–3848.

Singh,R., Valcárcel,J. and Green,M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, **268**, 1173–1176.

Valcárcel,J., Singh,R., Zamore,P.D. and Green,M.R. (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA. *Nature*, **362**, 171–175.

Vuister,G.W. and Bax,A. (1993) Quantitative *J* correlation: a new approach for measuring homonuclear three-bond *J* ($H^{N}H^{\alpha}$) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.*, **115**, 7772–7777.

Wincott,F. *et al*. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.*, **23**, 2677–2684.

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

Wüthrich,K., Billeter,M. and Braun,W. (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton–proton distance constraints with nuclear magnetic resonance. *J. Mol. Biol.*, **169**, 949–961.

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.

Zamore,P.D. and Green,M.R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl Acad. Sci. USA*, **86**, 9243–9247.

Zamore,P.D. and Green,M.R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.*, **10**, 207–214.

Zamore,P.D., Patton,J.G. and Green,M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**, 609–614.

Zhang,M., Zamore,P.D., Carmo-Fonseca,M., Lamond,A.I. and Green, M.R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl Acad. Sci. USA*, **89**, 8769–8773.

Zorio,D.A.R., Lea,K. and Blumenthal,T. (1997) Cloning of *Caenorhabditis* U2AF^{65}: an alternatively spliced RNA containing a novel exon. *Mol. Cell. Biol.*, **17**, 946–953.

Received March 15, 1999; revised June 11, 1999;
accepted June 21, 1999