

# A novel occluded RNA recognition motif in Prp24 unwinds the U6 RNA internal stem loop

Stephen Martin-Tumas<sup>1</sup>, Ashley C. Richie<sup>1</sup>, Lawrence J. Clos II<sup>1</sup>, David A. Brow<sup>2,\*</sup> and Samuel E. Butcher<sup>1,\*</sup>

<sup>1</sup>Department of Biochemistry and <sup>2</sup>Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA

Received April 21, 2011; Revised May 16, 2011; Accepted May 17, 2011

## ABSTRACT

The essential splicing factor Prp24 contains four RNA Recognition Motif (RRM) domains, and functions to anneal U6 and U4 RNAs during spliceosome assembly. Here, we report the structure and characterization of the C-terminal RRM4. This domain adopts a novel non-canonical RRM fold with two additional flanking  $\alpha$ -helices that occlude its  $\beta$ -sheet face, forming an occluded RRM (oRRM) domain. The flanking helices form a large electro-positive surface. oRRM4 binds to and unwinds the U6 internal stem loop (U6 ISL), a stable helix that must be unwound during U4/U6 assembly. NMR data indicate that the process starts with the terminal base pairs of the helix and proceeds toward the loop. We propose a mechanistic and structural model of Prp24's annealing activity in which oRRM4 functions to destabilize the U6 ISL during U4/U6 assembly.

## INTRODUCTION

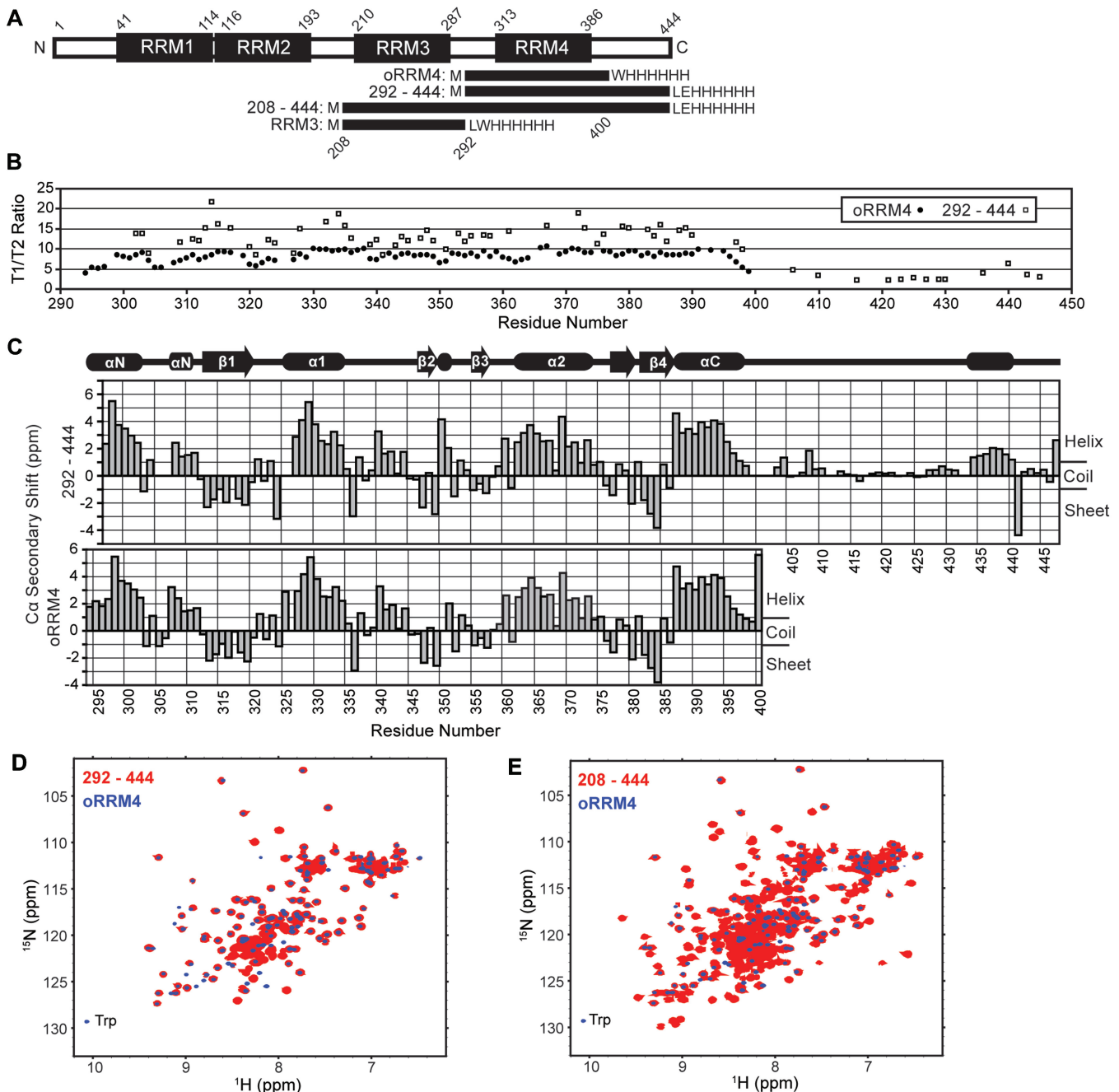
Proper gene expression in eukaryotes requires production of functional messenger RNA (mRNA). Through the essential process of pre-mRNA splicing, introns are removed from pre-mRNA, and exons are joined together. Many proteins and RNAs are required for splicing. Key to splicing are five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs), each of which combines with multiple proteins to form a small nuclear ribonucleoprotein (snRNP) (1). The five snRNPs and additional proteins form a dynamic complex known as the spliceosome that catalyzes splicing. During splicing, many proteins enter or leave the spliceosome, and both the pre-mRNA and the snRNAs undergo essential dynamic rearrangements.

*Saccharomyces cerevisiae* Prp24 [known as p110 or Sart3 in humans (2)] is an essential component of the U6 snRNP (3–9). For U6 to enter the splicing cycle, its internal stem loop (ISL) must be unwound, and U6 must base pair to U4 RNA to form the U4/U6 di-snRNP (10,11). Extensive evidence demonstrates that Prp24 greatly accelerates annealing of the U4 and U6 RNAs, but the mechanism by which it does so is as yet unclear (3–8,12,13). In addition to Prp24, the U6 snRNP includes a 7-membered ring of Lsm proteins (Lsm2–8) (14). Prp24 binds both the U6 RNA proximal to the ISL and the Lsm ring (5,6,8,15–17). The Lsm ring binds Prp24 and the uridine rich 3'-end of U6 RNA (8,12,18).

Prp24 is a 51 kDa protein containing three known and one predicted RNA recognition motif (RRM) domains (Figure 1A) (3,8,16). RRMs are common protein motifs that typically bind single-stranded RNA along a four stranded antiparallel  $\beta$ -sheet through two conserved RNP motifs enriched in aromatic residues (19,20). Previous work has demonstrated that the N-terminal 40 amino acids of Prp24 are unstructured and that the first three RRMs fold canonically (16). The first two RRMs pack together tightly and form a single RNA binding surface, with RRM2 binding U6 RNA canonically and RRM1 binding non-canonically through an electropositive surface (16,21). In contrast, RRM3 makes no stable contacts with RRM2 in solution (21), and its RNA binding behavior is unknown.

While the extreme C-terminal region of Prp24 contains a conserved Lsm interaction site (referred to as the C-terminal motif or SNFFL box) (8,15), little else is known about the structure or function of residues C-terminal of RRM3. The RNP motifs of the fourth RRM are so divergent from the canonical sequences that it was only identified as an RRM through comparison with fungal homologs and secondary structure prediction (3,8). RRM4 is functionally important, as a triple alanine mutation of RNP1 positions 1, 3 and 5 results in a

\*To whom correspondence should be addressed. Tel: +608 262 1475; Fax: +608 262 5253; Email: dabrow@wisc.edu  
Correspondence may also be addressed to Samuel E. Butcher. Tel: +608 263 3890; Fax: +608 262 3453; Email: butcher@biochem.wisc.edu



**Figure 1.** Structured regions in the C-terminal portion of Prp24. **(A)** Domains in Prp24. Black boxes indicate the three known and one predicted RRM. The four lower black bars represent the protein constructs used in this study. Non-native sequence is shown as text. **(B)** NMR T1/T2 relaxation data of 292–444 and oRRM4. The higher variability of the T1/T2 ratios for 292–444 likely results from decreased spectral quality and difficulties in precisely measuring the shorter T2 relaxation times for the larger protein. **(C)** C $\alpha$  secondary shift data for 292–444 and oRRM4. The predicted secondary structure is shown across the top.  $\alpha$ N and  $\alpha$ C indicate the non-canonical helices. **(D)** Overlay of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of 292–444 (red) and oRRM4 (blue). The 292–444 spectrum has wider contour lines than the oRRM4 spectrum to enhance its visibility. **(E)** Overlay of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of 208–444 (red) and oRRM4 (blue). The 208–444 has been given wider contour lines. The additional oRRM4 peak at 10 ppm arises from the non-native tryptophan residue.

temperature sensitive phenotype (8), and deletion of RRM4 (residues 317–391) is lethal even when Prp24 is expressed at or above wild-type levels (S.S. Kwan and D.A.B., unpublished data).

Here, we present the structural and functional characterization of the C-terminal region of Prp24. While

the extreme C-terminus is unstructured, there is a 108 residue structured region centered on the predicted fourth RRM. This domain is a non-canonical RRM in which the expected  $\beta\alpha\beta\beta\alpha\beta$  RRM-fold is flanked by N- and C-terminal  $\alpha$ -helices. These  $\alpha$ -helices are rigidly bound to the  $\beta$ -sheet face of the RRM fold, forming

a domain, which we refer to as an occluded RRM (oRRM). A substantial electropositive surface is present on the solvent exposed sides of the flanking  $\alpha$ -helices, which may constitute an alternative RNA binding surface. Indeed, oRRM4 is capable of binding and, unexpectedly, unwinding the U6 ISL *in vitro*. We present a structural model for the complex between Prp24 RRM3 1–4 and the ISL region of U6 RNA, which suggests a mechanism for the nucleation of U4/U6 pairing.

## MATERIALS AND METHODS

### Construct design

All constructs were derived from an *Escherichia coli*-based pET21b expression vector containing Prp24 208–444 prepared as described (15). 292–444, RRM3 and oRRM4 were obtained by truncation through a modification of the QuikChange™ Site-Directed Mutagenesis protocol (22,23). The additional tryptophan in RRM3 and oRRM4 (Figure 1A) was incorporated to facilitate spectrophotometric quantification. Reactions contained 1× Pfu Turbo Buffer (Agilent Technologies), 0.2 mM dNTPs (Agilent Technologies), 2  $\mu$ M forward or reverse primer (Integrated DNA Technologies), 100 ng template plasmid and 2U Pfu Turbo DNA polymerase (Agilent Technologies) in a 50  $\mu$ l volume. Reactions were incubated for 3 min at 94°C in a thermocycler (Eppendorf Mastercycler), then four cycles of 1 min 94°C, 1 min 52°C, 12 min 68°C. Twenty-five micro litre of each reaction were combined with an additional 2U Pfu Turbo and subjected to the same thermocycler program, except with 15 cycles and a final 1 h 68°C incubation. The remainder followed the manufacturer's protocols, using XL2-Blue Ultracompetent cells (Agilent Technologies) for the transformation.

### Protein and RNA preparation

All proteins were prepared essentially as described (24), with the following changes. The HisPur cobalt spin column (Pierce Biotechnology) elution buffer used 50 mM potassium phosphate pH 6 instead of 50 mM sodium phosphate pH 7.4. NMR Buffer contained 20 mM potassium phosphate pH 6, 50 mM potassium chloride and 1 mM dithiothreitol. Gel filtration was performed in NMR Buffer, and buffer changes were performed through buffer exchange in a centrifugal filter device (Millipore Amicon, 3 kDa cutoff).

The U6 nucleotide 49–88 and nucleotide 58–97 RNAs were prepared through *in vitro* transcription using purified His<sub>6</sub>-tagged T7 RNA polymerase. Synthetic DNA oligonucleotides (Integrated DNA Technologies) containing either the T7 promoter sequence, two non-natural G nucleotides and the appropriate U6 RNA sequence or the reverse complement were used (Supplementary Figure S1A and B). Transcribed RNA was purified using denaturing 10% polyacrylamide gel electrophoresis and identified through UV shadowing. Gel fragments containing RNA were excised, and the RNA was diffused out of the gel into 0.3 M sodium acetate pH 5.0 and precipitated with cold ethanol. RNA was purified by

anion exchange chromatography (Bio-Rad Bio-Scale Mini Macro-Prep High Q Cartridge on a BioLogic LP chromatography unit), then buffer exchanged into NMR Buffer.

Fluorescently labeled RNAs were purchased from Dharmacon (U2) or IDT (U4 and U6), and prepared according to the manufacturer's protocols. U2: 5'-FITC-ACGAAUCUCUUUGCCUUUUGGCUUAGAUC AAGU GUAGUAUCUGUUCUUUUUC-3'; U4: 5'-Cy3-AUCCU UAUGCACGGGAAAUUUUGCUGGUU and U6: 5'-Cy5-AGAGAUGAUCAGCAGU UCCCCUGCAUA AGGAUGAACCGUU-3'.

### NMR resonance assignments

All NMR experiments, unless stated otherwise, were acquired at 25°C in 90% NMR Buffer/10% D<sub>2</sub>O using 600  $\mu$ M <sup>13</sup>C/<sup>15</sup>N labeled protein. Standard 2D and 3D spectra of oRRM4 were acquired as follows: 2D <sup>1</sup>H, <sup>15</sup>N-HSQC, 3D CBCACONH, 3D NOESY-<sup>1</sup>H, <sup>15</sup>N-HSQC, 3D HNCACB and 3D HNCO on a Varian 900 MHz spectrometer; 3D HBHACONH, 3D CCONH and 3D HCCONH on a Varian 600 MHz spectrometer; 2D aliphatic and aromatic <sup>1</sup>H, <sup>13</sup>C-HSQC, 2D NOESY (unlabeled protein), 3D HCCH-TOCSY and 3D aliphatic and aromatic NOESY-<sup>1</sup>H, <sup>13</sup>C-HSQC on a Varian 900 MHz spectrometer in 99% deuterated NMR Buffer. Deuterated NMR Buffer was obtained by lyophilizing NMR Buffer to dryness and resuspending in an equal volume of D<sub>2</sub>O three times.

Resonance assignments of RRM3 and 292–444 were obtained using 2D <sup>1</sup>H, <sup>15</sup>N-HSQC, 3D CBCACONH, 3D HNCACB, 3D HBHACONH and 3D HNCO spectra, all acquired on a Varian 900 MHz spectrometer in 90% NMR Buffer/10% D<sub>2</sub>O.

Spectra were processed using nmrPipe (25) and analyzed using Sparky (University of California San Francisco).

### NMR relaxation measurements

<sup>15</sup>N NMR relaxation experiments of 292–444 and oRRM4 were acquired on a 600 MHz Varian spectrometer. Longitudinal (T<sub>1</sub>) experiments had relaxation delays of 0.1, 0.3, 0.5, 0.7, 0.9 or 1.1 s; transverse (T<sub>2</sub>) experiments had relaxation delays of 0.01, 0.03, 0.05, 0.07, 0.09 or 0.13 s. Spectra were processed using nmrPipe (25), and relaxation curves fit using the Sparky relaxation peak heights tool.

### RNP sequence alignment and statistical analysis

Fifty RRMs of known structure and function were identified from the PDB by searching for 'RRM'. Their RNP motifs were manually aligned based on sequence and structure (Supplementary Table S1). They were classified as follows: canonical RRMs are known to bind nucleic acid along their  $\beta$ -sheet face, either through presence of nucleic acid in the structure or through other evidence presented in the 'Primary Citation' linked to the structure in the PDB (i.e. NMR chemical shift perturbation, mutations, etc.); all other RRMs are non-canonical. Quasi-RRM domains were excluded from the analysis.



Thirty-four ‘canonical’ RRM4s and 16 ‘non-canonical’ RRM4s (including oRRM4) were identified. Each sequence was checked for conservation at every RNP position. Note that tryptophan was accepted at conserved F/Y positions, and methionine at conserved I/L/V positions. The data were used to construct a contingency table for each position, showing the number of canonical and non-canonical RRM4s, which either have or do not have sequence conservation at that position. A two-sided Fisher’s exact chi-square test (GraphPad Prism) was used to calculate *P*-values.

### Residual dipolar coupling experiments

RDC values for  $^{13}\text{C}^{15}\text{N}$  labeled oRRM4 were obtained and analyzed as described (21).

### Structure determination

Initial structures of oRRM4 were obtained using the ATNOS-CANDID module of Unio 08 (26–28). Twenty structures were calculated based on the assigned peak list, the 3D NOESY spectra, and dihedral angle and hydrogen bond restraints derived from TALOS<sup>+</sup> (29). Ultimately, 1289 NOE distance restraints were obtained. Ninety additional NOE distance restraints were manually identified. NOEs from ATNOS-CANDID were restrained as 1.8 Å lower bound to 0.5 Å above the assigned distance upper bound, while manual NOEs were restrained as 1.8–6 Å. All 20 ATNOS-CANDID structures underwent water refinement using CNS (30) under the HADDOCK 2.0 interface (31,32). As no structure had any distance restraint violation >0.5 Å, all 20 structures were aligned and had their pairwise RMSD calculated using an in-house script. Structure statistics are shown in Table 1. Structure statistics and buried surface area values were obtained from HADDOCK output files, except for RDC Q factors, which were obtained from PALES (33). Ramachandran plot statistics for all 20 structures (based on the PDB validation suite) are: most favored 78.4%, additional allowed 19.7%, generously allowed 1.7% and disallowed 0.2%.

The structural model of Prp24 bound to U6 RNA was calculated starting from structures of Prp24 residues 41–399 and U6 nucleotides 49–91. HADDOCK 2.0 (31,32) was used to dock Prp24 and U6. However, HADDOCK does not typically allow the large rearrangements necessary to bring RRM3 and oRRM4 into contact with U6. Therefore, the lowest energy HADDOCK structure was refined in XPLOR-NIH 2.21 (34) to produce the final model. In order to maintain the structures of the individual domains, all distance and dihedral angle restraints from the NMR structures of oRRM4, RRM3 1 and 2, and the extended U6 ISL (BMRB IDs: 17490, 7070 and 6320) were included in the calculations, along with distance restraints generated from the crystal structure of RRM3 (PDB ID: 2GHP). Missing residues from the crystal structure (197–205 and 248–250) were manually added using the ‘build residue’ and ‘sculpting’ functions in PyMol (DeLano Scientific). oRRM4 was added by using the ‘translate’ and ‘create bond’ functions in PyMol to join the NMR structure of oRRM4 to the crystal structure of

**Table 1.** NMR structure determination statistics

NMR distance and angle constraints	
Distance constraints	1432
Total NOE	1379
Intraresidue	460
Interresidue	919
Sequential ( $ i-j  = 1$ )	363
Medium range ( $ i-j  < 4$ )	210
Long range ( $ i-j  > 3$ )	346
Hydrogen bonds	53
Total dihedral angle restraints	177
$\phi$	92
$\psi$	85
Total residual dipolar coupling restraints	120
NH	79
CH	41
Structure statistics	
Violations (mean and SD)	
Distance constraints	0 ± 0
Dihedral angle constraints	2.8 ± 1.4
Max. dihedral angle violation (°)	12.8
Max. distance constraint violation (Å)	0.34
Q-factor (%)	16 ± 1
Deviations from idealized geometry	
Bond lengths (Å)	0.0064
Bond angles (°)	0.94
Improper (°)	1.1
Average pairwise r.m.s. deviation (20 structures) (Å)	
Heavy	1.78
Backbone	0.95

RRM3 1–3. U6 RNA nucleotide 49–91 was generated as described (21). Intermolecular distance restraints were derived from the previous model of RRM3 1 and 2 bound to U6 RNA (21), chemical shift perturbations on RRM3 induced by U6 RNA (Supplementary Figure S1) and perturbations on the U6 ISL induced by oRRM4 (Figure 5D). Because RRM3 appears to bind the ISL, and oRRM4 was shown to preferentially destabilize the lower ISL, RRM3 was restrained to bind the upper ISL.

Structure figures were prepared using PyMol, and electrostatic surfaces were calculated using APBS (35) displaying the solvent accessible surface from  $-4\text{ kT}$  to  $4\text{ kT}$ .

### Fluorescence anisotropy RNA binding assay

RNA of 10 nM in Assay Solution (150 mM potassium chloride, 1 mM magnesium chloride, pH 6) was titrated with increasing amounts of oRRM4 as shown in Figure 5B. Fluorescence polarization measurements were taken in a Varian Cary Eclipse spectrofluorimeter fitted with automated polarizers. At each concentration of protein, anisotropy was measured five times using a 10 s averaging time. The mean and standard deviations for three independent experiments were fit using the following one-site equation:

$$Y = Y_f + \left[ (Y_o - Y_f) \left( \frac{X}{K_d + X} \right)^n \right],$$

where  $Y_f$  is the final anisotropy,  $Y_o$  is the initial anisotropy,  $X$  is the protein concentration,  $K_d$  is the apparent dissociation constant and  $n$  is the Hill Coefficient.



$K_d$  and  $n$  values were determined through non-linear least squares fitting using GraphPad Prism 4.0. Values for the Hill coefficient were all between 0.9 and 1.0 suggesting no cooperativity. Anisotropy values were normalized to between 0 and  $\pm 1$ .

#### UV-monitored RNA unwinding assay

All measurements were acquired at 25°C in Assay Solution. U6 RNA of 2  $\mu$ M was titrated with increasing amounts of protein, as indicated in Figure 5C. For oRRM4+U6 RNA, the final data point was obtained by adding solid sodium chloride to each sample to bring them to 1 M NaCl. Absorbance at 260 nm was followed in triplicate for the following samples for each protein tested: A<sub>1</sub>, U6 RNA+protein; A<sub>2</sub>, Assay Solution+protein (control for protein absorbance); A<sub>3</sub>, U6 RNA+mock protein addition (control for dilution effects) and A<sub>4</sub>, Assay Solution+mock protein addition (blank for A<sub>3</sub>). Mock protein addition entailed addition of an equal volume of Assay Solution instead of protein. All absorbance values were baseline corrected using the no protein value. Changes in absorbance were calculated as:

$$\Delta A_{260} = (A_1 - A_2) - (A_3 - A_4),$$

where the first term reflects the absorbance of U6 in the presence of protein, and the second term reflects the absorbance of U6 in the absence of protein. The range of concentrations tested was limited by the relatively poor solubility of the protein/RNA complex. Denaturing polyacrylamide gel electrophoresis analysis of the RNA before and after addition of protein showed no degradation (data not shown).

#### NMR-monitored RNA unwinding assay

One dimensional proton with flip-back water suppression NMR spectra were acquired of a 10  $\mu$ M sample of U6 RNA in 90% NMR Buffer/10% D<sub>2</sub>O with 0  $\mu$ M, 10  $\mu$ M or 20  $\mu$ M of oRRM4 on a Bruker 500 MHz spectrometer at 25°C. Due to the low sample concentration, 6144 scans were necessary (~3.5h). In order to accentuate relative differences and control for signal loss due to partial aggregation, signal intensity was normalized to the peak for nucleotides 77/78, which showed the least absolute change. Spectra were analyzed and processed using TopSpin (Bruker). Resonance assignments were based on BMRB ID: 6320.

#### Chemical shift perturbation

Chemical shift perturbation of RRM3 was performed as described previously (21) using 167  $\mu$ M <sup>15</sup>N RRM3 in 90% NMR Buffer/10% D<sub>2</sub>O on a 900 MHz spectrometer at 25°C. For U6 nucleotide 49–88, either 0  $\mu$ M or 333  $\mu$ M RNA was used. For U6 nucleotide 58–97 either 0  $\mu$ M or 15  $\mu$ M RNA was used. The putative binding site was identified as residues with a change in peak position of >0.08 ppm when U6 nucleotide 49–88 RNA was added.

## RESULTS

### Structural characterization of the Prp24 C-terminal region

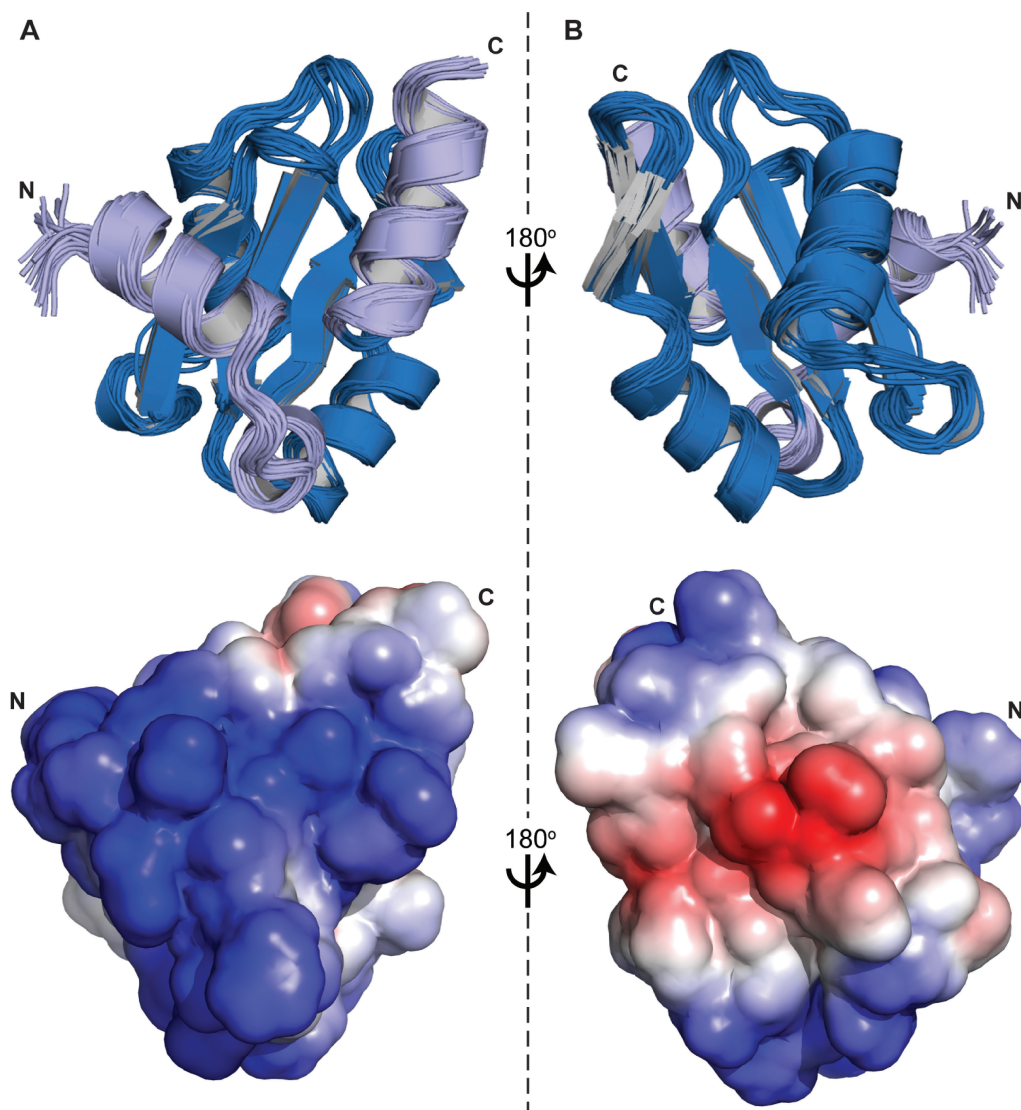
The previously determined crystal structure of Prp24 included residues 1–291, covering the first three RRM domains (16). We investigated the structure of the remaining C-terminal region of Prp24 corresponding to residues 292–444 (Figure 1A). NMR relaxation times were measured in order to identify which regions are structured in solution (Figure 1B). There is a highly structured region overlapping the predicted fourth RRM (residues 302–398, average T1/T2 of 13.3), while the C-terminal tail is unstructured (residues 406–end, average T1/T2 of 3.2). The secondary structure of the C-terminal region was predicted based on C $\alpha$  secondary shifts (Figure 1C) (36). In addition to a canonical  $\beta\alpha\beta\beta\alpha\beta$  RRM-fold (residues 312–387), additional flanking  $\alpha$ -helices are predicted (residues 297–302, 308–311 and 388–397). An  $\alpha$ -helix is also predicted in the far C-terminal region (residues 434–442, overlapping the conserved SNFFL box), but subsequent NMR investigation was unable to confirm the existence of this helix (data not shown).

Because the large unstructured C-terminal tail results in extensive spectral overlap, a shorter construct incorporating only the structured regions was studied (oRRM4, Figure 1A). NMR relaxation time measurements showed that the entire sequence, with the exception of the extreme termini, is structured in solution (average T1/T2 of 9.4, Figure 1B). Both the C $\alpha$  secondary shifts (Figure 1C) and <sup>1</sup>H-<sup>15</sup>N-HSQC spectra (Figure 1D) of oRRM4 are consistent with the larger 292–444 construct, demonstrating that the presence of the C-terminal tail does not significantly affect the structured region. We also investigated potential interdomain contacts between oRRM4 and RRM3, as Prp24's first two RRMs are known to make extensive interdomain contacts (16,21). The chemical shifts of the isolated oRRM4 do not change in a construct containing RRM3 (Figure 1E), indicating that no stable contacts occur between oRRM4 and RRM3.

### oRRM4 is a non-canonical RRM

The solution structure of oRRM4 was determined by NMR (Table 1). It displays the  $\beta\alpha\beta\beta\alpha\beta$  RRM fold, but with N-terminal and C-terminal  $\alpha$ -helices occluding the  $\beta$ -sheet face (Figure 2A), consistent with the secondary structure predictions (Figure 1C). The solvent exposed faces of the flanking  $\alpha$ -helices are rich in basic residues, and therefore could serve as a non-canonical RNA binding site. Together, they form a substantial electropositive surface (Figure 2A). As the opposite side of oRRM4 is neutral to electronegative (Figure 2B), any RNA binding by oRRM4 would likely be through the electropositive patch on the N- and C-terminal  $\alpha$ -helices.

The N- and C-terminal  $\alpha$ -helices appear to be rigidly attached to the  $\beta$ -sheet face of oRRM4, as evidenced by T1/T2 ratios (Figure 1B). An extensive hydrophobic core forms between the flanking  $\alpha$ -helices and the  $\beta$ -sheet face of oRRM4 (Figure 3A and B). The majority



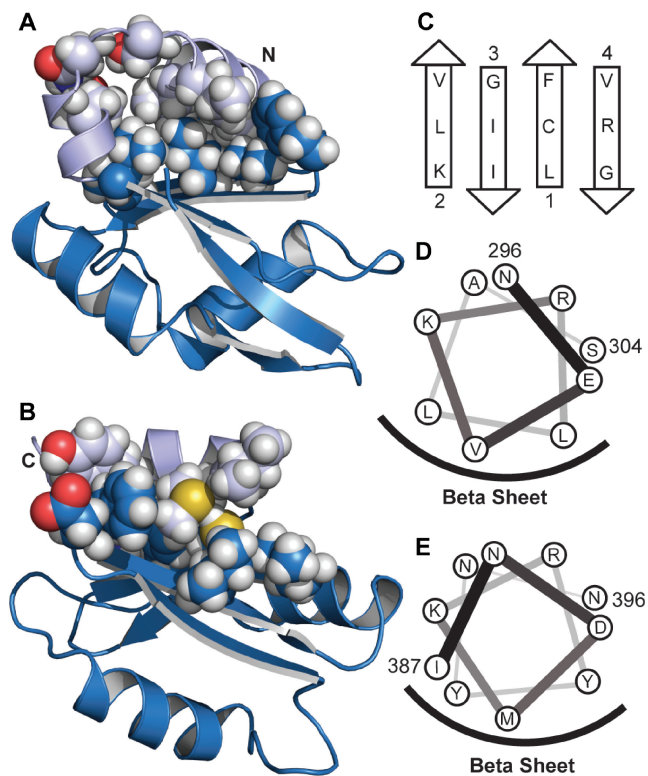
**Figure 2.** The solution structure of oRRM4. (A) Overlay of the 20 calculated structures, with the core RRM colored dark blue and the N-terminal and C-terminal helices colored light blue. The solvent accessible electrostatic surface is shown below, with blue representing electropositive and red electronegative. (B) As in (A), but rotated 180° about the vertical axis.

of the residues on oRRM4's  $\beta$ -sheet face that would be exposed in a canonical RRM are hydrophobic (Figure 3C). In addition, the flanking  $\alpha$ -helices are amphipathic with a hydrophobic face oriented toward the  $\beta$ -sheet and a hydrophilic face oriented toward solution (Figure 3D and E). Approximately 1000 Å<sup>2</sup> of surface area on the  $\beta$ -sheet face is buried by the flanking helices consistent with an extensive and stable interaction.

The structure of oRRM4 reveals why it lacks all of the canonical RNA binding side chains in the RNP1 ( $\beta$ 3) and RNP2 ( $\beta$ 1) motifs. The canonical basic and aromatic residues are instead hydrophobic residues that contribute to the binding surface for the flanking  $\alpha$ -helices. In contrast, RNP1 and RNP2 motif residues that pack internally to stabilize the overall fold of oRRM4 match the RRM consensus. To see if this conservation pattern is a general feature of non-canonical RRMs, we identified 50 RRMs of known structure and RNA binding activity

from the Protein Data Bank (PDB) (Supplementary Table S1), and classified them as canonical ( $n = 34$ ) or non-canonical ( $n = 16$ ) based on whether or not they are reported to bind nucleic acid along their  $\beta$ -sheet face. Statistically significant deviations from the consensus identity in the RNP1 and RNP2 motifs of non-canonical RRMs were observed only at the four canonical RNA binding positions (Figure 4).

The results of our analysis suggest that RRMs that deviate from the consensus at these four positions are unlikely to bind RNA in a canonical fashion. In these cases, the RRM fold is maintained for some other purpose, such as non-canonical RNA binding or a protein interaction. However, such a conclusion is not definitive, since there are examples of RRMs binding RNA canonically despite conserving only one or none of the RNA binding positions [Polypyrimidine Tract Binding Protein RRM2-4 (37)], and of RRMs conserving three

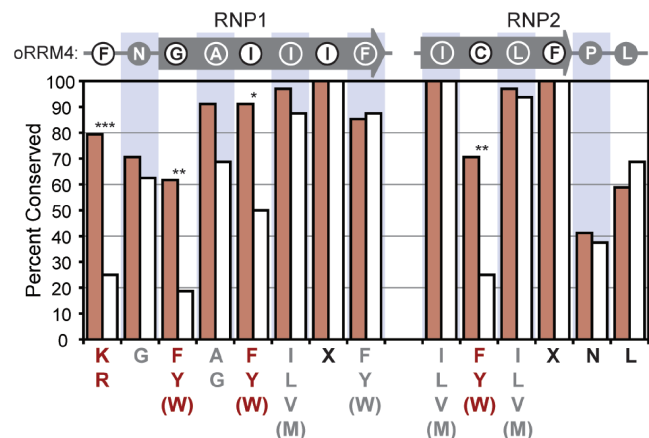


**Figure 3.** The N- and C-terminal  $\alpha$ -helices are rigidly attached to the  $\beta$ -sheet face of oRRM4. (A) Buried surface area between the  $\beta$ -sheet of oRRM4 and the N-terminal helices. The C-terminal helix was removed for clarity. Residues that pack against each other are highlighted with spheres. Atoms are colored as follows: hydrogen white, carbon blue (same shade as the associated domain), oxygen red, nitrogen dark blue and sulfur yellow. (B) As in (A), but showing the C-terminal helix. The N-terminal helices were removed for clarity. (C) Residues on the  $\beta$ -sheet face of oRRM4, which would be solvent exposed in the absence of the flanking helices. (D) Helical wheel representation of the larger N-terminal  $\alpha$ -helix. (E) As in (D), but for the C-terminal helix.

of the RNA binding positions and not binding RNA canonically [Prp24-RRM1 (16), La-RRM1 (38), Y14 (39) and p14 (40)].

#### oRRM4 binds RNA and unwinds the U6 ISL

Despite the occlusion of the canonical  $\beta$ -sheet RNA binding site, oRRM4's large electropositive surface suggests that it may be capable of RNA binding. To investigate this, a fluorescence anisotropy assay was used to determine its binding affinity for segments of three different yeast spliceosomal RNAs. RNAs containing the U4/U6 base pairing sequence from U4 (nucleotide 1–18 + 54–64 + 5'-Cy3) and U6 (nucleotide 49–88 + 5'-Cy5) were used (Figure 5A and Supplementary Figure S2). In addition, the region of U2 that base pairs with U6 (nucleotide 1–51 + 5'-fluorescein isothiocyanate) was assayed. oRRM4 is capable of binding all three RNAs with an apparent dissociation constant ( $K_d$ ) ranging between 3.1  $\mu$ M for U6 and 7.1  $\mu$ M for U4 (Figure 5B). The similarity of the  $K_d$ 's for three different RNAs suggest that RNA binding is primarily non-specific, consistent



**Figure 4.** Sequence conservation of the RNP1 and RNP2 motifs from 50 RRM domains (see also Supplementary Table S1). oRRM4's RNP motifs are indicated at the top, with solvent exposed side chains as black text and internal side chains as white text. The canonical RNP sequences are at the bottom: canonical RNA binding residues (red), internal side chains (gray), solvent exposed or off the  $\beta$ -sheet (black). Parentheses indicate residues that are not in the canonical motif, but were counted as canonical in this analysis. RRMs classified as canonical RNA binding are shown as red rectangles; non-canonical as white rectangles. Positions with statistical differences between canonical and non-canonical RRMs based on two-sided Fisher's exact chi-square test are indicated as \* $P < 0.05$ , \*\* $P < 0.005$ , \*\*\* $P < 0.0005$ .

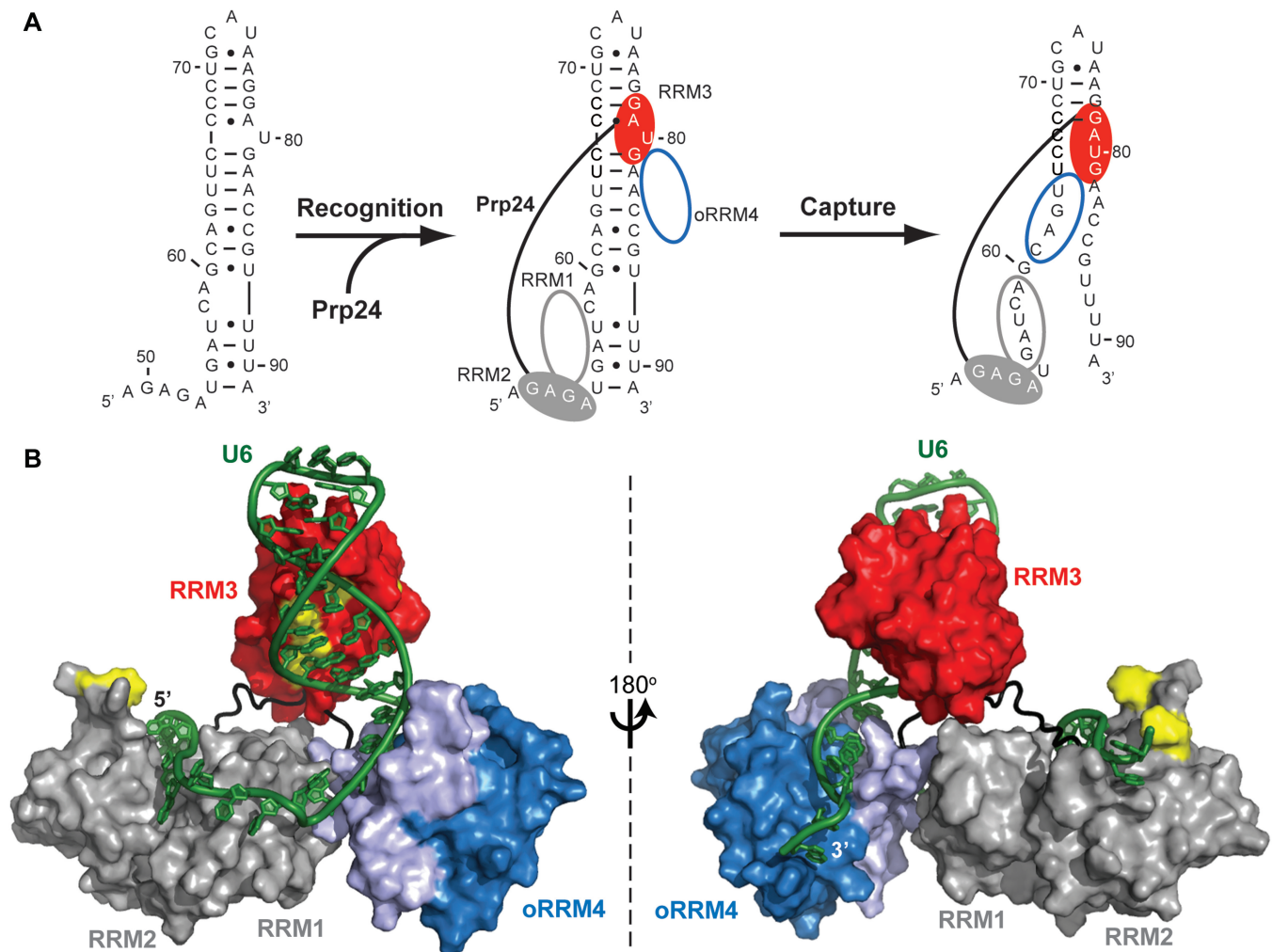
with oRRM4 binding RNA predominantly through its electropositive surface.

A surprising result of the binding studies is that oRRM4 binding to U6 RNA causes a decrease rather than the expected increase in fluorescence anisotropy (Figure 5B). Our U6 construct forms the U6 ISL (nucleotide 59–88, Figure 5A), and previous work has found that Cy5 can bind nucleic acid helices either at a terminal base pair (41) or in the major groove (42). The decrease in anisotropy upon oRRM4 binding indicates that the fluorophore is able to rotate more freely when the protein is bound. There are two possible explanations for this behavior: oRRM4 binding either displaces the fluorophore from the intact U6 ISL, or disrupts the U6 ISL so that the fluorophore can no longer interact with it.

To discriminate between these two possibilities, we assayed the effect of oRRM4 binding on U6 RNA (nucleotide 49–88) secondary structure by monitoring the UV absorbance at 260 nm. While adding bovine serum albumin (BSA) or RRM3 had no effect on the UV absorbance of U6 RNA, adding oRRM4 caused a dramatic increase in UV absorbance (Figure 5C), consistent with unwinding of the ISL. This large change in absorbance is similar to that observed for the heat-induced melting of U6 RNA (data not shown). Adding sodium chloride to 1 M to the oRRM4/U6 complex returns the UV absorbance of the sample to that of folded U6 RNA in the absence of oRRM4 (Figure 5C). This result indicates that the protein-induced destabilization of the ISL is reversible. It is also consistent with an electrostatic interaction between the protein and RNA that is disrupted at high-ionic strength. Attempts at mapping the interaction







**Figure 6.** Modeling the Prp24-U6 complex. (A) Schematic model showing a potential mechanism by which Prp24 could recognize and unwind the U6 ISL. (B) Structural model of a complex between Prp24 and U6 RNA (nucleotides 49–91). Colors as in (A), except U6 is now green and the electropositive helices on oRRM4 are light blue. Yellow indicates sites of U6 A62G suppressor mutations (7). This complex corresponds to the final stage shown in (A).

neutral in charge. Finally, two of the qRRMs in *H. sapiens* hnRNP F have C-terminal helices, but they are neither electropositive nor involved in RNA binding (49).

Our data show that RRM3 is capable of binding the U6 ISL (Supplementary Figure S1). Analysis of RRM3 binding to partially randomized libraries of single-stranded RNA using Scaffold Independent Analysis (50) followed by NMR titration experiments indicated that RRM3 binds non-specifically to single-stranded RNA with poor affinity ( $K_d \sim 1$  mM) (data not shown). While some residues on the  $\beta$ -sheet of RRM3 are perturbed in the presence of the U6 ISL, the largest changes occur in the  $\beta$ 2- $\beta$ 3 loop (Supplementary Figure S1), showing the interaction between RRM3 and the ISL is at least partially non-canonical. Interestingly, a previous structure of an RRM binding double stranded RNA (*H. sapiens* RBMY) also observed extensive interactions between the  $\beta$ 2- $\beta$ 3 loop and RNA, coupled with a canonical interaction on the  $\beta$ -sheet (51).

We previously proposed that *S. cerevisiae* Prp24 is composed of two ‘matchmaker domains’, each of which

binds a site on U6 RNA through one RRM responsible for sequence specific recognition and a second RRM, which interacts electrostatically to facilitate unwinding and/or annealing of RNA helices (21). The N-terminal half of Prp24 appears to fit this description—RRM2 binds U6 RNA sequence specifically, positioning a weakly base paired region of U6 near a large electropositive patch on RRM1 (21). Our results are consistent with the C-terminal half of Prp24 also acting as a matchmaker domain. oRRM4 appears to bind RNA through a large electropositive surface, and is capable of destabilizing the U6 ISL, which suggests that it may serve a similar function to RRM1. Although, the U6 ISL is an attractive target for oRRM4’s destabilization activity, we cannot rule out that it may act on other secondary structure elements in the context of full-length U6 RNA, such as the telestem (nucleotide 29–39 + 92–103 in the most recent model) (7,17) or the previously proposed central stem loop (nucleotide 29–59) (11).

Based on previous data (21) and results presented here, we propose a mechanistic (Figure 6A) and structural

(Figure 6B) model for all four domains of Prp24 bound to U6 RNA. This model extends our previous model of RRM1 and 2 bound adjacent to the U6 ISL (21). RRM1 and 2 are bound to U6 nucleotide 49–60; RRM2 binds sequence specifically, while RRM1 destabilizes a weakly paired region (nucleotide 54–61 + 86–91) of U6 that extends from the ISL (17,21). RRM3 may provide additional affinity by interacting with the ISL, consistent with chemical shift perturbation studies (Supplementary Figure S1). Finally, the electropositive flanking helices of oRRM4 are positioned at the base of the ISL, where they would be able to disrupt base pairs or capture helical fraying motions. This model shows that the individual domains of Prp24 are physically capable of interacting with U6 RNA in a manner consistent with available data. For example, our model is consistent with previously observed hydroxyl radical footprinting results studying an *in vitro* complex between Prp24 and U6 snRNA (17). The separation between RRM2 and 3 in the model is made possible by a flexible 10 residue linker (16); the presence (but not sequence) of this linker appears universally conserved (data not shown).

U6-A62G is a cold-sensitive mutation thought to act through hyperstabilizing the U6 ISL (11). A screen for spontaneous suppressors of A62G found mutations in U6, U4 and RRM2 and 3 of Prp24 (7,11). When we map the locations of the A62G suppressor mutations in Prp24 onto our structural model, we find that they are all located near the U6 RNA (Figure 6B). This finding suggests that the source of suppression may be a modulation of Prp24's RNA binding activity.

Beyond demonstrating a potential mechanism by which Prp24 helps anneal U4 and U6, our model suggests a mechanism, by which Prp24 is released from the U4/U6 di-snRNP. The helices in the U4/U6 di-snRNA complex, once formed, are predicted to have a substantially more favorable free energy than the U6 ISL (−37.9 versus −7.0 kcal/mol, respectively, at 1 M NaCl, 37°C) (52,53). If RRM1 and oRRM4 bind preferentially to single stranded or frayed regions of RNA, the increased stability of the U4/U6 complex would eliminate their binding sites. Coupled with the loss of the RRM3 binding site due to ISL unwinding, this would result in release of Prp24. Furthermore, the model makes testable predictions related to the orientation of protein–RNA contacts, and may help guide further investigations into the mechanism of Prp24-mediated assembly of the U4/U6 di-snRNP.

## ACCESSION NUMBERS

The structure of oRRM4 has been deposited to the Protein Data Bank (PDB ID: 2L9W) and to the BioMagResBank (BMRB ID: 17490). Backbone resonance assignments for 292–444 and RRM3 have been deposited to the BioMagResBank (BMRB IDs: 17491 and 17589, respectively).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge Marco Tonelli and Mark Anderson for assisting with NMR data collection. We also appreciate helpful discussions with and support from members of the Butcher and Brow laboratories, and comments on the manuscript from James Keck.

## FUNDING

The National Institutes of Health (NIH) (grant number GM065166). NIH pre-doctoral training grant (grant number GM007215 to S.M.-T.); William R. and Dorothy E. Sullivan Wisconsin Distinguished Graduate Fellowship to S.M.-T. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by the NIH (grant numbers P41RR02301 and P41GM66326); Additional NMRFAM equipment was purchased with funds from the University of Wisconsin, the NIH (grant numbers RR02781 and RR08438); the National Science Foundation (grant numbers DMB-8415048, OIA-9977486 and BIR-9214394), and the United States Department of Agriculture. Funding for open access charge: Grant number R01 GM065166 (to S.E.B. and D.A.B.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Wahl, M.C., Will, C.L. and Luhrmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
- Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.*, **21**, 2724–2735.
- Shannon, K.W. and Guthrie, C. (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. *Genes Dev.*, **5**, 773–785.
- Ragunathan, P.L. and Guthrie, C. (1998) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science*, **279**, 857–860.
- Ghetti, A., Company, M. and Abelson, J. (1995) Specificity of Prp24 binding to RNA: a role for Prp24 in the dynamic interaction of U4 and U6 snRNAs. *RNA*, **1**, 132–145.
- Jandrositz, A. and Guthrie, C. (1995) Evidence for a Prp24 binding site in U6 snRNA and in a putative intermediate in the annealing of U6 and U4 snRNAs. *EMBO J.*, **14**, 820–832.
- Vidaver, R.M., Fortner, D.M., Loos-Austin, L.S. and Brow, D.A. (1999) Multiple functions of *Saccharomyces cerevisiae* splicing protein Prp24 in U6 RNA structural rearrangements. *Genetics*, **153**, 1205–1218.
- Rader, S.D. and Guthrie, C. (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. *RNA*, **8**, 1378–1392.
- Trede, N.S., Medenbach, J., Damianov, A., Hung, L.H., Weber, G.J., Paw, B.H., Zhou, Y., Hersey, C., Zapata, A., Keefe, M. *et al.* (2007) Network of coregulated spliceosome components revealed by zebrafish mutant in recycling factor p110. *Proc. Natl Acad. Sci. USA*, **104**, 6608–6613.
- Brow, D.A. and Guthrie, C. (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature*, **334**, 213–218.
- Fortner, D.M., Troy, R.G. and Brow, D.A. (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev.*, **8**, 221–233.
- Ryan, D.E., Stevens, S.W. and Abelson, J. (2002) The 5' and 3' domains of yeast U6 snRNA: Lsm proteins facilitate binding of Prp24 protein to the U6 telestem region. *RNA*, **8**, 1011–1033.



13. Verdone, L., Galardi, S., Page, D. and Beggs, J.D. (2004) Lsm proteins promote regeneration of pre-mRNA splicing activity. *Curr. Biol.*, **14**, 1487–1491.
14. Karaduman, R., Dube, P., Stark, H., Fabrizio, P., Kastner, B. and Luhrmann, R. (2008) Structure of yeast U6 snRNPs: arrangement of Prp24p and the LSm complex as revealed by electron microscopy. *RNA*, **14**, 2528–2537.
15. Kwan, S.S. and Brow, D.A. (2005) The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 RNA binding. *RNA*, **11**, 808–820.
16. Bae, E., Reiter, N.J., Bingman, C.A., Kwan, S.S., Lee, D., Phillips, G.N. Jr, Butcher, S.E. and Brow, D.A. (2007) Structure and interactions of the first three RNA recognition motifs of splicing factor Prp24. *J. Mol. Biol.*, **367**, 1447–1458.
17. Karaduman, R., Fabrizio, P., Hartmuth, K., Urlaub, H. and Luhrmann, R. (2006) RNA structure and RNA-protein interactions in purified yeast U6 snRNPs. *J. Mol. Biol.*, **356**, 1248–1262.
18. Vidal, V.P., Verdone, L., Mayes, A.E. and Beggs, J.D. (1999) Characterization of U6 snRNA-protein interactions. *RNA*, **5**, 1470–1481.
19. Maris, C., Dominguez, C. and Allain, F.H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.*, **272**, 2118–2131.
20. Kenan, D.J., Query, C.C. and Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.
21. Martin-Tumasz, S., Reiter, N.J., Brow, D.A. and Butcher, S.E. (2010) Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24. *RNA*, **16**, 792–804.
22. Wang, W. and Malcolm, B.A. (1999) Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques*, **26**, 680–682.
23. Zheng, L., Baumann, U. and Reymond, J.L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
24. Martin-Tumasz, S. and Butcher, S.E. (2009) (1)H, (13)C and (15)N resonance assignments of a ribonucleoprotein complex consisting of Prp24-RRM2 bound to a fragment of U6 RNA. *Biomol. NMR Assign.*, **3**, 227–230.
25. 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.
26. Herrmann, T., Guntert, P. and Wuthrich, K. (2002) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR*, **24**, 171–189.
27. Fiorito, F., Herrmann, T., Damberger, F.F. and Wuthrich, K. (2008) Automated amino acid side-chain NMR assignment of proteins using (13)C- and (15)N-resolved 3D [(1)H, (1)H]-NOESY. *J. Biomol. NMR*, **42**, 23–33.
28. Volk, J., Herrmann, T. and Wuthrich, K. (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J. Biomol. NMR*, **41**, 127–138.
29. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
30. Brunger, A.T. (2007) Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.*, **2**, 2728–2733.
31. Dominguez, C., Boelens, R. and Bonvin, A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
32. de Vries, S.J., van Dijk, A.D., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., Wassenaar, T. and Bonvin, A.M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
33. Zweckstetter, M. and Bax, A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. *J. Am. Chem. Soc.*, **122**, 3791–3792.
34. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. and Clore, G.M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, **160**, 65–73.
35. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
36. Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E. and Dyson, H.J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.*, **123**, 2970–2978.
37. Oberstrass, F.C., Auweter, S.D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D.L. et al. (2005) Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science*, **309**, 2054–2057.
38. Teplova, M., Yuan, Y.R., Phan, A.T., Malinina, L., Ilin, S., Teplov, A. and Patel, D.J. (2006) Structural basis for recognition and sequestration of UUU(OH) 3' termini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen. *Mol. Cell*, **21**, 75–85.
39. Fribourg, S., Gatfield, D., Izaurralde, E. and Conti, E. (2003) A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat. Struct. Biol.*, **10**, 433–439.
40. Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Luhrmann, R., Glover, J.N. and MacMillan, A.M. (2006) Crystal structure of a core spliceosomal protein interface. *Proc. Natl Acad. Sci. USA*, **103**, 1266–1271.
41. Iqbal, A., Wang, L., Thompson, K.C., Lilley, D.M. and Norman, D.G. (2008) The structure of cyanine 5 terminally attached to double-stranded DNA: implications for FRET studies. *Biochemistry*, **47**, 7857–7862.
42. Dolgih, E., Roitberg, A.E. and Krause, J.L. (2007) Fluorescence resonance energy transfer in dye-labeled DNA. *J. Photochem. Photobiol. A-Chem.*, **190**, 321–327.
43. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D. Biol. Crystallogr.*, **60**, 2256–2268.
44. Eulalio, A., Tritschler, F., Buttner, R., Weichenrieder, O., Izaurralde, E. and Truffault, V. (2009) The RRM domain in GW182 proteins contributes to miRNA-mediated gene silencing. *Nucleic Acids Res.*, **37**, 2974–2983.
45. Tresaugues, L., Dehe, P.M., Guerois, R., Rodriguez-Gil, A., Varlet, I., Salah, P., Pamblanco, M., Luciano, P., Quevillon-Cheruel, S., Sollier, J. et al. (2006) Structural characterization of Set1 RNA recognition motifs and their role in histone H3 lysine 4 methylation. *J. Mol. Biol.*, **359**, 1170–1181.
46. Allain, F.H., Howe, P.W., Neuhaus, D. and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5772.
47. Mittermaier, A., Varani, L., Muhandiram, D.R., Kay, L.E. and Varani, G. (1999) Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.*, **294**, 967–979.
48. Perez Canadillas, J.M. and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J.*, **22**, 2821–2830.
49. Dominguez, C., Fiset, J.F., Chabot, B. and Allain, F.H. (2010) Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.*, **17**, 853–861.
50. Beuth, B., Garcia-Mayoral, M.F., Taylor, I.A. and Ramos, A. (2007) Scaffold-independent analysis of RNA-protein interactions: the Nova-1 KH3-RNA complex. *J. Am. Chem. Soc.*, **129**, 10205–10210.
51. Skrisovska, L., Bourgeois, C.F., Steff, R., Grellescheid, S.N., Kister, L., Wenter, P., Elliott, D.J., Stevenin, J. and Allain, F.H. (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep.*, **8**, 372–379.
52. Markham, N.R. and Zuker, M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, W577–W581.
53. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.