Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24

STEPHEN MARTIN-TUMASZ,1 NICHOLAS J. REITER,1,3 DAVID A. BROW,2 and SAMUEL E. BUTCHER1

1Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
2Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

ABSTRACT

U6 RNA plays a critical role in pre-mRNA splicing. Assembly of U6 into the spliceosome requires a significant structural rearrangement and base-pairing with U4 RNA. In the yeast Saccharomyces cerevisiae, this process requires the essential splicing factor Prp24. We present the characterization and structure of a complex containing one of Prp24’s four RNA recognition motif (RRM) domains, RRM2, and a fragment of U6 RNA. NMR methods were used to identify the preferred U6 binding sequence of RRM2 (5'-GAGA-3'), measure the affinity of the interaction, and solve the structure of RRM2 bound to the hexaribonucleotide AGAGAU. Interdomain contacts observed between RRM2 and RRM3 in a crystal structure of the free protein are not detectable in solution. A structural model of RRM1 and RRM2 bound to a longer segment of U6 RNA is presented, and a partial mechanism for Prp24’s annealing activity is proposed.

Keywords: U6 snRNA; Prp24; RRM; ribonucleoprotein; NMR; RNA

INTRODUCTION

Pre-mRNA splicing is essential for proper gene expression in eukaryotes. In order to make a functional messenger RNA (mRNA), introns must be removed from pre-mRNA and exons ligated (spliced) together. Splicing involves many cellular components, including 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) (Will and Luhrmann 2006; Wahl et al. 2009). The five snRNPs, along with additional proteins, form the spliceosome, which catalyzes pre-mRNA splicing. Both the pre-mRNA and the snRNAs undergo essential dynamic rearrangements during splicing, many of which may be facilitated by RNA chaperone or matchmaker proteins. One such protein is Saccharomyces cerevisiae Prp24 (known as p110 in humans) (Bell et al. 2002), an essential component of the U6 snRNP that is required to anneal the U4 and U6 RNAs during splicing (Shannon and Guthrie 1991; Ghetti et al. 1995; Jandrositz and Guthrie 1995; Raghunathan and Guthrie 1998; Vidaver et al. 1999; Rader and Guthrie 2002; Trede et al. 2007).

The U6 snRNP contains the U6 RNA and eight proteins: Prp24 and a seven-membered ring of Lsm proteins (Lsm2-8) (Karaduman et al. 2008). Prp24 is thought to bind U6 RNA at a predominately single-stranded region upstream of the internal stem–loop (ISL) (Ghetti et al. 1995; Jandrositz and Guthrie 1995; Kwan and Brow 2005; Karaduman et al. 2006; Bae et al. 2007), while the Lsm ring binds the uridine-rich 3’ end of U6 RNA (Vidal et al. 1999; Rader and Guthrie 2002; Ryan et al. 2002). To enter the splicing cycle, U6 RNA must base-pair to U4 RNA and form the U4/U6 di-snRNP. For this to occur, the ISL must unwind (Brow and Guthrie 1988; Fortner et al. 1994). While the precise mechanism of this process is unknown, there is extensive evidence showing that Prp24 plays an essential role in annealing the U4 and U6 RNAs (Shannon and Guthrie 1991; Ghetti et al. 1995; Jandrositz and Guthrie 1995; Raghunathan and Guthrie 1998; Vidaver et al. 1999; Rader and Guthrie 2002; Ryan et al. 2002; Verdone et al. 2004).

Prp24 contains four RNA recognition motif (RRM) domains (Fig. 1A; Shannon and Guthrie 1991; Rader and Guthrie 2002; Bae et al. 2007). RRMs are common protein...
motifs that typically bind single-stranded RNA along a four-stranded anti-parallel β-sheet. Two conserved sequences enriched in aromatic residues define an RRM domain, the ribonucleoprotein (RNP) 1 and 2 motifs (Fig. 1B; Kenan et al. 1991; Maris et al. 2005). A number of structures of RRMs bound to nucleic acids have been solved, allowing for a canonical interaction to be established (Maris et al. 2005; Clery et al. 2008). The core of the interaction occurs between two nucleotides in the RNA and specific aromatic and electropositive residues in the protein’s RNP1 and RNP2 motifs (Maris et al. 2005).

Two partial structures of Prp24 in isolation have been determined: a crystal structure of the first three RRMs (referred to as RRM123) and an NMR structure of the first two (RRM12) (Bae et al. 2007). In both structures, the β-sheet surfaces of RRM1 and RRM2 are blocked by neighboring RRMs, occluding the canonical RNA binding sites. Therefore, it was hypothesized that RNA binding by Prp24 may be noncanonical and/or accompanied by a change in protein conformation (Bae et al. 2007). Only the N-terminal half of Prp24 (RRM12) is necessary for high-affinity binding to U6 RNA (Bae et al. 2007). Scaffold independent analysis (SIA) requires a single RNA binding site, so our investigation required use of a single RRM domain. Of these two domains, we found that only RRM2 could be expressed and purified as an independently structured and moderately soluble domain (data not shown).

We report the preferred RNA binding site for Prp24-RRM2, which corresponds to a single-stranded region in U6 RNA. Structure determination of a complex containing RRM2 and this part of U6 reveals that RRM2 binds RNA across its β-sheet surface, in canonical fashion. Our results show that the packing between RRM2 and RRM3 observed in the crystal structure is not significantly present in solution. By combining the structure presented here with NMR studies of a complex containing RRM12 and a 12-nucleotide (nt) RNA, we generate a model of RRM12 bound to a longer U6 RNA fragment. This model suggests a partial mechanism for Prp24’s match-making activity—RRM2 positions U6 RNA through sequence-specific interactions, while RRM1 uses electrostatic interactions with the phosphate backbone to destabilize an internal helix in U6 RNA so that it can anneal to U4 RNA.

RESULTS

RRM2 binds U6 RNA at a conserved GAGA sequence

The optimal RNA binding sequence for RRM2 was identified through SIA (Beuth et al. 2007). SIA is an NMR-based method that provides an unbiased measure of the sequence specificity of a nucleic acid binding protein. It was performed using 16 tetramer RNA pools, each with one defined and three random positions (e.g., ANNN). For each pool, two 1H,15N-heteronuclear single quantum coherence (15N-HSQC) spectra were acquired of RRM2, either in the absence or presence of RNA. These spectra show peaks corresponding to most amides in the protein. Because the RRM2–RNA interaction is fast exchange on the NMR timescale, the magnitudes of changes in peak position induced by RNA binding correlate to the extent of RNA binding. By comparing changes resulting from each pool, the preferred binding sequence was identified (Fig. 2A,B).

RRM2 specifically recognizes G at the first RNA position and A at the second. The third position is preferred to be purine, while the fourth shows little discrimination between nucleotides (Fig. 2A). These results are consistent with the canonical RRM–RNA interaction, in which conserved protein residues interact predominantly with a di-nucleotide (Maris et al. 2005). In addition, they suggest that other domains of Prp24 and/or another factor (such as the Lsm ring) (Rader and Guthrie 2002) must provide additional specificity to produce a highly specific Prp24–U6 RNA interaction.

In order to identify the biologically relevant binding sequence of RRM2, the SIA results were compared to previous studies (Fig. 2C). Previous electrophoretic mobility shift studies identified nucleotides 47–58 (bold in Fig. 2C) as the high-affinity binding site for the first three RRMs of Prp24 (Kwan and Brow 2005). Many of the nucleotides in this region also show chemical protection in the U6 snRNP, implying close contact with a protein (Ghetti et al. 1995; Jandrositz and Guthrie 1995; Karaduman et al. 2006). Finally, a cross-link between U6 RNA and Prp24 has been identified in this region (Karaduman et al. 2006). These data strongly indicate that Prp24 binds U6 RNA in the region of nucleotides 40–60, although they also suggest that...
mapped onto the existing crystal structure of RRM2 (Bae et al. 2007), revealing a putative RNA binding site located on the β-sheet face of RRM2 and in a nearby flexible loop (Fig. 3A,B). These results are entirely consistent with a canonical RRM–RNA interaction. For example, aromatic residues in the RNP1 and RNP2 motifs show dramatic chemical shift changes (W120, F160, and Y162) (Figs. 1B, 3B).

These changes in chemical shift can be used to calculate the apparent dissociation constant ($K_d$) of the interaction. For RRM2 and AGAGAU, the apparent $K_d$ is 90 ± 10 μM (Fig. 3D). This affinity is comparable to other RRM–RNA interactions studied by NMR (Auweter et al. 2007) and it suggests that the nanomolar dissociation constants reported for larger Prp24–RNA complexes are due to the cooperative binding of the multiple RRM domains with U6 RNA (Kwan and Brow 2005). Based on the SIA results, RRM2 should have a much lower affinity for UUU UUU RNA. Indeed, this interaction has an apparent $K_d$ of 1.1 ± 0.2 mM (Fig. 3D), a decrease in binding affinity of more than 10-fold relative to AGAGAU.

Canonical RRM–RNA interactions have a defined orientation of the RNA, with the 3’ end toward the β2 strand of the β-sheet. In order to locate the 3’ end of AGAGAU, a nitroxide spin label was attached to the 3’ end of the RNA through a 4-thio-U substitution (Ramos and Varani 1998). $^{15}$N-HSQC spectra taken of RRM2 in the absence and presence of the spin-labeled RNA showed a dramatic loss of signal from peaks corresponding to three specific residues (Fig. 3B; Supplemental Fig. 1A). These residues (Q130, I147, and L149) are located on or near the β2 strand, demonstrating that RRM2 binds AGAGAU in the canonical orientation.

**The RRM2–AGAGAU interaction is maintained in the presence of RM3**

In the crystal structure of RRM123, the β-sheet face of RRM2 (the identified RNA binding site) is blocked by contacts with RM3 (Bae et al. 2007). Therefore, prior to determining the structure of RRM2 bound to AGAGAU, it was necessary to verify that the observed binding site is not an artifact caused by the absence of RM3.

The chemical shift perturbation and nitroxide spin label experiments described above were repeated using RRM23.

**Prp24 may additionally contact other upstream sequences.**

The only sequence in this region in close agreement with the SIA-derived sequence located in a region previously implicated in Prp24–U6 binding. Results from previous experiments are indicated as follows: Bold nucleotides are the Prp24-N123 high-affinity binding site (Kwan and Brow 2005). In purified yeast U6 snRNP, red nucleotides show some protection from hydroxyl radicals (ribose protection), blue nucleotides show some protection from base modifications, and purple residues show protection from both (Karaduman et al. 2006). The red arrows show the sites of observed cross-links between Prp24 and U6 RNA (Karaduman et al. 2006). The cross-linking and chemical protection studies covered nucleotides 2–89 in U6 RNA.

**RRM2 binds AGAGAU canonically**

Chemical shift perturbation was used to identify the AGAGAU binding site on RRM2. Adding AGAGAU to RRM2 resulted in significant shifts of specific amide peaks, while many other peaks did not shift at all (Fig. 3A). The locations of the residues corresponding to these peaks were consistent with a canonical RRM–RNA interaction. For example, aromatic residues in the RNP1 and RNP2 motifs show dramatic chemical shift changes (W120, F160, and Y162) (Figs. 1B, 3B).

These changes in chemical shift can be used to calculate the apparent dissociation constant ($K_d$) of the interaction. For RRM2 and AGAGAU, the apparent $K_d$ is 90 ± 10 μM (Fig. 3D). This affinity is comparable to other RRM–RNA interactions studied by NMR (Auweter et al. 2007) and it suggests that the nanomolar dissociation constants reported for larger Prp24–RNA complexes are due to the cooperative binding of the multiple RRM domains with U6 RNA (Kwan and Brow 2005). Based on the SIA results, RRM2 should have a much lower affinity for UUU UUU RNA. Indeed, this interaction has an apparent $K_d$ of 1.1 ± 0.2 mM (Fig. 3D), a decrease in binding affinity of more than 10-fold relative to AGAGAU.

Canonical RRM–RNA interactions have a defined orientation of the RNA, with the 3’ end toward the β2 strand of the β-sheet. In order to locate the 3’ end of AGAGAU, a nitroxide spin label was attached to the 3’ end of the RNA through a 4-thio-U substitution (Ramos and Varani 1998). $^{15}$N-HSQC spectra taken of RRM2 in the absence and presence of the spin-labeled RNA showed a dramatic loss of signal from peaks corresponding to three specific residues (Fig. 3B; Supplemental Fig. 1A). These residues (Q130, I147, and L149) are located on or near the β2 strand, demonstrating that RRM2 binds AGAGAU in the canonical orientation.

**The RRM2–AGAGAU interaction is maintained in the presence of RM3**

In the crystal structure of RRM123, the β-sheet face of RRM2 (the identified RNA binding site) is blocked by contacts with RM3 (Bae et al. 2007). Therefore, prior to determining the structure of RRM2 bound to AGAGAU, it was necessary to verify that the observed binding site is not an artifact caused by the absence of RM3.

The chemical shift perturbation and nitroxide spin label experiments described above were repeated using RRM23.
Chemical shift perturbation revealed an essentially identical binding site (Fig. 3C) with an even lower apparent $K_d$ of $30 \pm 10 \mu M$ (Fig. 3D). Nitroxide spin labeling demonstrated that the orientation of AGAGAU across the $\beta$-sheet face was unchanged (Fig. 3C; Supplemental Fig. 1B). Thus, the observed interaction with RRM2 in isolation also occurs in the larger RRM23 construct. This result suggests that either RNA binding is accompanied by a conformational change between the two RRMs, or the orientation in the crystal structure is not significantly present in solution.

Close interdomain contacts can be detected through chemical shift perturbation by examining spectra of different constructs, because residues at an interface should display chemical shift changes when the partner domain is removed. Peak positions for residues in RRM2 were compared in $^{15}$N-HSQC spectra of RRM12, RRM2, and RRM23 (Fig. 4A). When comparing peak positions in RRM12 versus RRM2, there are dramatic differences at peaks corresponding to specific residues that predominantly map to the interface between RRM1 and RRM2 in the previous NMR and crystal structures or to flexible regions (Bae et al. 2007). However, when comparing RRM2 and RRM23, there are no significant differences in chemical shifts of residues in structured regions, implying that the close contacts between RRM2 and RRM3 observed in the crystal structure are not present in solution (Fig. 4A).

Residual dipolar couplings (RDCs) (Bax 2003) were used to compare the orientation between RRM2 and RRM3 in solution to the orientation in the crystal structure. RDC values provide information on the orientation of bonds in a molecule. Experimentally obtained RDC values for RRM2 and RRM3 were consistent with the crystal structure of each individual domain (data not shown), demonstrating that the individual domain structures are similar in the crystalline and solution states. However, the RDC values observed for the RRM23 construct, either in the absence of RNA (Fig. 4B) or in the presence of AGAGAU RNA (Fig. 4C), were not consistent with RDC values calculated from the crystal structure of RRM123. In combination with the chemical shift perturbation results, these data show that the crystal structure conformation is not significantly populated in solution.

Structure of RRM2 bound to AGAGAU

The structure of RRM2 bound to AGAGAU was determined by NMR, revealing that the core of the interaction is between the $\beta$-sheet face of RRM2 and two nucleotides—G50 and A51 (Fig. 5; Table 1). These nucleotides make the canonical RNP1 and RNP2 contacts. G50 and A51 both stack on aromatic protein residues, W120 and Y162, respectively (Fig. 6A,B). R158 forms a hydrogen
Sequence specificity is derived from hydrogen bonding to Watson–Crick functional groups on the nucleotides. The G50 Watson–Crick face is recognized by three hydrogen bonds: the ε-amine on K191 bonds with the carbonyl group, while the side-chain carbonyl of N194 bonds with imino and amino protons (Fig. 6A). A51 has a single hydrogen bond between the side-chain hydroxyl of T118 and the amino group (Fig. 6B). A guanosine substitution at this position would result in a steric clash with Y162. These hydrogen bonds agree with the observed pattern of intermolecular NOEs. Both of these positions have hydrogen bonds between acceptor groups on the protein and RNA amino groups, explaining why U (which lacks an amino group) is the least tolerated nucleotide at these two positions (Fig. 2A). Both of these positions also contain nonspecific interactions with the RNA backbone, which may serve to discriminate in favor of purines. Pyrimidines, being smaller, would be unable to make both the specific contact(s) at the Watson–Crick face and the nonspecific contact through the ribose group or the phosphate backbone. G52 predominately interacts with the guanidinium group of R148 in a cation–π interaction (Fig. 6C). SIA revealed a preference for purines at this position (Fig. 2A), and purines are much more commonly involved in cation–π interactions than pyrimidines (Biot et al. 2002; Rooman et al. 2002). A53 is disordered, consistent with SIA results showing the least specificity at the fourth position (Fig. 2A).

This structure is consistent with the chemical shift perturbation results (Fig. 3A,B). All of the residues that showed significant changes in peak position in the presence of AGAGAU are located in the RNA binding site or in flexible regions (the N and C termini, or the loop between β2 and β3) (cf. Figs. 3B and 5A). Based on our structure, it appears that the changes in peak position in flexible regions are due to conformational changes rather than RNA binding. However, we cannot exclude the possibility that some of the residues in these regions (particularly the β2–β3 loop) are either transiently or stably interacting with RNA, but the relevant NOEs are too weak to observe.

The structure is also consistent with sequence conservation and mutagenesis. Roughly one-third of the residues in RRM2 are either similar or identical in at least 13 of 14 fungal species investigated (Fig. 1B). While many are conserved for internal structural reasons, several of the residues

**FIGURE 4.** Absence of RRM2–RRM3 contacts in solution. (A) Chemical shift perturbation of RRM2. Dynamic regions are bounded by dashed lines. The presence of RRM1 (filled circles) results in significant changes in structured regions of RRM2 corresponding to the RRM1/RRM2 interface. The presence of RRM3 (open circles) does not, implying there are no interdomain contacts. (B) Residual dipolar coupling (RDC) analysis of RRM23 without RNA. Comparing solution data to the crystal structure shows a poor fit, implying that the solution orientation does not match the crystal structure. (C) RDC analysis of RRM23 bound to AGAGAU RNA. As in panel B, there is a poor fit to the crystal structure. However, the RDC fit is significantly improved (Q of 1.2 bound to RNA vs. 3.0 in isolation), implying a conformational change on RNA binding.

**FIGURE 5.** Structure of RRM2 bound to AGAGAU RNA. (A) Bundle of the 10 lowest energy structures after AMBER refinement. See Table 1 for relevant statistics. The protein is in gray, and the RNA is colored by nucleotide: (red) A49; (orange) G50; (yellow) A51; (green) G52; (blue) A53; (pink) U54. (B) Representative structure. Key protein residues are displayed as sticks, and hydrogen bonds are indicated by colored atoms ([red] oxygen or hydroxyl; [blue] nitrogen, amino, or imino) and dashed lines. These interactions are described in detail in Figure 6.
appear to be conserved in order to maintain the correct RNA interaction. For example, T118 and R158 are identical, preserving intermolecular hydrogen bonds. W120 is always aromatic, maintaining base-stacking with G50, while Y162 is identical in order to maintain both base-stacking and an intermolecular hydrogen bond with A51. N194 can be either asparagine or aspartate, either of which can make the correct hydrogen bond with the Watson–Crick face of G50.

Previous work found that an R158S mutation is temperature-sensitive, while a triple alanine mutation of R158, F160, and Y162 is lethal (Vidaver et al. 1999). In the crystal structure, these residues are close to the RRM2–RRM3 interface, and two of them form interdomain hydrogen bonds, suggesting that the phenotypes resulted from altering interactions with RRM3 rather than by affecting RNA binding (Bae et al. 2007). However, we observe that these residues interact with RNA and propose that the phenotypes result from disrupting the RRM2–AGAGAU binding interaction. The apparent $K_d$ of RRM2–R158S for AGAGAU was measured by chemical shift perturbation to be $440 \pm 30\mu M$ for AGAGAU (Supplemental Fig. 2A), roughly five times that of wild-type RRM2. No change in the protein structure (judged by $^{15}$N-HSQC peak positions) or RNA binding site was observed (Supplemental Fig. 2B). The $^{15}$N-HSQC spectrum of RRM23–R158S also showed no significant chemical shift changes (Supplemental Fig. 2C), supporting the proposal that the hydrogen bond between R158 and S230 in the crystal structure is not significantly present in solution. In addition, we investigated the effect of the lethal mutations on RNA binding using an RRM2–R158A F160A Y162A construct. Based on analysis of $^{15}$N-HSQC spectra taken when AGAGAU RNA was titrated into this construct, RNA binding by RRM2 is completely abrogated by the mutation, while the protein retains a similar structure (Supplemental Fig. 2D).

### A model for more extensive U6–Prp24 interactions

The structure of RRM2 bound to AGAGAU facilitates the modeling of a larger complex. NMR and crystallographic data indicated that RRM1 and RRM2 are rigidly connected (Bae et al. 2007), making a significant domain reorientation on RNA binding unlikely. Chemical shift perturbation studies using a 40-nt segment of U6 RNA also identified a potential RNA binding site on RRM1 (Bae et al. 2007). Because RRM2 binds the AGAGAU sequence in U6 RNA, with the 3′ terminus adjacent to RRM1, it is reasonable to expect that RRM1 binds the immediately downstream GAUCAG sequence. This would be entirely consistent with the previous studies summarized in Figure 2C.
In order to map the interaction between RRM12 and AGAGAUGAUCAG (corresponding to S. cerevisiae U6 RNA positions 49–60), we identified residues that showed peak perturbation in the presence of a stoichiometric amount of AGAGAUGAUCAG RNA (Supplemental Fig. 3A). The binding site identified on RRM2 is essentially identical to the site identified by adding AGAGAU to RRM2 alone. We also obtained intermolecular distance restraints by attaching a nitrooxide spin label to either position 6, 9, or 12 in AGAGAUGAUCAG and looking for signal loss from residues in RRM12 (Supplemental Fig. 3B–D). Our model incorporates these restraints, as well as all intermolecular restraints used in generating the structure of RRM2 bound to AGAGAU.

The resulting model (Fig. 7) suggests an interesting functional role for RRM1. The presence of RRM1 forces the RNA to pass through a fairly narrow cleft between the two protein domains, making the bases relatively inaccessible. On RRM1, the RNA wraps around behind the β-sheet face in a noncanonical interaction (Fig. 7B). Many of the intermolecular contacts in the model are between protein residues and the backbone of the RNA. This results in the bases having relatively little contact with the protein, particularly toward the 3′ end. The region of RRM1 observed to bind U6 is part of a large electropositive patch (Fig. 7C; Bae et al. 2007), so it is reasonable for RNA binding to occur predominantly through the electronegative backbone. Additionally, the nucleotides interacting with RRM1 are predicted to be base-paired in other forms of U6 RNA: with other U6 nucleotides in free U6 RNA (Karaduman et al. 2006), and with U4 RNA in the U4/U6 di-snRNP (Brow and Guthrie 1988). Unwinding and reannealing these RNAs would be facilitated by the relatively high accessibility of certain U6 RNA Watson–Crick faces when bound to Prp24.

**DISCUSSION**

We have proposed the biologically relevant RNA target for RRM2 and determined the structure of a complex containing RRM2 and this fragment of U6 RNA. The structure provides an explanation for observed sequence conservation and previously reported phenotypes of mutations in RRM2. Based on the crystal structure, it was hypothesized that the mutations disrupted interdomain contacts between RRM2 and RRM3, but this structure reveals that they likely act through disrupting RNA binding. Building on this structure, we have developed a detailed structural model of RRM12 bound to a longer segment of U6 RNA.

**A mechanism for U4/U6 annealing**

The model of RRM12 bound to a 12-nt sequence in U6 RNA suggests a mechanism for Prp24’s function in facilitating base-pairing between U4 RNA and U6 RNA (Fig. 8). RRM2 binds a GAGA sequence in U6 RNA (nucleotides 50–53). This site is ideally positioned to facilitate conformational changes in U6 RNA and base-pairing with U4 RNA. It is single-stranded in the proposed secondary structures of free U6 RNA (Karaduman et al. 2006), the U6 snRNP (Karaduman et al. 2006), and the U4/U6 di-snRNP (Fig. 8A; Brow and Guthrie 1988; Ghetti et al. 1995). In free U6 RNA, the site is directly adjacent to the start of a short helix (nucleotides 54–57) (Fig. 8A) that is not present in the U6 snRNP (Karaduman et al. 2006). We propose that RRM2 binding GAGA provides initial sequence-specific recognition of U6 and positions RRM1 near the short helix (Fig. 8B). This helix contains only two Watson–Crick pairs, both of which would be expected to have very short lifetimes (Snoussi and Leroy 2001), so it is likely to have pronounced breathing motions. Through this breathing, we propose that RRM1 is able to capture the unpaired form of the helix by binding one of the strands (Fig. 8C).

Because RRM1 appears to bind RNA electrostatically through the phosphate backbone and ribose groups (Fig. 7C), the Watson–Crick faces of these nucleotides are exposed to solution, facilitating their base-pairing with U4 RNA (Fig. 8D). Base-pairing may occur concurrently with unwinding the ISL or after unwinding is completed. However, unwinding the ISL and annealing to U4 RNA concurrently would seem preferable. The energetic barrier...
to breaking single intra-U6 base pairs and immediately forming a new U4/U6 base pair would be considerably lower than the barrier to unwinding the entire ISL followed by annealing to U4. While it is possible to trap intermediates containing U4 RNA, U6 RNA, and Prp24 (Shannon and Guthrie 1991; Jandrositz and Guthrie 1995), Prp24 is not present in the wild-type U4/U6 di-snRNP (Fig. 8E). Therefore, Prp24 must be released at some point during or immediately after annealing, but it is unclear exactly how or when.

A significant caveat of schematic models is that they do not take into account the three-dimensional (3D) structures of the molecules involved. Therefore, we used XPLOR-NIH (Schwieters et al. 2003) to calculate structural models of our proposed RRM12–U6 (Fig. 8B,C) and RRM12–U4/U6 (Fig. 8D) complexes, based on the structural model of RRM12 bound to AGAGAUGAUCAG (Fig. 7) and the structure of the extended U6 ISL (Sashital et al. 2004). They demonstrate that RRM2 binding AGAGAU can position RRM1 next to the weakly paired base of the ISL (Supplemental Fig. 4A), where RRM1 can bind the ascending strand (Supplemental Fig. 4B) and facilitate its pairing to U4 (Supplemental Fig. 4C). Thus, our model is structurally reasonable.

This model is similar to other RNA matchmaking mechanisms, such as those proposed for gBP21 and MRP1/2 (Muller and Goringer 2002; Schumacher et al. 2006). Significantly, the model proposes a passive role for Prp24 in unwinding the U6 ISL. Rather than actively breaking base pairs like a helicase, we propose that Prp24 initiates unwinding by trapping an open conformation that occurs due to helical breathing. RRM1’s mode of capturing the open conformation of U6 facilitates initial base-pairing with U4, nucleating the complex. Unwinding of U6 and annealing to U4 can then proceed concurrently, essentially swapping intra-U6 base pairs for U4/U6 base pairs. The Lsm ring and other Prp24 domains likely also play a role in annealing U4 and U6, a potentially fruitful avenue of future research.

This model relies on the β-sheet face of RRM2 being exposed to solution. Interestingly, in the crystal structure, RRM3 occupied part of the RNA binding site on RRM2 and made several very similar contacts. For example, the observed hydrogen bond between the guanidinium group of R158 and the backbone carbonyl of S230 in the crystal structure is replaced by a hydrogen bond with the RNA phosphate backbone. This suggests that RRM3 is acting as an RNA mimic in the crystal structure, satisfying the binding potential of RRM2’s β-sheet face. Our results show different fits for RDC values obtained in the absence and presence of RNA (Fig. 4B,C), suggesting that the domain orientation may change when RRM2 binds AGAGAU. A closed conformation could exist transiently in solution or in vivo, and might serve a regulatory role by restricting RNA access to RRM2.

The role of other domains of Prp24

Triple alanine mutations of the RNP1 motif in each RRM domain of Prp24 support our model (Fig. 8). Only the RRM2 triple alanine mutation is lethal (Vidaver et al. 1999; Rader and Guthrie 2002), indicating that RRM2 is the only domain where a canonical RRM–RNA interaction is essential for Prp24’s activity. The triple alanine mutation in RRM1 alone has no phenotype (Vidaver et al. 1999), consistent with the structural model showing that the bulk of RRM1’s interaction with RNA occurs noncanonically, away from the β-sheet face. Also in agreement with the proposal that RRM2 is critical for positioning Prp24 on U6 RNA, the triple alanine mutations of RRM3 and RRM4 show a temperature-sensitive phenotype (Vidaver et al. 1999; Rader and Guthrie 2002), suggesting that canonical RNA recognition by these domains is functionally important but not essential.

The role of the other two domains of Prp24, RRM3 and RRM4, remains unclear. Previous work has shown that they are not required for in vitro high-affinity Prp24–U6 RNA binding (Kwan and Brow 2005; Bae et al. 2007). However,
other potential Prp24 binding sites on U6 RNA have been identified, in particular, an AAAC sequence (nucleotides 40–43) also located in the asymmetric bulge (Jandrositz and Guthrie 1995; Karaduman et al. 2006). Some evidence shows that Prp24 binds a double-stranded portion of Stem II in the U4/U6 di-snRNP (Ghetti et al. 1995). These sites may be recognized by RRM3 and/or RRM4. Because mutational studies have shown that the canonical RRM motifs in RRM3 and RRM4 are important for Prp24’s function (Vidaver et al. 1999; Rader and Guthrie 2002), and there is evidence for additional Prp24 binding sites in U6 RNA, we hypothesize that Prp24 binds the unpaired nucleotides, enlarging the asymmetric bulge and ultimately leading to the unwinding of U6 RNA’s internal stem-loop (“Capture”). (D) Because RRM1 binds predominantly to the phosphate backbone, the U6 RNA bases are free to pair with U4 RNA (“Nucleation”). (E) Prp24 is not found in the U4/U6 di-snRNP, so it must be released either during or after base-pairing (“Annealing”).

FIGURE 8. Schematic model of the functional role of RRM12. (A) The relevant region of the proposed secondary structure of free U6 RNA (Karaduman et al. 2006). (B) RRM2 binds the GAGA sequence in the bulge (“Recognition”). (C) As the helix breathes, RRM1 binds the unpaired nucleotides, enlarging the asymmetric bulge and ultimately leading to the unwinding of U6 RNA’s internal stem-loop (“Capture”). (D) Because RRM1 binds predominantly to the phosphate backbone, the U6 RNA bases are free to pair with U4 RNA (“Nucleation”). (E) Prp24 is not found in the U4/U6 di-snRNP, so it must be released either during or after base-pairing (“Annealing”).

Due to lack of evolutionary pressure.

Additional studies focusing on RNA binding and structural investigation of RRM3 and RRM4 should further clarify the mechanism of Prp24’s matchmaker activity. Of particular importance is the RNA binding target of RRM3 and RRM4 and whether they act as a single unit in solution like RRM1 and RRM2. Investigating the effects of Prp24, U4, and U6 mutations and truncations on annealing activity will be helpful in testing the mechanism proposed here, although it should be noted that U6’s many interactions and multiple conformations during the splicing cycle make it extremely challenging to decipher the molecular source of a phenotype. Finally, structural and functional studies of the RRM domains of human p110, and comparison to what is known about Prp24, may help identify which components are truly essential for U4/U6 di-snRNP formation.

MATERIALS AND METHODS

Protein and RNA preparation

All protein and RNA samples were prepared as described (Martin-Tumasz and Butcher 2009).
NMR experiments

All NMR experiments described here were performed on a 750 MHz Bruker spectrometer at 25°C, unless stated otherwise. Resonance assignments for RRM12 in the absence of RNA (Reiter et al. 2006) and RRM2 and RRM23 in the absence and presence of RNA (Martin-Tumaz and Butcher 2009) were used in this study. All spectra were processed using NMRPipe (Delaglio et al. 1995) and in-house scripts, and analyzed in SPARKY (TD Goddard and DG Kneller, unpubl.; SPARKY 3, http://www.cgl.ucsf.edu/home/sparky/).

Scaffold independent analysis

A single pool of 15N RRM2 with a final concentration of 240 μM in 90% NMR buffer and 10% D2O was used. Aliquots of each RNA sufficient to make 150 μL of 1.2 mM solution were lyophilized to dryness. For each RNA, two 15N-HSQC experiments were acquired on a 150 μL sample (protein alone, 5:1 RNA:protein). Peak positions for five residues (L119, W120, T122, Y162, and I163) were recorded for each spectrum. Peak movements were calculated and analyzed as described (Beuth et al. 2007).

NMR titration and chemical shift perturbation

15N-HSQC spectra were taken of 15N protein (typically, ~300 μM) in 90% NMR buffer and 10% D2O in the presence of various amounts of RNA (typically 0×, 0.2×, 0.5×, 2×, 5×, and 15× protein concentration). Peak positions were recorded for the five residues used in SIA. Changes in peak position were calculated using

$$\Delta\delta_{\text{Total}} = \sqrt{\Delta\delta_i^2 + \left(\Delta\delta_{300}/9.86204\right)^2},$$

where 9.86204 is the ratio of the gyromagnetic ratios of 1H and 15N, and then fit to the equation

$$\Delta\delta = \frac{\Delta\delta_{\text{max}}}{2P} \left[ (P + L + K_d) - \sqrt{(P + L + K_d)^2 - 4PL} \right],$$

where $P = [\text{protein}]$ and $L = [\text{RNA}]$, to estimate the dissociation constant (apparent $K_d$).

For chemical shift perturbation experiments, peak positions for all residues were recorded, and residues that showed significant changes in chemical shift when saturated with RNA (>0.1 ppm normalized to 1H) were mapped onto the crystal structures of free RRM12, RRM2, and RRM23.

Nitrooxide spin label experiments

Modified RNA [AGAGA(4-thio-U)G, AGAGA(4-thio-U)GACUCAG, AGAGAUGA(4-thio-U)CAG, and AGAGAUGA(4-thio-U)CAG] was suspended in 100 mM KPO4 (pH 8), 10% EtOH with an RNA concentration of 1 mM, 3-(2-Iodoacetamido)proxyl (Acros Organics) suspended in this buffer was added to a 100 mM concentration and incubated for 20 h at room temperature with agitation. Reaction progress was followed by checking the 4-thio-U absorbance at 320 nm (Ramos and Varani 1998). The RNA was then desalted. 15N-HSQC spectra were acquired of protein alone and with a 10:1 protein:RNA ratio, and residues were defined as within 15 Å of the spin label if the relevant signal dropped to baseline.

Residual dipolar coupling experiments

Samples of 12C14N RRM2 and 15N RRM23 were brought to 300 μM in 90% NMR buffer and 10% D2O. As needed, lyophilized AGAGAU RNA was added to a final concentration of 4.5 mM. For aligned samples, 170 μL of protein solution were mixed with 130 μL of 15% bicelles (3:1 molar ratio of DMPC/DHPC; Avanti Polar Lipids) by thorough vortexing for a final concentration of 6.5% bicelles. All spectra were taken at 31°C. NH RDC values were obtained using two-dimensional (2D) 15N-HSQC-IPAP, recording peak positions in hertz, and manually calculating the splitting in the nitrogen dimension and the difference between isotropic and aligned values. CH RDC values were obtained for RRM2 bound to AGAGAU RNA using a J-modulated constant time 13C-HSQC and in-house scripts. The program PALES (Zweckstetter and Bax 2000) was used to fit RDC values to structural models.

Restraint generation

Thirteen intermolecular NOEs were identified by analysis of 2D 1H–1H NOESY experiments (150 msec mixing time, obtained on Varian 800 MHz and 900 MHz spectrometers) on samples containing 5 mM 12C14N AGAGAU RNA and 500 μM 13C14N or 13C15N RRM2. Protein peaks in the 13C15N sample were split into two due to coupling, while they remained single peaks in the 13C14N sample, enabling unambiguous identification of protein and RNA peaks. Intermolecular NOEs were identified by searching for cross-peaks between the RNA and protein peaks. The 2D NOEY spectra were also used to identify intra-RNA NOEs.

Intramolecular protein NOEs were predominantly identified through manual inspection of 3D 13C-HSQC- and 15N-HSQC-NOESY spectra taken of 5 mM 12C14N AGAGAU and 500 μM 13C14N RRM2 (900 MHz Varian spectrometer, 50 msec mixing time). NOEs were manually assigned in the 13C-HSQC-NOESY. These NOEs and TALOS (Cornilescu et al. 1999) dihedral angle restraints were used in ATNOS-CANDID version 1.1 to simultaneously assign the 3D 13C-HSQC-NOESY and calculate structures of the RNA-bound form of RRM2 (Herrmann et al. 2002a,b). ATNOS-CANDID produced 20 structures with a backbone RMSD of 1.55 Å. Manual inspection of the 2D NOEY and 3D 13C-HSQC-NOEY yielded further NOEs. NOEs were classified as strong (1.7–4 Å), medium (2.5–5 Å), or weak (2.5–6.5 Å). The ATNOS-CANDID structures were also used to identify secondary structural features and assign intra α-helix and β-sheet hydrogen bonds.

Structure calculation methodology

The relatively low affinity and low solubility of the complex between RRM2 and AGAGAU necessitated some modifications to the usual NMR structure calculation approach. As described above, we first calculated a structure of the protein alone in its RNA-bound form using ATNOS-CANDID. Next, HADDOCK (Dominguez et al. 2003; de Vries et al. 2007) was used to generate initial structures of the complex. HADDOCK was used in order to compensate for the relatively low number of unambiguous
were restrained to a donor-to-acceptor distance of 1.5–2.5 Å. Allows for rotation of freely rotatable groups. Hydrogen bonds and acceptor atoms must be within 3.5 Å in all 10 structures (this was achieved by HADDOCK. To qualify as a hydrogen bond, the donor and acceptor atoms must be within 3.5 Å). Intermolecular hydrogen bonds were identified through manual inspection of the structures, combined with a list generated by HADDOCK analysis scripts. Allowing for local movement subject only to the unambiguous restraints and the force field. Finally, HADDOCK was used to analyze the structures and correct minor planarity distortions introduced by AMBER.

Structure calculation details

HADDOCK version 2.0 (Dominguez et al. 2003; de Vries et al. 2007) was used to dock the lowest-energy ATNOS-CANDID structure of RRM2 to 50 AGAGAU RNAs generated in CNS version 1.21 (Brunger 2007). The AGAGAU structures were randomly generated from an extended chain, using the RNA restraints in Table 1 (dihedral angle restraints are based on allowed RNA conformations) (Saenger 1984; Richardson et al. 2008). The HADDOCK calculations included all restraints in Table 1 except intermolecular hydrogen bonds. Chemical shift perturbation restraints were incorporated as ambiguous restraints between individual protein residues and RNA as recommended by the HADDOCK authors, and nitrooxide spin label–derived restraints were set to be 6–12 Å. HADDOCK docked and refined 100 structures. The 10 lowest-energy structures with no ambiguous NOE or spin label restraint violations >0.5 Å were selected. Intermolecular hydrogen bonds were identified through manual inspection of the structures, combined with a list generated by HADDOCK. To qualify as a hydrogen bond, the donor and acceptor atoms must be within 3.5 Å in all 10 structures (this allows for rotation of freely rotatable groups). Hydrogen bonds were restrained to a donor–to-acceptor distance of 1.5–2.5 Å.

The 10 best HADDOCK structures were refined in AMBER version 9.0 (Case et al. 2006) using a generalized Born solvation model. The refinement had 50,000 steps, with temperature increasing from 0 K to 100 K over the first 5000 steps, decreasing to 0 K over the next 40,000 steps, and remaining at 0 K for 5000 steps. These calculations incorporated all restraints in Table 1 except for chemical shift perturbation. Despite a large weight on the improper term, AMBER introduced some minor planarity distortions in the final structures. These distortions were corrected using a 50-step refinement in HADDOCK. In addition, the HADDOCK analysis scripts were used to analyze the final structures. Ramachandran statistics for the protein were 69% most favored, 21% additionally allowed, 8% generously allowed, and 2% disallowed.

The model of RRM12 bound to AGAGAUGAUCAG was calculated in HADDOCK, starting from the NMR structure of RRM12 and 50 randomly generated AGAGAUGAUCAG RNAs. In addition to the restraints in Table 1, chemical shift perturbation data based on RRM12 binding AGAGAUGAUCAG RNA (Supplemental Fig. 3A) and spin label restraints identified between RRM12 and spin-labeled 12-mer RNAs (Supplemental Fig. 3B–D) were used in the calculation. The lowest-energy structure with no unambiguous NOE violations >0.5 Å was selected.

Models of larger complexes were calculated as follows. Extended chain U6 RNA (nucleotides 49–91) and U4 RNA (nucleotides 59–64) were generated using CNS (Brunger 2007). For the structures of RRM12 bound to U6 nucleotides 49–91, the U6 RNA was first refined in XPLOR-NIH version 2.21 (Schwieters et al. 2003) using the restraints from the structure of the U6 extended ISL (Sashital et al. 2004). Base-pairing restraints to generate the helix between nucleotides 54–60 and 87–91 were also included (planarity restraints, hydrogen-bond restraints between the appropriate Watson–Crick functional groups, and standard A-form helix backbone restraints). HADDOCK (Dominguez et al. 2003; de Vries et al. 2007) was used to dock the NMR structure of RRM12 (Ba et al. 2007) to the U6 structure, incorporating the restraints used to generate the RRM12–AGAGAUGAUCAG complex model as well as the restraints used to calculate the U6 structure. Finally, the lowest-energy HADDOCK structure was refined in XPLOR using the same restraint set, with the coordinates for the protein fixed. (For the structure in Supplemental Fig. 4B, the restraints used to generate the helix between nucleotides 54–60 and 87–91 were removed.) The structure of RRM12 bound to U4 and U6 was generated by manually adding the coordinates of U4 RNA to the structural model of RRM12 bound to AGAGAUGAUCAG, then refining it in XPLOR using the restraints from the RRM12–AGAGAUGAUCAG model and base-pairing restraints to produce a helix between U6 nucleotides 55–60 and U4 nucleotides 59–64.

Molecular graphics programs

All images of structures were prepared in PyMOL (DeLano 2002), while structure alignment was done in MOLMOL (Koradi et al. 1996).

Coordinates

The 10 lowest-energy structures of RRM2 bound to AGAGAU with no NOE violations >0.5 Å were deposited to the Protein Data Bank (PDB) (code 2kh9). Restraint lists used in the calculations have been deposited to the Biological Magnetic Resonance Bank (BMRB) (entry 16230). See Table 1 for relevant statistics.

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Ashley Richie and Marco Tonelli with data collection, Elizabeth Curran with mutagenesis, Gabriel Cornilescu and Anna Füzery with RDC data collection and analysis, and Milo Westler with software. We also appreciate support from and helpful discussions with other members of the Butcher laboratory, Richard Stefl, Frédéric H-T. Allain, and Alexandre M.J.J. Bonvin. This study was funded by NIH grant GM065166, and S.M.-T. was supported by NIH pre-doctoral training grant GM007215. This study made use of the National Magnetic Resonance Facility at The University of Wisconsin–Madison, which is supported by NIH grants P41RR02301 (BRTP/NCRR) and P41GM66326 (NIGMS). Additional equipment was purchased with funds from the University...
REFERENCES


