Functional stabilization of an RNA recognition motif by a noncanonical N-terminal expansion

Catharina Netter, Gert Weber, Heike Benecke, et al.

RNA 2009 15: 1305-1313 originally published online May 15, 2009
Access the most recent version at doi:10.1261/rna.1359909

Functional stabilization of an RNA recognition motif by a noncanonical N-terminal expansion

CATHARINA NETTER,[1] GERT WEBER,[1,2] HEIKE BENECKE,[3] and MARKUS C. WAHL[1,2]

[1]Max-Planck-Institut für Biophysikalische Chemie, D-37077 Göttingen, Germany
[2]Freie Universität Berlin Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, D-14195 Berlin, Germany
[3]Ruhr-Universität Bochum, Lehrstuhl für Zellphysiologie, D-44780 Bochum, Germany

ABSTRACT

RNA recognition motifs (RRMs) constitute versatile macromolecular interaction platforms. They are found in many components of spliceosomes, in which they mediate RNA and protein interactions by diverse molecular strategies. The human U11/U12-65K protein of the minor spliceosome employs a C-terminal RRM to bind hairpin III of the U12 small nuclear RNA (snRNA). This interaction comprises one side of a molecular bridge between the U11 and U12 small nuclear ribonucleoprotein particles (snRNPs) and is reminiscent of the binding of the N-terminal RRMs in the major spliceosomal U1A and U2B″ proteins to hairpins in their cognate snRNAs. Here we show by mutagenesis and electrophoretic mobility shift assays that the β-sheet surface and a neighboring loop of 65K C-terminal RRM are involved in RNA binding, as previously seen in canonical RRMs like the N-terminal RRMs of the U1A and U2B″ proteins. However, unlike U1A and U2B″, some 30 residues N-terminal of the 65K C-terminal RRM core are additionally required for stable U12 snRNA binding. The crystal structure of the expanded 65K C-terminal RRM revealed that the N-terminal tail adopts an α-helical conformation and wraps around the protein toward the face opposite the RNA-binding platform. Point mutations in this part of the protein had only minor effects on RNA affinity. Removal of the N-terminal extension significantly decreased the thermal stability of the 65K C-terminal RRM. These results demonstrate that the 65K C-terminal RRM is augmented by an N-terminal element that confers stability to the domain, and thereby facilitates stable RNA binding.

Keywords: crystal structure; RNA recognition motif (RRM); RNP motif; U11/U12-65K protein; U11/U12 di-snRNP; U1A protein; U2B″ protein; X-ray crystallography

INTRODUCTION

Spliceosomes are multisubunit RNA–protein enzymes that catalyze pre-mRNA splicing, i.e., the removal of noncoding sequences (introns) from precursor messenger RNA (pre-mRNA) and the concomitant ligation of coding regions (exons) (for review, see Burge et al. 1999). Unique among other large, composite enzymes, spliceosomes assemble only in the presence of a substrate. Spliceosome assembly entails the stepwise binding of uridine-rich small nuclear ribonucleoprotein particles (U snRNPs) and numerous non-snRNP splicing factors on a pre-mRNA. After initial buildup of a pre-catalytic particle, major compositional and conformational rearrangements are required to bring about a catalytically activated spliceosome (for review, see Reed and Palandjian 1997; Burge et al. 1999; Brow 2002; Nilsen 2003; Will and Lührmann 2006). Thus, spliceosome function critically relies both on persistent as well as on transient interactions among its building blocks.

Two types of spliceosomes have been identified in higher eukaryotes (for review, see Burge et al. 1999; Will and Lührmann 2005): the U2-dependent (major) spliceosome and the U12-dependent (minor) spliceosome. The major spliceosome excises U2-type introns, which represent the vast majority of pre-mRNA introns. The minor spliceosome removes U12-type introns, which make up less than 1% of all noncoding sequences in humans (Burge et al. 1998; Levine and Durbin 2001). The two types of spliceosomes utilize different, but functionally analogous, snRNPs. The major spliceosomal U1, U2, and U4/U6 snRNPs are replaced, respectively, by the U11, U12, and U4atac/U6atac snRNPs in the minor spliceosome. The U5 snRNP is common to both spliceosomes.

Reprint requests to: Markus C. Wahl, Max-Planck-Institut für Biophysikalische Chemie, AG Makromolekulare Röntgenkristallographie Am Faßberg 11, D-37077 Göttingen, Germany; and Freie Universität Berlin Fachbereich Biologie, Chemie, Pharmazie, Institut für Chemie und Biochemie, AG Strukturbiochemie, Takustraße 6, D-14195 Berlin, Germany; e-mail: mwahl@gwdg.de; fax: 49-30-838-56702.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1359909.

The snRNPs are composed of a unique snRNA, a set of seven Sm or Sm-like (LSm) proteins, and a varying number of particle-specific proteins (for review, see Will and Lührmann 1997). Analogous major and minor spliceosomal snRNPs differ either with respect to their RNA components only (U4/U6 versus U4atac/U6atac) or with respect to both their snRNAs and their specific proteins (U1 and U2 versus U11/U12). Unlike U1 and U2 in the major spliceosome, U11 and U12 form a stable di-snRNP, even in the absence of the pre-mRNA (Frilander and Steitz 1999; Will et al. 2004). While the Sm proteins and the multicomponent splicing factor SF3b are present both in the major U1 and U2 snRNPs and the minor U11/U12 di-snRNP, seven proteins not found in U1 or U2, denoted 65K, 59K, 48K, 35K, 31K, 25K, and 20K, stably interact with the human U11/U12 di-snRNP (Will et al. 2004). Interestingly, some of these U11/U12-specific proteins exhibit similarities to U1-specific proteins. The domain structure of the U1C protein, for example, is mirrored in the U11/U12 20K protein; the U11/U12-specific 35K protein resembles the U1 70K protein and the organization of the U11/U12-65K protein is reminiscent of the U1A and U2B″ proteins (Benecke et al. 2005). These U11/U12-specific proteins may, therefore, functionally replace the respective U1- or U2-specific proteins in the minor spliceosome.

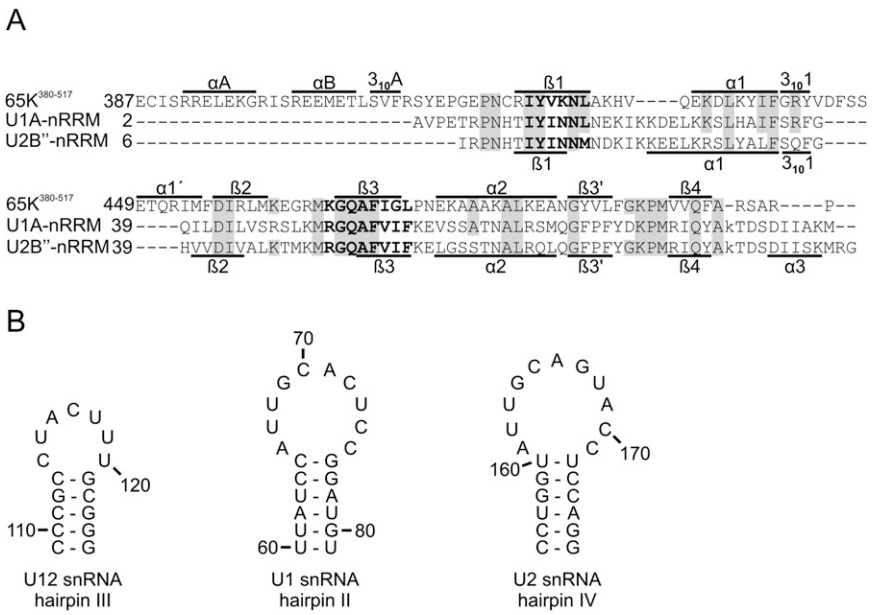

FIGURE 1. (*A*) Structure-based sequence alignment of 65K$^{380–517}$, U1A-nRRM, and U2B″-nRRM. The secondary structure elements (labeled black bars) of 65K$^{380–517}$ are depicted above the alignment, the secondary structure of U1A-nRRM and U2B″-nRRM is shown below the alignment. The RNP1 and RNP2 consensus sequences are shown in bold letters. Invariant residues are shown with a gray background. (*B*) Secondary structure of the RNA ligands of 65K-cRRM (U12 snRNA hairpin III), U1A-nRRM (U1 snRNA hairpin II), and U2B″-nRRM (U2 snRNA hairpin IV).

RNA recognition motifs (RRMs) constitute one of the most versatile macromolecular interaction modules, which can support RNA or protein binding by using diverse interaction surfaces (Maris et al. 2005). Spliceosomes make ample use of these multipurpose interaction modules. For example, the U11/U12-65K protein serves as a bridging factor between the U11 and U12 snRNPs (Benecke et al. 2005). It contains two RRMs connected by a linker that includes a proline-rich region. U11/U12-65K binds to the U11-associated 59K protein via its N-terminal half and to hairpin III of U12 snRNA using its C-terminal RRM (65K-cRRM) (Benecke et al. 2005). 65K-cRRM exhibits high-sequence similarity to the N-terminal RNA recognition motifs of the U1A and U2B″ proteins (U1A-nRRM, U2B″-nRRM) (Fig. 1A), suggesting that the three RRMs have evolved from a common ancestor by gene duplication and subsequent diversification (Bandziulis et al. 1989; Benecke et al. 2005).

The homology of the 65K-cRRM and the U1A/U2B″-nRRMs is paralleled by the similarity of their RNA targets (Fig. 1B). Human 65K-cRRM has been shown to bind to nucleotides 109–125 at the 3′-end of U12 snRNA, which forms hairpin III comprising a 5-base-pair (bp) stem and a 7-nucleotide (nt) loop (Benecke et al. 2005). The U1A and U2B″-nRRMs bind to the U1 snRNA hairpin II and the U2 snRNA hairpin IV, respectively, which exhibit similar global structures, but different sequences (Fig. 1B). Here, we have analyzed the RNA-binding determinants in human 65K-cRRM. Guided by the comparison of our crystal structure of an expanded 65K-cRRM with the U1A and U2B″-nRRMs, we introduced mutations and tested their effects on RNA binding in electrophoretic mobility shift assays (EMSA). Our results reveal that 65K-cRRM employs the same structural elements for RNA binding as the U1A and U2B″-nRRMs. A unique feature of the 65K-cRRM is an N-terminal expansion that apparently forms an integral element of the overall fold. Consistent with this noncanonical architecture, thermal unfolding and RNA-binding experiments suggested that the N-terminal expansion lends stability to the otherwise marginally stable 65K-cRRM structure, which, in turn, bestows the protein with the capacity to stably interact with its target RNA.

RESULTS AND DISCUSSION

The core of the U11/U12-65K C-terminal RRM fails to bind RNA

RRMs are about 80–90 amino acids in length and adopt an α/β-sandwich structure. The minimal domain comprises

four β-strands that form an anti-parallel β-sheet in the spatial order β4-β1-β3-β2 with two α-helices packed against one face of the sheet. In many cases an additional β-hairpin structure (comprising β3′ and β3″) is interspersed between strands β3 and β4. Canonical RRMs bind single-stranded RNA regions by splaying them out across their β-sheet surface. Two highly conserved sequence motifs, RNP1 and RNP2, located on the central strands β3 and β1, respectively, comprise aromatic residues that interact with the nucleic acid bases to build up an "intermolecular hydrophobic core" (Kranz and Hall 1999). Additionally, these residues participate in intramolecular interaction networks, which ensure proper orientation of other parts of the domain (Kranz and Hall 1998, 1999). Loop regions neighboring the β-sheet confer specificity for a particular RNA ligand (Scherly et al. 1990a; Allain et al. 1997; De Guzman et al. 1998; Maris et al. 2005). More recently, a number of RRMs and RRM-related proteins have been shown to bind RNA also by other strategies (Maris et al. 2005; Dominguez and Allain 2006; Skrisovska et al. 2007; Clery et al. 2008).

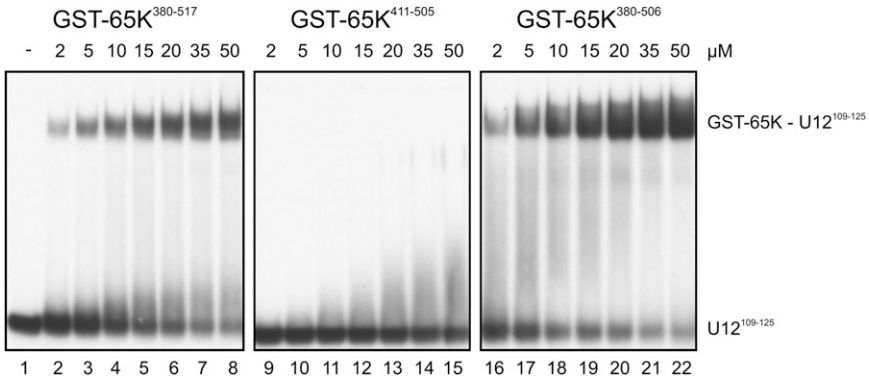

FIGURE 2. A role for the N-terminal expansion of the 65K C-terminal RRM in RNA binding. Band-shift analysis monitoring binding of [^{32}P]-labeled U12 snRNA residues 109–125 (stem–loop III) to GST fusions of 65K$^{380–517}$ (lanes *1–8*), 65K$^{411–505}$ (comprising the RRM core only; lanes *9–15*), and 65K$^{380–506}$ (comprising the RRM core and the N-terminal expansion; lanes *16–22*). The concentration of protein added to each reaction is indicated above each lane.

In the U11/U12-65K protein of the minor spliceosome, a fragment encompassing residues 380–517 (65K$^{380–517}$) was found to bind stem–loop III of the U12 snRNA (Benecke et al. 2005). RNA loop nucleotides and the loop-closing base pair, as well as the presence of a stem-structure, were shown to be required for binding by 65K cRRM (Benecke et al. 2005). It remained unclear, however, which protein elements were involved in the binding of the RNA. 65K$^{380–517}$ contains the 65K-cRRM and additional sequences at both termini. In order to further explore the RNA-binding characteristics of 65K-cRRM and compare them with the homologous U1A-nRRM and U2B″-nRRM, we generated a protein fragment, 65K$^{411–505}$, which encompasses the predicted core RRM fold, but lacks the terminal expansions. Using gel-shift assays, we tested the binding of a N-terminal GST fusion of 65K$^{411–505}$ to a RNA oligomer comprising U12 snRNA residues 109–125 (stem–loop III) (Fig. 1B), which was also used in previous interaction studies with 65K$^{380–517}$ (Benecke et al. 2005). Surprisingly, we failed to detect any binding of the shortened 65K$^{411–505}$ variant to the U12 stem–loop III oligo (Fig. 2, lanes 9–15).

65K$^{411–505}$ lacks 31 residues at the N terminus and 11 residues at the C terminus compared with the expanded 65K$^{380–517}$. We next tested which of these tails facilitates RNA binding. A N-terminal GST fusion of 65K$^{380–506}$ encompassing the core RRM plus the N-terminal expansion efficiently bound the U12 stem–loop III oligo with an apparent K_d of ~9 μM (Fig. 2, lanes 16–22). The C-terminal appendix, additionally contained in fragment 65K$^{380–517}$, failed to further enhance RNA binding (Fig. 2, lanes 1–8). Instead, the C-terminal tail exerted a slight inhibitory effect; 65K$^{380–517}$ bound the RNA with an apparent K_d of around 16 μM (after removal of the tag, we measured an apparent K_d of ~11 μM, comparable to the value reported previously) (Benecke et al. 2005). This inhibitory effect of the C-terminal tail is reminiscent of a number of other RRMs, in which residues C-terminal of the anti-parallel β-sheet fold as a short α-helix that partially occupies the RNA-binding site (Avis et al. 1996). Taken together, these results demonstrate that some 30 residues N-terminal of the 65K-cRRM core are necessary for stable interaction with U12 stem–loop III.

The crystal structure of 65K$^{380–517}$ reveals a N-terminal clamp

Previous studies of RRM-containing proteins failed to detect a requirement for an analogous N-terminal expansion for RNA binding. Among 65K proteins from different species, however, the 30 residues preceding the C-terminal RRM core are highly conserved (for example, the human and *Xenopus laevis* 65K proteins exhibit 68% sequence identity in that region) (Benecke et al. 2005). In order to explore the role of this structurally unique expansion for the function of the 65K-cRRM, we determined the crystal structure of human 65K$^{380–517}$. 65K$^{380–517}$ was crystallized by the sitting drop vapor diffusion technique, yielding small needles with a reservoir comprising PEG3350 and LiCl. The structure was solved by molecular replacement using the structure coordinates of U1A-nRRM (from PDB ID 1URN) (Oubridge et al. 1994) and refined at 2.5 Å resolution. The final model exhibited R_{work}/R_{free} factors of 19.4/25.2% and maintained good stereochemistry

TABLE 1. Crystallographic data and refinement

Data collection	
Wavelength (Å)	0.984
Temperature (K)	100
Space group	C2
Unit cell parameters (Å, °)	
a	91.0
b	33.5
c	49.3
β	96.6
Resolution (Å)	50.0–2.5 (2.59–2.50)[a]
Reflections	
Unique	5299 (527)
Completeness (%)	99.9 (100.0)
Redundancy	3.5 (3.3)
I/σ(I)	9.6 (2.2)
R_{sym}(I)	10.8 (56.3)
Refinement	
Resolution (Å)	30.0–2.5 (2.56–2.50)[a]
Reflections	
Number	5283
Completeness (%)	99.7 (99.5)
Test set (%)	10.7
R_{work}[c]	19.4 (23.6)
R_{free}[c]	25.2 (35.8)
ESU (Å)[d]	0.184
Contents of the asymmetric unit	
Protein molecules/residues/atoms	1/120/982
Water oxygens	84
Mean B-factors (Å^2)	
Wilson	33.8
Protein	32.8
Water	34.6
Ramachandran plot[e] (%)	
Preferred	96.6
Allowed	3.4
Disallowed	0
RMSD[f] from target geometry	
Bond lengths (Å)	0.008
Bond angles (°)	1.18
RMSD B-factors (Å^2)	
Main chain bonds	0.40
Main chain angles	0.65
Side chain bonds	1.12
Side chain angles	1.90
PDB ID	3EGN

[a]Data for the highest resolution shell in parentheses.
[b]$R_{sym}(I) = \Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i|I_i(hkl)|$; for n independent reflections and i observations of a given reflection; $\langle I(hkl)\rangle$, average intensity of the i observations.
[c]$R = \Sigma_{hkl}||F_{obs}|-|F_{calc}||/\Sigma_{hkl}|F_{obs}|$; R_{work} – hkl ∉ T; R_{free} – hkl ∈ T; T, test set.
[d](ESU) Estimated overall coordinate error based on maximum likelihood.
[e]Calculated with Molprobity (http://molprobity.biochem.duke.edu/).
[f](RMSD) Root-mean-square deviation.

(Table 1). Residues 387–506 could be unequivocally traced in the electron density of the protein (Fig. 3A).

In agreement with domain prediction algorithms (de Castro et al. 2006; Letunic et al. 2006), residues 417–501 comprise an archetypal core RRM fold. The minimal motif is augmented by an α-helical insertion (residues 449–454; helix α1′) after the first α-helix (Fig. 3A, cyan). Residues C-terminal of the core RRM lack regular secondary structure and are disordered beyond residue 506. The ordered portion of the C-terminal tail runs along one edge of the β-sheet and covers a peripheral portion of the β-sheet surface (Fig. 3A, black).

The amino acids comprising the N-terminal expansion form two short α-helices (Fig. 3A, helices αA, αB), one 3_{10}-helix (Fig. 3A, 3_{10}A), and a long loop. These elements run along one side of the β-sheet and reach around to the opposite α-helical side of the core RRM like a clamp (Fig. 3A). The N-terminal expansion appears to be stably fastened to the core RRM by extensive hydrophobic interactions and additional hydrogen bonds (Fig. 3B). Overall, 1766 A^2 of combined surface area are covered between the core RRM and the N-terminal expansion (residues 387–417). Thus, the structure suggests that the additional elements are an integral part of the domain fold. A search for structural homologs using the DALI server (Holm and Sander 1993) did not reveal any other known RRM structure with a similar N-terminal appendix.

65K C-terminal RRM employs canonical motifs to bind RNA

Since part of the N-terminal expansion of 65K$^{380–517}$ is neighboring the canonical RNA-binding β-sheet surface (Fig. 4A), it is conceivable that the N terminus is directly involved in RNA binding by the 65K-cRRM. Alternatively, it may play primarily a role in the structural integrity of the protein. To distinguish between these possibilities, we first asked whether the 65K-cRRM employs the canonical RRM structural elements for RNA binding. In U1A-nRRM, mutational and structural studies have revealed residues and regions that are critical for RNA binding (Nagai et al. 1990; Jessen et al. 1991; Oubridge et al. 1994). We therefore identified analogous residues in 65K-cRRM based on our structure-based sequence alignment (Fig. 1A), introduced the corresponding mutations, and tested RNA binding of the mutant proteins by EMSA.

In U1A-nRRM, mutations of aromatic residues on the β-sheet in RNP1 and RNP2 exert a large effect on RNA binding (Nagai et al. 1990; Jessen et al. 1991). Tyr423 from RNP2 of 65K-cRRM corresponds to the essential Tyr13 of U1A (Figs. 1A, 4A). Even the conservative mutation Tyr13Phe abolished RNA binding of U1A-nRRM entirely (Oubridge et al. 1994). We therefore converted Tyr423 of 65K$^{380–517}$ to an alanine and a phenylalanine and tested RNA binding of the mutant proteins. In complete analogy to the U1A-nRRM, both mutations essentially abolished binding to the U12 stem–loop III (Fig. 4B, cf. lanes 1–7 and lanes 8–14). These results confirm that the β-sheet surface of 65K-cRRM is involved in RNA binding.

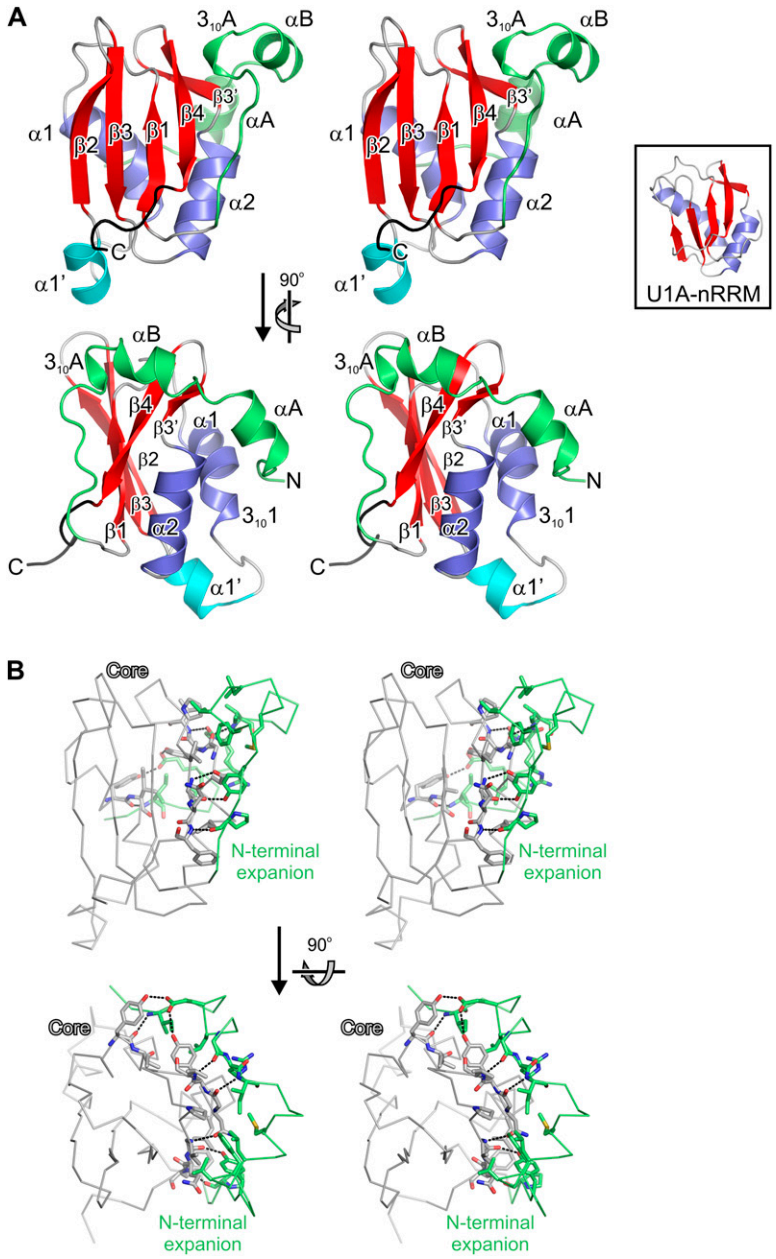

FIGURE 3. Structure of $65K^{380–517}$. (*A*) Stereo ribbon plot of the expanded 65K-cRRM. (*Top*) Front view; (*bottom*) side view. Relative orientations are indicated. Secondary structure elements and termini are labeled. α-helices, blue; β-strands, red. An additional α-helix (α1′) in $65K^{380–517}$ is shown in cyan. The N-terminal expansion (380–417) is in green. The *inset* shows the N-terminal RRM of U1A (Nagai et al. 1990) in the same orientation for comparison. (*B*) Stereo plot illustrating the interaction of the N-terminal expansion (green) with the RRM core (gray). (*Top*) Front view as in the *top* of *A*; (*bottom*) *top* view. Relative orientations are indicated. Interacting side chains are shown as sticks and color coded by atom type. Carbon, as the respective fragments; nitrogen, blue; oxygen, red; sulfur, yellow. Dashed lines indicate hydrogen bonds, most of which involve backbone functionalities.

In many RRMs, less-conserved residues located primarily in loops of variable length that connect the β-strands are responsible for sequence-specific recognition of the RNA targets (Scherly et al. 1990a; Allain et al. 1997; De Guzman et al. 1998). As shown for U1A and U2B′′-nRRMs, the loop between strands β2 and β3 protrudes through the RNA loop and aids in splaying out the single-stranded portion of the RNA. A number of positively charged residues in this loop contact the phosphodiester backbone or the bases of the RNA. In U1A, mutation of Lys50 in the β2–β3 loop to a glutamine leads to a 10- to 50-fold reduction in RNA affinity (Nagai et al. 1990; Oubridge et al. 1994). Lys50 is conservatively replaced by Arg464 in 65K-cRRM (Figs. 1A, 4A). We mutated Arg464 to a glutamine and observed a ~30-fold reduction in the apparent K_d determined by EMSA (Fig. 4B, lanes 15–21).

Arg52 in the β2-β3 loop of U1A interacts with the RNA loop-closing base pair in a sequence-specific manner. Upon mutation to a glutamine, RNA affinity of U1A-nRRM is lost, while the more conservative replacement with a lysine retains partial RNA binding (Nagai et al. 1990; Oubridge et al. 1994). In 65K-cRRM, Arg52 is replaced by Lys466 (Figs. 1A, 4A). We tested the function of Lys466 in 65K-cRRM by introducing a glutamine at this position. Analogous to the situation in U1A, the Lys466Gln mutation entirely abrogated RNA binding, as monitored by gel shifts (Fig. 4B, lanes 22–28).

The β2-β3 loop in 65K-cRRM is shorter (six residues) than the corresponding loops in U1A or U2B′′ (nine residues). However, it has been shown that the length of this loop varies according to the size of the loop of the RNA ligand (Katsamba et al. 2002). Consistently, the RNA target of 65K-cRRM exhibits a correspondingly shorter loop (7 nt) compared with the RNA targets of U1A and U2B′′-nRRMs (10 nt) (Fig. 1B). We conclude that similar to the situation in U1A and U2B′′-nRRMs, the RNA-binding activity of the β2–β3 loop of 65K-cRRM is involved in RNA binding.

Taken together, the above data suggest that 65K-cRRM binds hairpin III of U12 snRNA via structural elements that are also employed by canonical RRMs, such as U1A-nRRM, for RNA binding (i.e., the β-sheet surface and the β2-β3 loop). However, we cannot exclude the possibility that the RNA comes to lie on these

elements in a different orientation as, for example, in U1A-nRRM, and additional regions of 65K-cRRM may contribute to RNA binding. Novel modes of RNA binding have recently been observed in a number of noncanonical RRMs (for a recent review, see Clery et al. 2008).

Notably, the overall RNA-affinity of 65K-cRRM (which lies in the micromolar range) is significantly lower than the affinity of U1A-nRRM to its cognate RNA (which is in the nanomolar range) (Nagai et al. 1990; Hall and Stump 1992; Katsamba et al. 2002). Presently, we cannot pinpoint the exact source for this differential affinity. Most likely it arises from a number of changes on both the protein and the RNA.

The N-terminal expansion influences RNA binding primarily by stabilizing the fold of the core RRM

To test whether the N-terminus directly participates in RNA binding, potential RNA-contacting residues were mutated, and the mutants were tested in gel-shift assays. In the $65K^{380–517}$ structure, part of the N-terminal expansion borders the canonical RNA-binding surface (Fig. 4A). In particular, Val409 and Arg411 from the N-terminal extension may approach an RNA ligand and foster direct contacts (Fig. 4A). We separately mutated Val409 to alanine and Arg411 to glutamine. Both mutations caused a small decrease in RNA affinity (apparent K_d's of the GST-fusion proteins were estimated at ~50 and ~30 μM, respectively). The moderate effects on RNA binding suggest that some residues in the N-terminal extension may directly contact the RNA ligand. However, the putative contact points do not appear to be essential.

We next asked whether the N-terminal expansion is important for the overall stability of the domain. Using a fluorescence-based thermal melting assay (Semisotnov et al. 1991), we compared the stabilities of 65K-cRRM constructs including ($65K^{380–517}$ and $65K^{380–506}$) or lacking ($65K^{411–505}$) the N-terminal

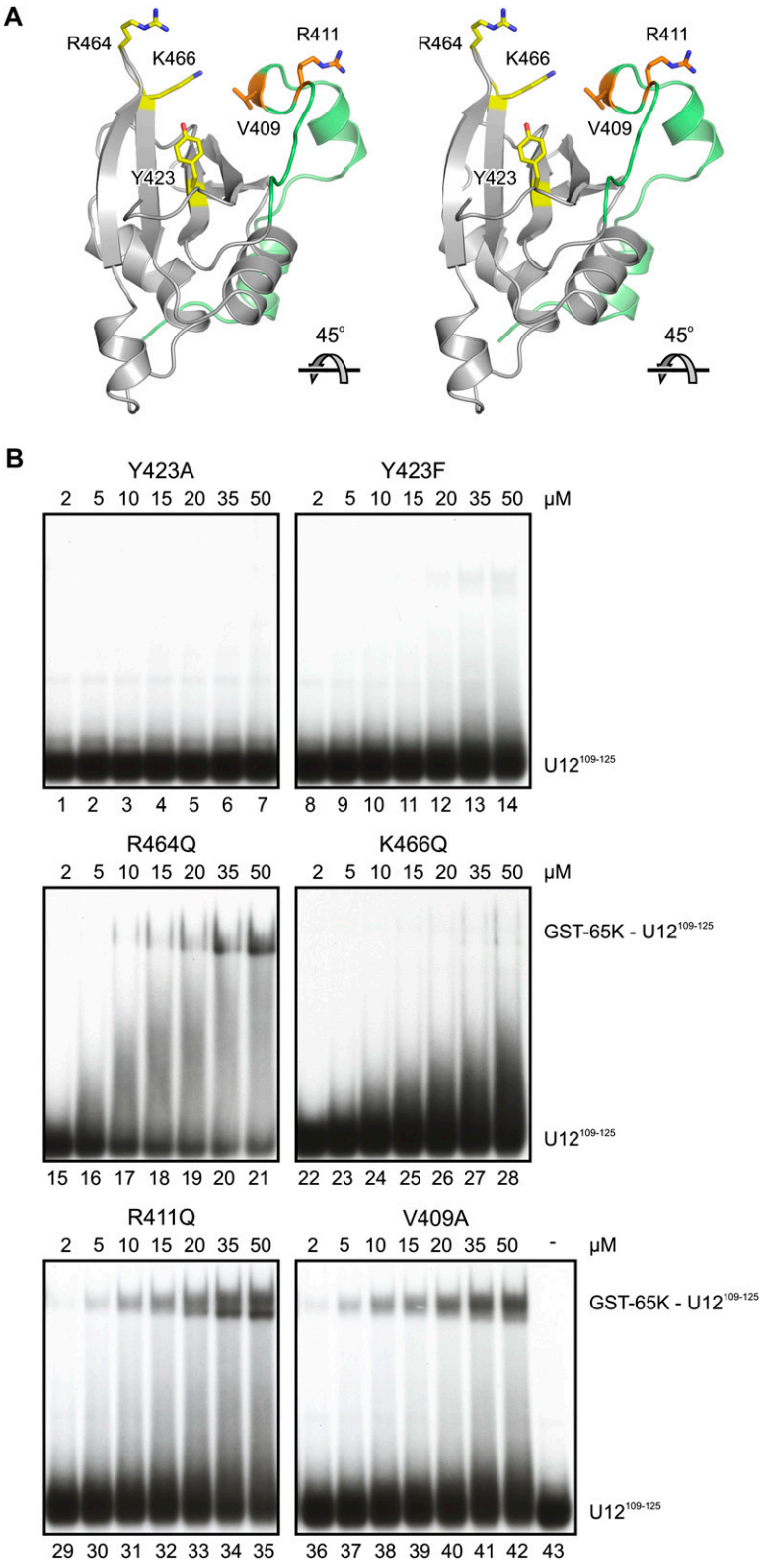

FIGURE 4. (Legend on next page)

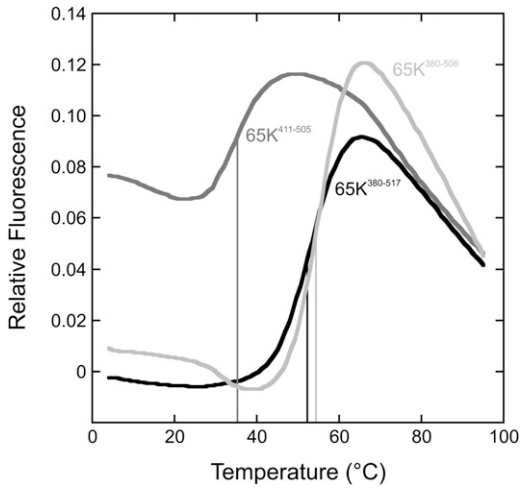

FIGURE 5. Fluorimetric melting analysis. Melting profiles of $65K^{411–505}$ (gray curve), $65K^{380–506}$ (light gray curve), and $65K^{380–517}$ (black curve). The curves were obtained by heating protein samples in the presence of SYPRO Orange and monitoring the fluorescence of the dye. Melting temperatures, indicated by vertical lines, were taken as the midpoints of the cooperative transitions.

expansion (Fig. 5). $65K^{380–517}$ and $65K^{380–506}$ exhibited melting temperatures (T_m) of 52°C and 54°C, respectively. In sharp contrast, $65K^{411–505}$, lacking the N-terminal expansion, melted with a T_m of 36°C. Furthermore, the initial fluorescence of the $65K^{411–505}$ sample was significantly higher than that of $65K^{380–506}$ or $65K^{380–517}$, suggesting that the integrity of its hydrophobic core was corrupted already at lower temperatures. These data show that the N-terminal tail strongly stabilizes the core RRM. We suggest that this stabilization is important for RNA binding by supporting the appropriate folding and orientation of RNA-binding elements.

Outlook: Modifying protein function by domain expansions

Disorder predictions (Linding et al. 2003; Dosztanyi et al. 2005; Prilusky et al. 2005) suggest that the N-terminal expansion of the 65K-cRRM is intrinsically unstructured. Natively unfolded proteins or intrinsically unstructured regions in proteins undergo disorder-to-order transitions upon binding to an interaction partner such as another protein or a nucleic acid ligand (Wright and Dyson 1999). The N-terminal expansion of the 65K-cRRM seems to behave intramolecularly in an analogous fashion. Upon binding to the core domain, it adopts a regular secondary structure.

Recently, the structures of predicted domains in some other spliceosomal proteins were investigated and similar expansions that confer stability on the core domains or safeguard the domains against aggregation have been found (Pena et al. 2007; Zhang et al. 2007). Reminiscent of the present case, these sequence additions are stably grafted onto the protein cores and have apparently become integral elements of the protein folds. For example, the Pml1p protein of the pre-mRNA retention and splicing complex (RES) exhibits a forkhead-associated (FHA) domain in its C-terminal half, while the N-terminal portion is intrinsically unstructured on its own (Trowitzsch et al. 2008). However, in the crystal structure (Trowitzsch et al. 2009), a portion of this N terminus is fixed on one flank of the FHA domain, and deletion of this tail dramatically reduces the solubility of the protein. In another example, a Jab1/MPN domain in the C-terminal region of the large Prp8 protein was found to be augmented by a number of internal insertions and terminal appendices (Pena et al. 2007; Zhang et al. 2007). These additional elements again exhibited stabilizing effects on the core domain. In addition, they endowed the core fold with novel protein–protein interaction capabilities (Pena et al. 2007; Zhang et al. 2007). It remains to be seen whether the N-terminal appendix of the 65K-cRRM also bestows novel functions on the protein. Similar to the Prp8 case, it may facilitate protein–protein interactions, or it may be involved in fine tuning or regulation of RNA binding.

Interestingly, the RNA affinity and the specificity of the nRRM of U2B″ are increased by formation of a complex with U2A′ (Scherly et al. 1990a,b). Structural analysis has shown that U2A′ binds to the side opposite the β-sheet surface of U2B″ and fosters some distal, auxiliary RNA contacts (Price et al. 1998). Thus, the role of U2A′ in supporting U2B″ in *trans* resembles the function of the N-terminal expansion of U11/U12-65K cRRM in *cis*.

MATERIALS AND METHODS

Cloning and expression

A DNA fragment encoding human $65K^{380–517}$ was PCR-amplified and cloned into pGEX-6p-1 (GE Healthcare) using BamHI and XhoI restriction sites. Mutagenesis was performed according to the QuikChange protocol (Stratagene). DNA primers were obtained from MWG Biotech.

For expression, a single colony of transformed *E. coli* Rosetta 2 (DE3) cells was used to inoculate 60 mL of LB medium supplemented with 100 μM Ampicillin and 34 μM Chloramphenicol and incubated overnight at 37°C. Cells were harvested and used to

FIGURE 4. Mutational analysis of $65K^{380–517}$. (*A*) Stereo ribbon plot of $65K^{380–517}$ with the RRM core region in gray and the N-terminal expansion in green. Mutated residues (labeled) are shown as sticks and color coded by atom type. Carbons of mutated residues in the RRM core, yellow; carbons of mutated residues in the N-terminal expansion, orange. The view relative to Figure 2A is indicated. (*B*) Electrophoretic mobility-shift assays monitoring binding of U12 snRNA residues 109–125 (stem–loop III) to GST fusions of $65K^{380–517}$ mutants. Panels are labeled by the mutants used in the assays. The concentrations of the respective protein used are indicated above each lane.

inoculate 6 L of auto-inducing medium (Studier 2005) containing antibiotics. The cultures were grown to an OD_{600} of 0.6 at 37°C, cooled to 16°C, and incubated until cessation of growth. Cells were pelleted and resuspended in phosphate-buffered saline (pH 7.3), 2 mM DTT (PBS buffer) supplemented with 1 tablet/30 mL Complete protease inhibitor (Roche).

Protein purification

Cells were ruptured by sonification, debris was removed by centrifugation, and the soluble fusion protein was captured on glutathione–Sepharose 4 FF beads (GE Healthcare) pre-swollen in PBS buffer. Beads were washed with PBS buffer and subsequently with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM DTT (buffer A). The fusion protein was eluted with buffer A plus 100 mM reduced glutathione (Sigma-Aldrich). The GST-tag was removed by adding 1/40 (mg/mg) Prescission protease and incubating overnight at 4°C. Soluble protein was applied on a heparin Sepharose column (GE Healthcare) and eluted using a salt gradient from 150 mM to 1 M NaCl in buffer A. Peak fractions were pooled and subjected to gel-filtration on a Superdex 75 26/60 column (GE Healthcare) with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM DTT (crystallization buffer). This strategy yielded >98% pure protein as judged by SDS-PAGE analysis. The protein was concentrated to 28 mg/mL using a 5 kDa MWCO concentrator (VIVAscience). Protein concentrations were determined via a Bradford assay (Bio-Rad).

Crystallographic procedures

We aimed at crystallizing 65K-cRRM in complex with RNA. Therefore, the protein was supplemented with 1.5 mM $MgCl_2$ (buffer B) and incubated for 1 h with an RNA oligonucleotide mimicking hairpin III of U12 snRNA (5′-CCCGCCUACUUUGCGGG-3′). Crystals were grown by the sitting drop vapor diffusion technique with a reservoir containing 20% (w/v) PEG3350 and 0.2 M LiCl.

For diffraction data collection, crystals were transferred to mother liquor plus 15% (v/v) propylene glycol and flash-frozen in liquid nitrogen. Data were collected at 100 K at the PXI beamline of SLS (Villigen) using a microfocused beam. Data were processed with the HKL2000 suite of programs (Otwinowski and Minor 1997).

Crystals belonged to space-group C2 and exhibited a solvent content of 47%, assuming one 65K-cRRM molecule without RNA ligand per asymmetric unit. The structure was solved by molecular replacement with the program Molrep (Vagin and Teplyakov 2000) using the structure coordinates of the U1A-nRRM (from PDB ID 1URN) (Oubridge et al. 1994). Iterative cycles of manual model building in Coot (Emsley and Cowtan 2004) and automatic refinement using the programs Phenix (Adams et al. 2002) and REFMAC5 (Murshudov et al. 1997) yielded the final model (Table 1).

The structure factors and coordinates have been deposited in the Protein Data Bank (http://www.pdb.org) under entry code 3EGN, and will be released upon publication.

Electrophoretic mobility shift assays (EMSA)

EMSA experiments were conducted in duplicates. A total of 0–300 pmol untagged protein were incubated in buffer B with 0.25 pmol of [^{32}P]-end-labeled RNA oligonucleotide (5′-CCCGCCUACUUUGCGGG-3′), 1 μL of 10 mg/mL of *E. coli* tRNA, and 0.25 μL of RNasin (Promega) in a final volume of 10 μL for 45 min at 4°C. After addition of 6 μL of loading dye (50% glycerol, 0.5 x TBE buffer), RNA–protein complexes were separated on a native 6% (29:1 acrylamide:bis-acrylamide) polyacrylamide gel at 8 W and 4°C. For GST-tagged proteins, the same procedure was applied, but mixtures were incubated in 20 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 0.2 mM EDTA, 200 mM KCl, 0.1% Triton X-100.

Bands were visualized by autoradiography. Quantification of the bands was performed with ImageQuant (GE Healthcare). Apparent K_d-values were estimated from the protein concentrations required to elicit a 50% shift of the RNA.

Thermal denaturation experiments

A total of 18 μL of purified proteins at 30 μM concentration in 20 mM Na/KPO_4 (pH 7.5), 1.5 mM $MgCl_2$, 150 mM NaF, 2 mM DTT were mixed with 2 μL of 5× SYPRO Orange (Sigma-Aldrich). The changes in fluorescence of the samples as a function of temperature (4°C–95°C in steps of 1°C/min) were monitored in a real-time PCR machine. The fluorescence was recorded once per degree with 30-sec intervals between reads.

ACKNOWLEDGMENTS

We thank Reinhard Lührmann for generous support, Elke Penka for excellent technical assistance, the team at the beamline PXI (SLS, Villigen, Switzerland) for support during data collection, Cindy Will for critical reading of the manuscript, and members of our group for helpful discussions. This work was supported by a Ph.D. scholarship from the Boehringer Ingelheim Fonds (to C.N.) and the Max-Planck-Society (to M.C.W.).

Received September 11, 2008; accepted March 16, 2009.

REFERENCES

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. 2002. PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58:** 1948–1954.

Allain FH, Howe PW, Neuhaus D, Varani G. 1997. Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J* **16:** 5764–5772.

Avis JM, Allain FH, Howe PW, Varani G, Nagai K, Neuhaus D. 1996. Solution structure of the N-terminal RNP domain of U1A protein: The role of C-terminal residues in structure stability and RNA binding. *J Mol Biol* **257:** 398–411.

Bandziulis RJ, Swanson MS, Dreyfuss G. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev* **3:** 431–437.

Benecke H, Lührmann R, Will CL. 2005. The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *EMBO J* **24:** 3057–3069.

Brow DA. 2002. Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36:** 333–360.

Burge CB, Padgett RA, Sharp PA. 1998. Evolutionary fates and origins of U12-type introns. *Mol Cell* **2:** 773–785.

Burge CB, Query CC, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In: *The RNA world*, 2nd ed. (eds. RF Gesteland et al.) pp 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Clery A, Blatter M, Allain FH. 2008. RNA recognition motifs: Boring? Not quite. *Curr Opin Struct Biol* **18:** 290–298.

de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N. 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* **34:** W362–W365.

De Guzman RN, Turner RB, Summers MF. 1998. Protein–RNA recognition. *Biopolymers* **48:** 181–195.

Dominguez C, Allain FH. 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: A novel mode of RNA recognition. *Nucleic Acids Res* **34:** 3634–3645.

Dosztanyi Z, Csizmok V, Tompa P, Simon I. 2005. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21:** 3433–3434.

Emsley P, Cowtan K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

Frilander MJ, Steitz JA. 1999. Initial recognition of U12-dependent introns requires both U11/5′ splice-site and U12/branchpoint interactions. *Genes & Dev* **13:** 851–863.

Hall KB, Stump WT. 1992. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res* **20:** 4283–4290.

Holm L, Sander C. 1993. Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233:** 123–138.

Jessen TH, Oubridge C, Teo CH, Pritchard C, Nagai K. 1991. Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J* **10:** 3447–3456.

Katsamba PS, Bayramyan M, Haworth IS, Myszka DG, Laird-Offringa IA. 2002. Complex role of the β 2-β 3 loop in the interaction of U1A with U1 hairpin II RNA. *J Biol Chem* **277:** 33267–33274.

Kranz JK, Hall KB. 1998. RNA binding mediates the local cooperativity between the β-sheet and the C-terminal tail of the human U1A RBD1 protein. *J Mol Biol* **275:** 465–481.

Kranz JK, Hall KB. 1999. RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal binding surface. *J Mol Biol* **285:** 215–231.

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. 2006. SMART 5: Domains in the context of genomes and networks. *Nucleic Acids Res* **34:** D257–D260.

Levine A, Durbin R. 2001. A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res* **29:** 4006–4013.

Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. 2003. Protein disorder prediction: Implications for structural proteomics. *Structure* **11:** 1453–1459.

Maris C, Dominguez C, Allain FH. 2005. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272:** 2118–2131.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.

Nagai K, Oubridge C, Jessen TH, Li J, Evans PR. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348:** 515–520.

Nilsen TW. 2003. The spliceosome: The most complex macromolecular machine in the cell? *Bioessays* **25:** 1147–1149.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276:** 307–326.

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372:** 432–438.

Pena V, Liu S, Bujnicki JM, Lührmann R, Wahl MC. 2007. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. *Mol Cell* **25:** 615–624.

Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394:** 645–650.

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. 2005. FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21:** 3435–3438.

Reed R, Palandjian L. 1997. Spliceosome assembly. In: *Eukaryotic mRNA processing* (ed. AR Krainer), pp 103–129. IRL Press, Oxford, UK.

Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990a. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. *Nature* **345:** 502–506.

Scherly D, Dathan NA, Boelens W, van Venrooij WJ, Mattaj IW. 1990b. The U2B″ RNP motif as a site of protein–protein interaction. *EMBO J* **9:** 3675–3681.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI. 1991. Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31:** 119–128.

Skrisovska L, Bourgeois CF, Stefl R, Grellscheid SN, Kister L, Wenter P, Elliott DJ, Stevenin J, Allain FH. 2007. The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep* **8:** 372–379.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41:** 207–234.

Trowitzsch S, Weber G, Lührmann R, Wahl MC. 2008. An unusual RNA recognition motif acts as a scaffold for multiple proteins in the pre-mRNA retention and splicing complex. *J Biol Chem* **283:** 32317–32327.

Trowitzsch S, Weber G, Lührmann R, Wahl MC. 2009. Crystal structure of the Pml1p subunit of the yeast precursor mRNA retention and splicing complex. *J Mol Biol* **385:** 531–541.

Vagin A, Teplyakov A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56:** 1622–1624.

Will CL, Lührmann R. 1997. snRNP structure and function. In: *Eukaryotic mRNA processing* (ed. AR Krainer), pp 130–173. IRL Press, Oxford, UK.

Will CL, Lührmann R. 2005. Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem* **386:** 713–724.

Will CL, Lührmann R. 2006. Spliceosome structure and function. In: *The RNA world*, 3rd ed. (eds. RF Gesteland et al.), pp 369–400. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Will CL, Schneider C, Hossbach M, Urlaub H, Rauhut R, Elbashir S, Tuschl T, Lührmann R. 2004. The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* **10:** 929–941.

Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure–function paradigm. *J Mol Biol* **293:** 321–331.

Zhang L, Shen J, Guarnieri MT, Heroux A, Yang K, Zhao R. 2007. Crystal structure of the C-terminal domain of splicing factor Prp8 carrying retinitis pigmentosa mutants. *Protein Sci* **16:** 1024–1031.