Crystal structure of a core spliceosomal protein interface

Matthew J. Schellenberg*, Ross A. Edwards*, Dustin B. Ritchie*, Oliver A. Kent*, Monika M. Golas†, Holger Stark†, Reinhard Lührmann†, J. N. Mark Glover*‡, and Andrew M. MacMillan*‡

*Department of Biochemistry, University of Alberta, Edmonton, AB, Canada T6G 2H7; and †Department of Cellular Biochemistry and Research, Group of 3D Electron Cryomicroscopy, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

Edited by Joan A. Steitz, Yale University, New Haven, CT, and approved December 7, 2005 (received for review September 14, 2005)

The precise excision of introns from precursor mRNAs (pre-mRNAs) in eukaryotes is accomplished by the spliceosome, a complex assembly containing five small nuclear ribonucleoprotein (snRNP) particles. Human p14, a component of the spliceosomal U2 and U11/U12 snRNPs, has been shown to associate directly with the pre-mRNA branch adenosine early in spliceosome assembly and within the fully assembled spliceosome. Here we report the 2.5-Å crystal structure of a complex containing p14 and a peptide derived from the p14-associated U2 snRNP component SF3b155. p14 contains an RNA recognition motif (RRM), the surface of which is largely occluded by a C-terminal α-helix and a portion of the SF3b155 peptide. An analysis of RNA·protein crosslinking to wild-type and mutant p14 shows that the branch adenosine directly interacts with a conserved aromatic within a pocket on the surface of the complex. This result, combined with a comparison of the structure with known RRMs and pseudoRRMs as well as model-building by using the electron cryomicroscopy structure of a spliceosomal U11/U12 di-snRNP, suggests that p14·SF3b155 presents a noncanonical surface for RNA recognition at the heart of the mammalian spliceosome.

RNA-binding proteins | RNA splicing | spliceosome

Precursor mRNA (pre-mRNA) splicing occurs through two sequential transesterification reactions. The first step involves displacement of the 5′ exon by the nucleophilic attack of the 2′ hydroxyl of the conserved branch region adenosine. In the second step, the free 5′ exon attacks the 3′ splice site to yield the ligated mRNA product. Assembly of the spliceosomal catalytic machinery containing the U1, U2, and U4/U5/U6 small nuclear ribonucleoproteins (snRNPs) is directed by conserved sequence elements at the splice sites and within the intron including the branch region (1–3). Spliceosome assembly occurs in an ordered, stepwise fashion through a series of discrete intermediates and involves sequential steps of pre-mRNA recognition by protein and snRNA components of the spliceosome as well as formation and rearrangement of snRNA·snRNA interactions. The fully assembled spliceosome contains a U2/U6 snRNA structure that has been proposed to form the active site for catalysis of the transesterifications (4, 5). Although pre-mRNA splicing is therefore believed to be intrinsically RNA catalyzed, RNA·protein crosslinking studies have established direct association of human p14 with the reactive branch adenosine of the pre-mRNA (6, 7). Thus, the catalytic heart of the spliceosome may include protein as well as RNA components.

Initial recognition of the pre-mRNA includes a stable U1 snRNP·5′ splice-site base pairing and the association of non-snRNP protein factors with the branch region/3′ splice site to form the early or commitment complex. The ATP-dependent transition to A complex involves the stable association of U2 snRNP with the pre-mRNA, an interaction that includes the formation of a duplex between U2 snRNA and the pre-mRNA branch region. The bulging of the branch adenosine from this duplex specifies it as the nucleophile for the first step of splicing (8); thus, a structure critical to catalysis involving recognition of the pre-mRNA by components of U2 snRNP is established early in spliceosome assembly. Final steps in this assembly include association of the U4/U5/U6 tri-snRNP with the pre-mRNA to produce B complex, which undergoes a series of rearrangements to yield C complex, the fully assembled spliceosome.

The dynamic nature of spliceosome assembly has been established by a wealth of studies that have described a complex series of RNA·RNA and RNA·protein associations during the course of assembly (1–3). Site-specific modification of the pre-mRNA branch adenosine with benzophenone followed by photocrosslinking revealed the specific, sequential association of several factors with the branch adenosine from very early in spliceosome assembly (6). Of particular interest was a strong crosslink to p14 that appeared in A complex and persisted within the fully assembled spliceosome. Subsequent experiments showed that this protein crosslinked directly to the branch nucleotide in A through C complexes indicating an intimate association between protein and RNA at the heart of the mammalian spliceosome (7). This finding contrasts with the ribosome in which high-resolution structural analysis has shown that the active site is exclusively composed of RNA with the closest ribosomal protein found ≈18 Å distant from the active site (9).

p14 was initially isolated from purified mammalian spliceosomes and subsequently identified as a constituent of U2 snRNP and U12 snRNP, a component of the minor spliceosome responsible for splicing a relatively rare subset of introns (10). p14 is an evolutionarily highly conserved protein with orthologs across diverse species; the human protein is a 125-aa polypeptide containing a central region with strong homology to the well characterized RNA-recognition motif (RRM) domain (11). In particular, p14 contains consensus RNP1 and RNP2 motifs, which mediate RRM·RNA interactions, as evidenced by high-resolution structural analyses (12–16). Despite the presence of an RRM, the association of p14 with U2 snRNP is mediated, at least in part, through protein·protein interactions. Specifically, p14 is part of the heteromeric protein complex SF3b, a salt-dissociable component of U2 snRNP, and a strong interaction between p14 and the SF3b protein SF3b155 has been demonstrated (10). The intimate association of p14 with the branch nucleotide in the fully assembled spliceosome suggests that this protein may contribute to the architecture of the active site.

To further define the role of p14 in the spliceosome, we have solved the high-resolution x-ray structure of human p14 in a complex with a portion of SF3b155. Analysis of this structure

---

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: pre-mRNA, precursor mRNA; snRNP, small nuclear ribonucleoprotein; cryo-EM, electron cryomicroscopy; RRM, RNA recognition motif; MBP, maltose-binding protein.

Data deposition: The atomic coordinates for the wild-type and mutant complexes have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 2F9D and 2F9J, respectively).

‡To whom correspondence may be addressed. E-mail: andrew.macmillan@ualberta.ca or mark.glover@ualberta.ca.

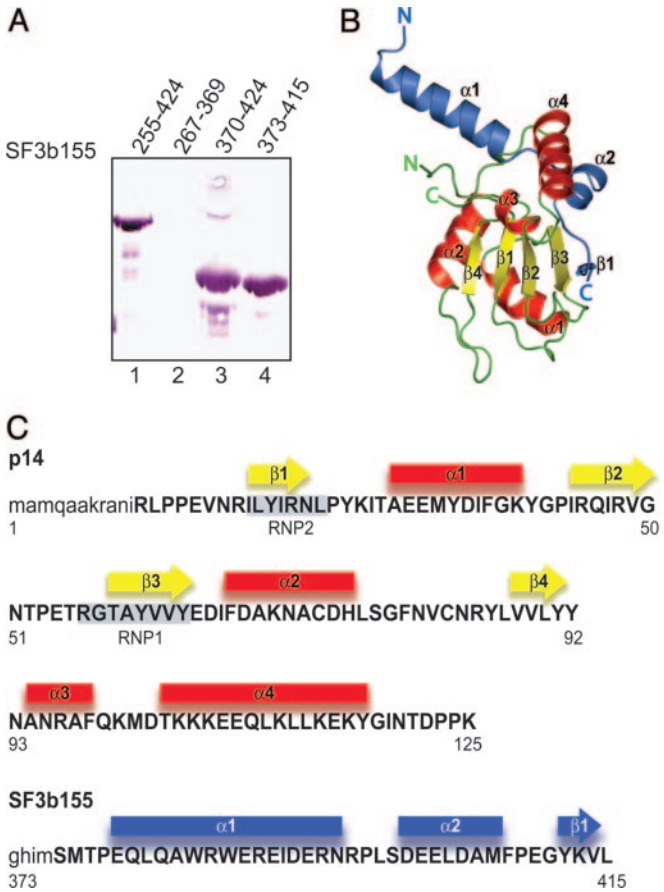

Fig. 1. Structure of p14·SF3b155 peptide complex. (*A*) Determination of a minimal interaction surface between p14 and SF3b155. Pulldowns using amylose resin of the indicated GST-deletion constructs of SF3b155 after incubation with MBP-p14. Reactions were analyzed by 16% 29:1 SDS/PAGE using anti-GST horseradish peroxidase conjugate to detect SF3b155. (*B*) Ribbon diagram of p14-SF3b155 peptide complex. α-Helices and β-strands of p14 are colored red and yellow, respectively; SF3b155 peptide is colored blue. (*C*) Secondary structure diagram depicting α-helices (red) and β-strands (yellow) of p14. RNP motifs are highlighted with gray boxes. Secondary structural elements of SF3b155 peptide are colored blue.

combined with previous crosslinking data and analyses of extant RRM structures and electron cryomicroscopy (cryo-EM) data allow us to propose a model for branch region recognition in the fully assembled spliceosome.

Results

Structure of the p14·SF3b155 Complex. Initial investigations of the properties of p14 suggested that the protein in isolation was partially unfolded as evidenced by poor solubility, aggregation, and ^{1}H NMR (M.J.S., L. Spyrocopolous, and A.M.M., unpublished data). We therefore focused further studies on the complex formed between p14 and SF3b155 (10). By using a pulldown assay to examine a series of deletion constructs, we were able to establish the existence of a strong interaction between p14 and a peptide representing amino acids 373–415 of SF3b155 (Fig. 1*A*). The complex formed between p14 and this peptide is very soluble and elutes by gel filtration at the volume expected for a 1:1 complex. We crystallized a selenomethionine-substituted complex and determined its structure to 2.5 Å resolution by using multiwavelength anomalous diffraction methods (Fig. 1*B* and Table 1).

As predicted, p14 contains a central RRM domain spanning residues 20–91 (Fig. 1*C*). In addition, the C terminus of p14

Table 1. Data collection, phasing, and refinement statistics

	SeMet Derivative		Y22M Mutant
Data collection			
Space group	*C* 2 2 21		*P* 4_1
Cell dimensions			
a/*b*/*c*, Å	101.87/115.23/ 82.54		71.27/71.27/ 102.28
α/β/γ, °	90/90/90		90/90/90
	Remote	Peak	Native
Wavelength	1.019859	0.979547	1.115869
Resolution, Å	2.5	3.1	3.0
R_{sym} or R_{merge}	0.076 (0.399)	0.090 (0.317)	0.0059 (0.460)
$I/\sigma I$	25.5 (2.4)	17.2 (5.0)	23.3 (3.3)
Completeness, %	99.5 (95.6)	100	99.8 (100)
Redundancy	7.6	4.5	4.2
Refinement			
Resolution, Å	76–2.50		71–3.0
No. of reflections	16225		10288
R_{work}/R_{free}	0.221/0.278		0.227/0.290
No. of atoms			
Protein	2560		2416
Ligand/ion	0		0
Water	67		14
β-factors			
Protein	45.3		76.4
Ligand/ion	NA		NA
Water	42.5		53.5
rms deviations			
Bond lengths, Å	0.014		0.011
Bond angles, °	1.56		1.35

Peak and remote data for selenomethionine (SeMet) were collected from a single crystal. Refinement was performed against the remote data. Values in parentheses are for the highest-resolution shell (SeMet, 2.59–2.50 for remote and 3.21–3.10 for peak; Y22M mutant, 3.11–3.00). NA, not applicable.

contains two additional α-helices: a short 5-aa helix (amino acids 94–98) and a second 15-residue helix (amino acids 103–117). The bound SF3b155 peptide consists of a long N-terminal α-helix (amino acids 380–396) and a second shorter helix (amino acids 401–407). Interestingly, the C terminus of the SF3b155 fragment contains a short β-strand that interacts with β-3 of the p14 RRM β-sheet, which is connected to the shorter α-helix by a loop that makes extensive contacts with both the shorter C-terminal helix and RRM of p14. Amino Acids 1–11 of p14 and 373–376 of SF3b155 are disordered in the structure. Because the crosslinks formed between spliceosomal versus recombinant p14 and RNA differ in size by ≈1 kDa, we speculate that spliceosomal p14 lacks ≈10 aa of the predicted full-length protein; therefore, the former residues are probably not relevant to the structure of p14 within the spliceosome (M.J.S. and A.M.M., unpublished data).

The most striking feature of the p14·peptide complex is that the p14 β-sheet is occluded by one of the C-terminal helices of p14, the central helix of the SF3b155 peptide, and the loop connecting SF3b155 to the C-terminal β-strand, resulting in a buried surface area of ≈1,300 Å. The interface between the two proteins is extensive and includes a hydrophobic core (Fig. 2*A*) surrounded by a set of hydrogen bonds and salt bridges (Fig. 2*B*). The blocking of one face of p14 is significant because the four-stranded β-sheet of the canonical RRM represents the RNA-binding surface of the domain including the highly conserved RNP1 and RNP2 motifs (11–16). In the p14·peptide complex, residues of RNP1 and RNP2 are largely buried. Intriguingly, a portion of RNP2 is exposed within a pocket on the otherwise occluded surface: a highly conserved aromatic residue

BIOPHYSICS

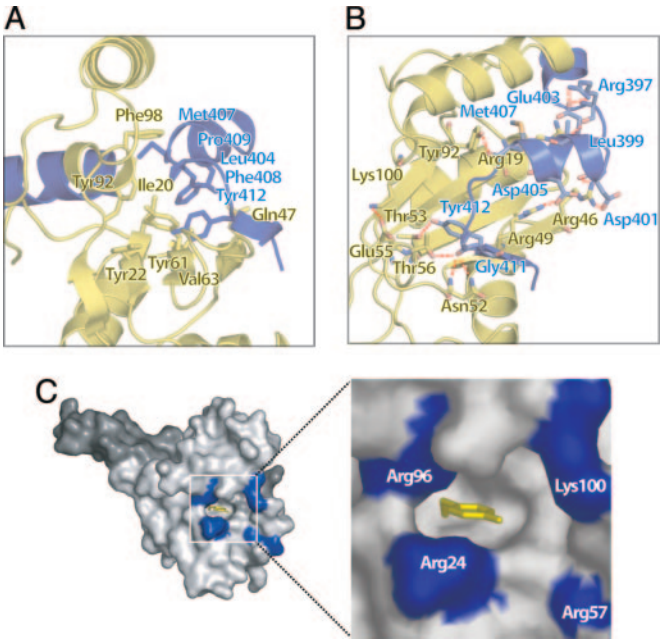

Fig. 2. Details of p14·SF3b155 interface. (*A*) Hydrophobic core of the p14·SF3b155 interface. p14 is colored yellow, and SF3b155 peptide is colored blue. (*B*) Hydrogen-bonding network and salt bridges surrounding the hydrophobic core. p14 is colored yellow, and SF3b155 peptide is colored blue. Hydrogen bonds involved in secondary structural elements are omitted. (*C*) Surface representation of the p14·SF3b155 complex showing Y22 of RNP2 exposed within a surface pocket surrounded by conserved basic residues. p14 is shaded light gray, SF3b is shaded dark gray, Y22 is colored yellow, and R24, R57, R96, and K100 are colored blue.

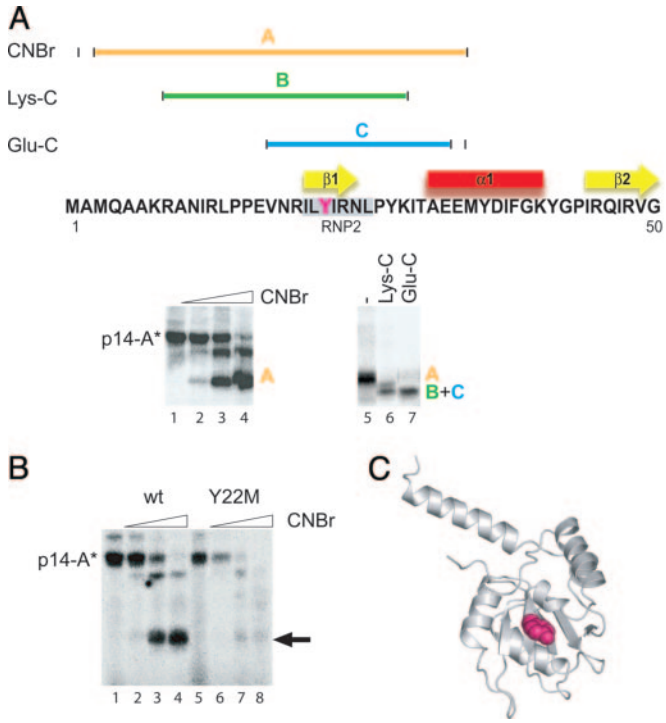

Fig. 3. Mapping of the p14·branch nucleotide interaction. (*A*) Treatment of p14·SF3b155 peptide complex–RNA crosslink with nuclease P1 yields a single radiolabeled band on SDS/PAGE corresponding to p14 (lane 1). Cleavage with CNBr produces fragment A, which appears immediately (lane 2) and persists as a final digestion product (lanes 3 and 4). Further digestion of purified CNBr fragment A (lane 5) with endoproteinase Lys-C produces fragment B (lane 6). Alternatively, cleavage of fragment A with endoproteinase Glu-C produces fragment C. Cleavage with endoproteinase Glu-C of a Δ1–11 deletion mutant of p14 produces an identically migrating fragment, indicating that fragment C is not the N-terminal Glu-C fragment of p14 (data not shown). (*B*) CNBr treatment of p14·RNA (lanes 1–4) and Y22M p14·RNA (lanes 5–8) crosslinks. (*C*) Ribbon diagram of p14·SF3b155 peptide complex with Y22 depicted in a space-filling representation (purple).

within RNP2, Y22, forms the base of this pocket on the p14·peptide surface (Fig. 2*C*). The surface surrounding this pocket includes four basic residues: R24, R57, R96, and K100 (Fig. 2*C*). The side chains of two of these residues (R24 and R57) project from the surface of the p14 β-sheet, the third (R96) is found at the end of α3, and the fourth (K100) in the loop between α3 and α4. The identities of R24 and R57 are highly conserved among p14 orthologs but not between p14 and other RRMs.

RNA Protein Interactions in a Minimal Complex. The direct interaction of p14 with the branch nucleotide bulged from a base-paired duplex is intriguing because in all previously characterized RNA·RRM complexes, single-stranded RNA binds across the β-sheet of the RRM (11–16). Mobility shift assays using either p14 or the p14·peptide complex show only weak RNA·protein association (>50 μM dissociation constants) and no preference for single-stranded, duplex, or bulged-duplex RNA (M.J.S., A.M.M., unpublished data). These results probably reflect the cooperative nature of RNA·protein interaction within the spliceosome, including interactions of the pre-mRNA with SF3b155 (17); whereas p14 is the only protein that directly interacts with the branch adenosine in the fully assembled spliceosome, as evidenced by photocrosslinking studies (7), SF3b155 has been shown to directly interact with nucleotides at the −6 position, just 5′ to the branch region·U2 duplex and at the +5 position, 3′ to the branch (17).

To determine the part of p14 that interacts with the branch adenosine, we mapped the location of the crosslink formed between the branch nucleotide and p14. We first synthesized an RNA containing a hairpin representing the pre-mRNA branch region duplex; this RNA contained a single adenosine at the branch position and could thus be uniquely labeled at this position by carrying out transcriptions in the presence of [α-^{32}P]ATP. We then performed crosslinking experiments in a minimal system containing this RNA and the p14·peptide complex. After crosslinking by irradiation at 254 nm, the reactions were digested with Nuclease P1 and analyzed by SDS/PAGE, which showed that the branch adenosine crosslinks to p14 but not the SF3b155 peptide (Fig. 3*A* and data not shown). By using cyanogen bromide and endoproteases Glu-C and Lys-C, we were able to establish that the branch nucleotide crosslinks to p14 between amino acids 16 and 29 (Fig. 3*A*); this region includes β-1 of the RRM and the RNP2 consensus region. Thus, the pre-mRNA branch interacts with a portion of the RRM of p14.

We next examined the interaction with RNA of a mutant p14, Y22M, in which the conserved aromatic of RNP2 has been changed to methionine. This mutation has little effect on the structure of the p14·peptide complex; the 3.0-Å x-ray structure of the Y22M p14·SF3b peptide complex is essentially superimposable on the native complex, with the side chain of M22 forming the base of the pocket in the same fashion as Y22 in the wild-type (Table 1 and data not shown). We performed crosslinking experiments in reactions containing the Y22M p14·peptide complex and the bulged duplex RNA and observed the formation of a crosslink between the branch nucleotide and Y22M p14 (Fig. 3*B*). When we attempted to perform a crosslink mapping experiment, we observed that the crosslink was sensitive to cyanogen bromide, as evidenced by the disappearance of labeled protein upon treatment with the reagent (Fig. 3*B*). This

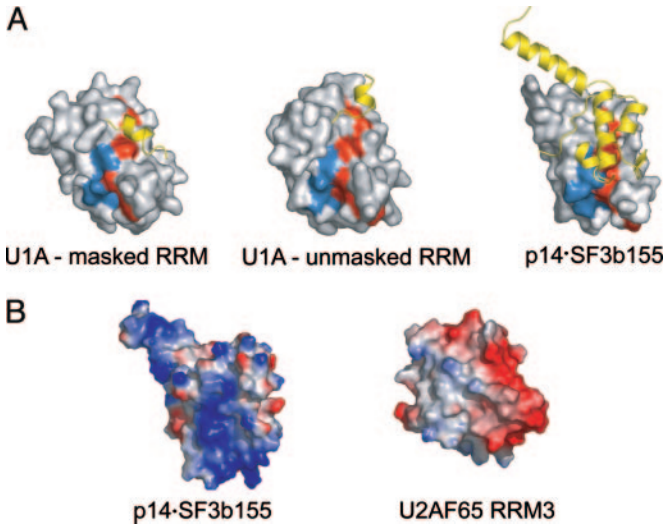

Fig. 4. Comparison between p14·SF3b155 peptide surface and structures of U1A and U2AF65 RRM3. (*A*) (*Left*) NMR structure of free U1A with C-terminal helix (in yellow) positioned across RNA-binding surface (19). (*Middle*) X-ray structure of free U1A showing C-terminal helix rotated to unmask the RNA binding surface (20). (*Right*) Structure of p14·SF3b155 complex showing C-terminal helices of p14 and Sf3b155 peptide (in yellow). RNP1 and RNP2 are colored red and blue, respectively. (*B*) Representation of surface charges of p14·SF3b155 and U2AF homology motif (RRM3) of U2AF^{65}. The color scheme is as follows: blue, basic; red, acidic; white, neutral.

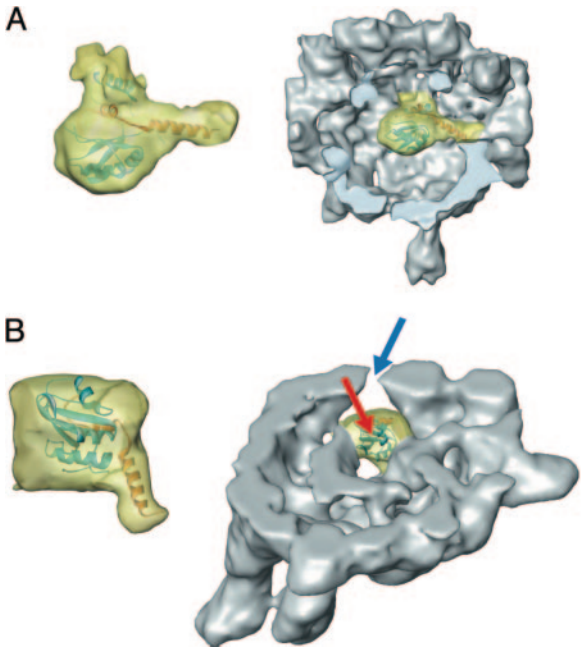

Fig. 5. p14·SF3b155 peptide structure within SF3b and U11/U12 snRNP. (*A*) Fitting of p14·SF3b155 peptide density into the SF3b cryo-EM structure (24). (*Left*) cryo-EM density corresponding to p14·SF3b155 peptide (yellow), with the p14·SF3b155 x-ray structure represented as a ribbon diagram. (*Right*) cryo-EM density corresponding to p14·SF3b155 peptide within the context of a cutaway view of the overall SF3b cryo-EM structure. (*B*) Fitting of p14·SF3b155 peptide density into the U11/U12 cryo-EM structure (25). (*Left*) cryo-EM density corresponding to p14·SF3b155 peptide (yellow), with the p14·SF3b155 x-ray structure represented as a ribbon diagram. (*Right*) cryo-EM density corresponding to p14·SF3b155 peptide within the context of a cutaway view of the overall U11/U12 cryo-EM structure. The red arrow indicates the proposed RNA-binding surface of p14·SF3b155; the blue arrow indicates a possible path of pre-mRNA to groove on U11/U12 surface.

result is consistent with the crosslink occurring between the branch nucleotide and the terminal methyl group of the side chain of M22 (18), which suggests that the conserved aromatic of RNP2 interacts directly with the bulged nucleotide (Fig. 3*C*). Given the fact that single-stranded RNA containing a single adenosine crosslinks to p14 in the same fashion, it is also possible that the p14·peptide complex can recognize a structure other than the bulged duplex; nevertheless, the crosslinking experiments strongly support specific interaction of this residue of RNP2 with an unpaired nucleotide at the branch position.

Comparison with Canonical and PseudoRRMs. The structure of the p14·peptide complex is reminiscent of a number of other unusual RRM and RRM-like structures (19–22). An α-helix C-terminal to the RRM of the spliceosomal U1A protein has been shown to adopt two different conformations with respect to the RNA-binding face of the RRM (Fig. 4*A*). In one orientation, the β-sheet of the RRM is masked (19); an $\approx$135° rotation of this helix reveals the surface of the RRM (20) in a structure similar to that observed in the U1A protein·RNA complex (12). In contrast to U1A, the C-terminal helices of p14 are held rigidly in position through an extensive network of hydrophobic and hydrophilic interactions, which are both intramolecular and between p14 and SF3b (Figs. 2 *A* and *B* and 4*A*). It is unlikely that a rearrangement of structure similar to that proposed for U1A would expose the p14 RNP motifs for branch duplex recognition.

A number of proteins contain atypical RRMs that differ from canonical RRMs in conserved sequence elements, such as the RNP motifs, in the length of conserved secondary structural elements, and in the predicted charge of the surface of the RRM β-sheet. Recent high-resolution structural studies of atypical RRMs from the splicing factors U2AF^{65} and U2AF^{35} provide a model of the interaction between U2AF^{65} and U2AF^{35} as well as that between U2AF^{65} and the splicing factor SF1; in both cases, the RRM serves as a scaffold for interactions with a partner protein mediated by contacts to the rear (α-helical) surface of the RRM (21, 22). These pseudoRRMs have been labeled U2AF homology motifs (UHMs) and are part of a larger family of putative protein·protein interaction domains (23); thus, RRMs or RRM-related motifs can serve as scaffolds for noncanonical ligand interactions. A comparison of both the p14 RNP sequence and the structure of the p14 RRM with standard RRMs and UHMs suggests that p14 is best classified as containing a canonical RRM. In addition, the overall positive charge of the p14 RRM and the p14·SF3b155 peptide complex is consistent with either the RRM surface or the complex representing an RNA interaction surface, in contrast, for example, to the UHM of U2AF^{65} (Fig. 4*B*).

p14 Within SF3b and the U11/U12 di-snRNP. Recent cryo-EM studies have provided a 10-Å resolution model of SF3b alone and within the context of the U11/U12 di-snRNP, a component of the minor spliceosome; p14 was modeled on the basis of its predicted RRM and assigned density in both structures (24, 25). Fitting of the x-ray structure of the p14·SF3b155 peptide into the EM density of isolated SF3b resulted in an excellent agreement with respect to the overall shape (Fig. 5*A*). The cocrystal structure and isolated SF3b show a globular domain corresponding to the RRM domain of p14 and two connecting bridges corresponding to the N-terminal α-helix of the SF3b155 peptide and the C-terminal α-helix of the SF3b155 peptide plus the C-terminal α-helix of p14. Similarly, the cocrystal structure of the p14·SF3b155 peptide fits well into the outer globular domain of the U11/U12 di-snRNP previously suggested (25) to represent p14. In this fit, the β-sheet of the p14 RRM is oriented toward

the outer surface of the U11/U12 di-snRNP, whereas the two α-helices of the p14 RRM are located toward the interior of the complex (Fig. 5*B*). The C-terminal α-helix of p14 is located on the outside, consistent with immunoprecipitation data in which the U11/U12 di-snRNP was precipitated by an anti-p14 antibody directed against the C terminus (10). In contrast to isolated SF3b (24), where density representing p14 was found to be caged in a central cavity, within the U11/U12 structure a rotation of one SF3b shell half by 90°, combined with some smaller movements, opens up SF3b and reveals the surface of p14 (25). Overall, there are no indications in the EM data of isolated SF3b or the U11/U12 di-snRNP that the conformation of the p14·SF3b155 peptide interface is changed.

The positioning of p14 within the U11/U12 structure suggests that the pre-mRNA enters into a cleft on the surface of the U11/U12 di-snRNP to interact with p14 (Fig. 5*B*). This cleft is enough to accommodate even double-stranded RNA because the space between the outer wall of the U11/U12 di-snRNP and the p14 RRM is >2 nm wide. It is possible, however, that pre-mRNA association involves conformational changes involving the p14 region, particularly given the requirement for duplex formation between U12 (U2) snRNA and the branch region. Because RNA density has not been assigned in the U11/U12 di-snRNP, it is not possible at this point to model the spatial relationship between p14 and the U12 snRNA.

Discussion

The results of the crosslinking experiments reported here suggest that the pre-mRNA branch adenosine interacts directly with the conserved aromatic of RNP2 in the RRM of the p14·SF3b155 peptide complex. High-resolution structures of RNA·RRM complexes show a conserved mode of RNA·protein interaction in which RNA is bound across the RRM β-sheet, making interactions with a variety of protein side chains, including conserved residues of RNP1 and RNP2. Because a significant portion of the p14 β-sheet is occluded by a C-terminal α-helix and portions of the SF3b peptide, either a rearrangement must occur upon RNA binding or the branch duplex interacts with p14·SF3b155 in a noncanonical fashion. Modeling of the p14·peptide structure into the cryo-EM structures suggests that no rearrangement occurs upon incorporation of SF3b into U12 (or U2) snRNP, although this could also occur later in spliceosome assembly.

It is important to note that p14 is unique among RRM-containing proteins characterized to date in that it interacts with a bulged duplex as opposed to single-stranded RNA. The canonical interaction between an RRM and RNA includes critical stacking interactions between nucleobases and the conserved aromatic residues of RNP1 and RNP2. Given the fact that the bulged branch duplex contains only one unpaired residue, it seems reasonable that only one of the two RNP aromatics would interact with bound RNA.

A bulged duplex model of the pre-mRNA·U2 structure that features an extruded adenosine at the branch position (26) can be docked to the p14·SF3b155 interface by positioning the adenine base of the bulged nucleotide within the surface pocket to allow a stacking interaction with Y22; a small movement of the side chain of one of the solvent-exposed basic residues surrounding the pocket, R96, would allow this interaction. By anchoring the branch nucleotide at this position, two orientations of the duplex rotated 180° with respect to one another are possible. Interestingly, either model positions the phosphate backbone of the duplex proximal to the conserved, positively charged residues of p14 that surround the pocket (Fig. 2*B*). Thus, specificity for the branch duplex could be determined by an adenosine-binding pocket and appropriately positioned basic residues. We propose that these features of the p14·SF3b155 interface present a surface for RNA recognition at the heart of the spliceosome.

Materials and Methods

Determination of SF3b155 p14 Interaction Domain. Maltose-binding protein (MBP)·p14 fusion protein (300 pmol) was incubated with 300 pmol of GST-SF3b155 constructs and 10 μl of amylose beads (GE Healthcare) in buffer (10 mM Hepes, pH 7.9/60 mM KCl/2 mM $MgCl_2$/0.1 mM EDTA/0.5 mM DTT) for 30 min at 23°C. Beads were pelleted by centrifugation at 1,000 $\times$ *g* and washed twice with buffer after removal of supernatant. Protein was eluted with buffer containing 100 mM maltose, run on a 16% SDS/PAGE gel, and transferred to nitrocellulose (Millipore) for 3 h at 200 mA. Membranes were blocked in 5% milk powder and probed with a 1:5,000 dilution of anti-GST-horseradish peroxidase conjugate. After being washed, the membrane was developed by addition of 1 mg/ml 4-chloro-1-naphthol and 0.1% H_2O_2 in Tris-buffered saline with 20% methanol.

Protein Expression and Purification. DNA encoding full-length human p14 was cloned into the EcoRI and PstI sites of pMALc2x (NEB, Beverly, MA) by using PCR primers to insert a tobacco etch virus (TEV) protease cleavage site between MBP and p14. Mutagenesis of p14 was carried out by PCR and confirmed by sequencing. The resulting MBP·p14 fusion proteins were expressed in *Escherichia coli* and purified by sequential amylose resin and anion exchange chromatography. DNA encoding amino acids 373–415 of SF3b155 was cloned into the EcoRI and BamHI sites of pGEX6p1 (GE Healthcare) by using PCR primers to insert a TEV protease cleavage site between GST and SF3b 373–415. Fusion protein was expressed in *E. coli* and purified by glutathione Sepharose chromatography (GE Healthcare). After cleavage of the GST tag using TEV protease, the SF3b155 peptide was purified on a Superdex-75 26/60 column (GE Healthcare).

To prepare the complex, SF3b155 peptide was incubated with MBP·p14 followed by cleavage of the fusion protein with TEV protease. The p14·SF3b155 peptide complex was then purified by cation exchange on a Source15S HR10/10 column (GE Healthcare) followed by gel filtration with a Superdex-75 26/60 column (GE Healthcare).

Crystallization. Crystals of p14·SF3b155 peptide were grown at 23°C by using the hanging drop vapor diffusion technique. Crystals of native complex and complex containing selenomethionine-substituted SF3b155 peptide were grown by mixing 2 μl of 10 mg/ml protein solution (10 mM Tris, pH 8.0/60 mM KCl/1 mM EDTA/5 mM 2-mercaptoethanol/0.02% NaN_3) with 2 μl of precipitant (14–18% polyethylene glycol 3350/100 mM MOPS, pH 6.0/200 mM $NaHCO_2$). Y22M-containing crystals were grown by mixing 2 μl of protein solution with 2 μl of precipitant (1.8 M 60%/40% NaH_2PO_4/K_2HPO_4 and 500 mM β-alanine). Crystals were transferred to precipitant containing 20% glycerol and frozen in liquid nitrogen for data collection.

Data Collection and Processing. For the wild-type and mutant complexes, data were collected at beamline 8.3.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory. Data for the wild-type complex were collected from a single selenomethionine derivatized crystal; a two-wavelength multiwavelength anomalous diffraction experiment was performed, collecting data in an inverse-beam mode at the midpoint energy between the experimentally determined Se/K edge and the inflection point. The crystal was subsequently translated, and further data were collected at a lower energy wavelength, remote from the Se absorption edge. Data were processed and scaled with the HKL package (Table 1) (27).

Model Building and Refinement. For the native complex, the program SOLVE (28) was used to determine the positions of four

of the expected six Se atoms (the remaining two Se atoms were later found to be in a disordered region of the structure). Initial phases to 3.1 Å were extended to 2.5 Å with maximum likelihood density modification in RESOLVE (29). An initial model automatically built by RESOLVE (30) served as the basis for model building and refinement. Iterative cycles of refinement in REFMAC (31) against the low-energy remote data, manual model building using XFIT (32), and automatic model building using RESOLVE were used to complete and refine the model. Two-fold noncrystallographic symmetry restraints were maintained throughout refinement but were relaxed for regions of the structure showing deviation due to different packing environments. The structure of the Y22M mutant complex was solved by molecular replacement. Refinement statistics for both structures are summarized in Table 1.

Crosslink Mapping. RNA (5′-GGGCGGUGGUGCCCUGGUGGGUGCUGACCGCCC-3′) was prepared and labeled at the branch nucleotide by T7 transcription [using 3,000 Ci/mmol (1 Ci = 37 GBq) [α-^{32}P]ATP] from a synthetic DNA, followed by purification on a 15% denaturing PAGE gel. Recombinant p14·SF3b155 373–415 (20 μM) was incubated with 1 pmol of RNA in buffer (10 mM Hepes, pH 7.9/60 mM KCl/2 mM $MgCl_2$/0.1 mM EDTA/0.5 mM DTT) in the presence of 1 μg of tRNA (Roche) for 40 min at 23°C. Reactions were irradiated on ice with a 5-W, 254-nm UV lamp (Ultraviolet Products, San Gabriel, CA) at a distance of 8 mm for 30 min. Nuclease P1 was added to a concentration of 50 ng/μl, and RNA was digested at 55°C for 2 h. For CNBr cleavage, reactions were brought to 70% formic acid and 50 mg/ml CNBr and incubated overnight at 23°C in the dark. Formic acid was removed *in vacuo*, and protein was precipitated with acetone to remove traces of acid. For Glu-C digestions, the purified CNBr fragment was resuspended in 100 mM NH_4HCO_3, pH 7.8, with 10% acetonitrile. Glu-C protease (500 ng) was added, and cleavage was allowed to proceed overnight at 30°C. For Lys-C digestions, the purified CNBr fragment was resuspended in 25 mM Tris, pH 8.5/1 mM EDTA/10% acetonitrile. Lys-C (100 ng) was added, and cleavage was allowed to proceed overnight at 37°C. Crosslinking and cleavage reactions were separated by 16% 15:1 SDS/PAGE and visualized using a Molecular Dynamics PhosphorImager.

Comparison of X-ray and cryo-EM Structures. For fitting, the EM densities of isolated SF3b (24) and of the U11/U12 di-snRNP (25) were low-pass-filtered to their respective resolution. Thresholds of the EM densities were chosen based on the theoretical molecular mass of the particle. Fitting was performed manually by using the 3D alignment tool of the software AMIRADEV 2.3 (TGS Europe, Merignac, France). The cage-like structure of SF3b contains a limited number of larger-density elements that are potential locations for an RRM. Three of nine elements revealed good fits with an RRM; p14 was distinguished from the SF3b component SF3b49 based on the presence of two adjacent RRMs in SF3b49 and only one in p14. Subsequent to manual fitting of the x-ray structure of the p14/SF3b155 peptide, an exhaustive real-space refinement in a small translational and rotational range was performed to optimize the fit. The quality of the fits was judged visually and by normalized cross-correlation coefficients.

We thank Bart Hazes and Joanne Lemieux for help with data collection and Howard Young for helpful discussions. This work was supported by a Canadian Institutes of Health Research grant (to A.M.M.) and by funding from the Alberta Synchrotron Institute.

1. Krämer, A. (1996) *Annu. Rev. Biochem.* **65,** 367–409.
2. Staley, J. P. & Guthrie, C. (1998) *Cell* **92,** 315–326.
3. Burge, C. B., Tuschl, T. & Sharp, P. A. (1999) in *The RNA World*, eds. Gesteland, R. F., Cech, T. R. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Woodbury, NY), 2nd Ed., pp. 525–560.
4. Sharp, P. A. (1991) *Science* **254,** 663.
5. Valadkhan, S. & Manley, J. L. (2001) *Nature* **413,** 701–707.
6. MacMillan, A. M., Query, C. C., Allerson, C. A., Chen, S., Verdine, G. L. & Sharp, P. A. (1994) *Genes Dev.* **8,** 3008–3020.
7. Query, C. C., Strobel, S. A. & Sharp, P. A. (1996) *EMBO J.* **15,** 1392–13402.
8. Query, C. C., Moore, M. J. & Sharp P. A. (1994) *Genes Dev.* **8,** 587–597.
9. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000) *Science* **289,** 920–930.
10. Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., Lührmann, R. & Query, C. C. (2001) *EMBO J.* **20,** 4536–4546.
11. Kenan, D. J., Query, C. C. & Keene, J. D. (1991) *Trends Biochem. Sci.* **16,** 214–220.
12. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994) *Nature* **372,** 432–438.
13. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. (1999) *Cell* **98,** 835–845.
14. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. & Yokoyama, S. (1999) *Nature* **398,** 579–585.
15. Allain, F. H. T., Bouvet, P., Dieckmann, T. & Feigon, J. (2000) *EMBO J.* **19,** 6870–6871.
16. Wang, X. & Hall, T. M. T. (2001) *Nat. Struct. Biol.* **8,** 141–145.
17. Gozani, O., Potashkin, J. & Reed, R. (1998) *Mol. Cell. Biol.* **18,** 4752–4760.
18. Clément, M., Martin, S. S., Beaulieu, M.-E., Chamberland, C., Lavigne, P., Leduc, R., Guillemette, G. & Escher, E. (2005) *J. Biol. Chem.* **280,** 27121–27129.
19. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K. & Neuhaus, D. (1996) *J. Mol. Biol.* **257,** 398–411.
20. Rupert, P. B., Xiao, H. & Ferre-D'Amare, A. R. (2003) *Acta Crystallogr. D* **57,** 1521–1524.
21. Kielkopf, C. L., Rodionova, N. A., Green, M. R. & Burley, S. K. (2001) *Cell* **106,** 595–605.
22. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A. & Sattler, M. (2003) *Mol. Cell.* **11,** 965–976.
23. Kielkopf, C. L., Lucke, S. & Green, M. R. (2004) *Genes Dev.* **18,** 1513–1526.
24. Golas, M. M., Sander, B., Will, C. L., Lührmann, R. & Stark, H. (2003) *Science* **300,** 980–984.
25. Golas, M. M., Sander, B., Will, C. L., Lührmann, R. & Stark, H. (2005) *Mol. Cell* **17,** 869–883.
26. Berglund, J. A., Rosbash, M. & Schultz, S. C. (2000) *RNA* **7,** 682–691.
27. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
28. Terwilliger, T. C. & Berendzen, J. (1999). *Acta Crystallogr. D* **55,** 849–861.
29. Terwilliger, T. C. (2000) *Acta Crystallogr. D* **56,** 965–972.
30. Terwilliger, T. C. (2002) *Acta Crystallogr. D* **59,** 34–44.
31. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Crystallogr. D* **53,** 240–255.
32. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165.