

Complex assembly mechanism and an RNA-binding mode of the human p14-SF3b155 spliceosomal protein complex identified by NMR solution structure and functional analyses

Kanako Kuwasako,¹ Naoshi Dohmae,² Mio Inoue,¹ Mikako Shirouzu,^{1,3} Seiichi Taguchi,⁴ Peter Güntert,^{1,5} Bertrand Séraphin,⁶ Yutaka Muto,^{1*} and Shigeyuki Yokoyama^{1,7*}

¹ Protein Research Group, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan

² Biomolecular Characterization Team, RIKEN, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan

³ RIKEN SPring-8 Center, Harima Institute, Hyogo 679-5148, Japan

⁴ Division of Biotechnology and Macromolecular Chemistry, Graduate School of Engineering, Hokkaido University, N13W8, Kita-ku, Sapporo, Hokkaido 060-8628, Japan

⁵ Tatsuo Miyazawa Memorial Program, RIKEN Genomic Sciences Center, Yokohama 230-0045, Japan

⁶ Équipe Labelisée “La Ligue”, Centre de Génétique Moléculaire, Gif sur Yvette, France

⁷ Department of Biophysics and Biochemistry, Graduate School of Sciences, The University of Tokyo, Tokyo 113-0033, Japan

ABSTRACT

The spliceosomal protein p14, a component of the SF3b complex in the U2 small nuclear ribonucleoprotein (snRNP), is essential for the U2 snRNP to recognize the branch site adenosine. The elucidation of the dynamic process of the splicing machinery rearrangement awaited the solution structural information. We identified a suitable complex of human p14 and the SF3b155 fragment for the determination of its solution structure by NMR. In addition to the overall structure of the complex, which was recently reported in a crystallographic study (typical RNA recognition motif fold $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ - $\beta 4$ of p14, and αA - βA fold of the SF3b155 fragment), we identified three important features revealed by the NMR solution structure. First, the C-terminal extension and the nuclear localization signal of p14 ($\alpha 3$ and $\alpha 4$ in the crystal structure, respectively) were dispensable for the complex formation. Second, the proline-rich segment of SF3b155, following βA , closely approaches p14. Third, interestingly, the $\beta 1$ - $\alpha 1$ loop and the $\alpha 2$ - $\beta 4$ β -hairpin form a positively charged groove. Extensive mutagenesis analyses revealed the functional relevance of the residues involved in the protein-protein interactions: two aromatic residues of SF3b155 (Phe408 and Tyr412) play crucial roles in the complex formation, and two hydro-

phobic residues (Val414 and Leu415) in SF3b 155 serve as an anchor for the complex formation, by cooperating with the aromatic residues. These findings clearly led to the conclusion that SF3b155 binds to p14 with three contact points, involving Phe408, Tyr412, and Val414/Leu415. Furthermore, to dissect the interactions between p14 and the branch site RNA, we performed chemical-shift-perturbation experiments, not only for the main-chain but also for the side-chain resonances, for several p14-SF3b155 complex constructs upon binding to RNA. These analyses identified a positively charged groove and the C-terminal extension of p14 as RNA-binding sites. Strikingly, an aromatic residue in the $\beta 1$ - $\alpha 1$ loop, Tyr28, and a positively charged residue in the $\alpha 2$ - $\beta 4$ β -hairpin, Arg85, are critical for the RNA-binding activity of the positively charged groove. The Tyr28Ala and Arg85Ala point mutants and a deletion mutant of the C-terminal extension clearly revealed that their RNA binding activities were independent of each other. Collectively, this study provides details for the protein-recognition mode of p14 and insight into the branch site recognition.

Proteins 2008; 71:1617–1636.
© 2007 Wiley-Liss, Inc.

Key words: NMR; p14; RRM; SF3b155; U2 snRNP.

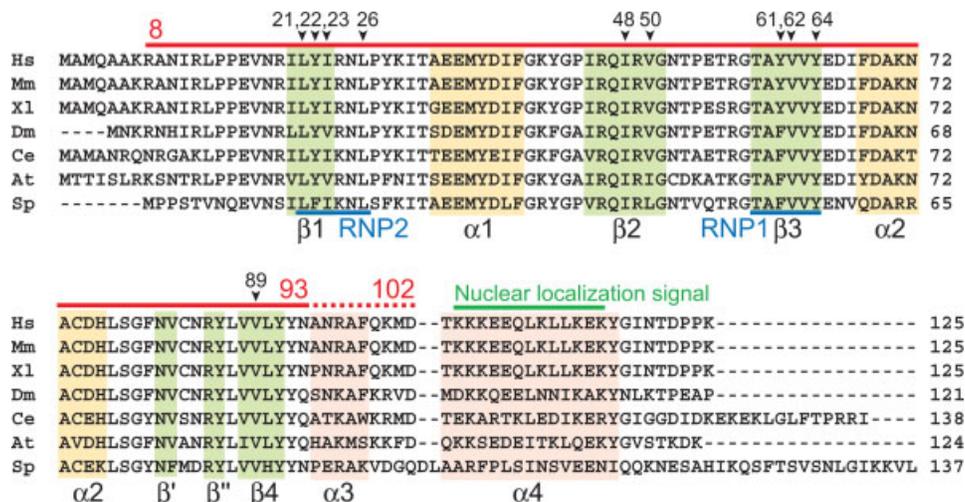
The Supplementary Material referred to in this article can be found online at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/>

Grant sponsors: RIKEN Structural Genomics/Proteomics Initiative (RSGI), National Project on Protein Structural and Functional Analyses of the Ministry of Education, Culture, Sports, Science and Technology of Japan, Human Frontier Science Program (HFSP).

*Correspondence to: Shigeyuki Yokoyama, Project Director, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. E-mail: yokoyama@biochem.s.u-tokyo.ac.jp or Yutaka Muto, Team Leader, Stable Isotopes Team, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan. E-mail: ymuto@gsc.riken.jp

Received 3 April 2007; Revised 4 September 2007; Accepted 20 September 2007

Published online 12 December 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21839

**Figure 1**

Multiple sequence alignment of p14 proteins. The p14 sequences from *Homo sapiens* (Hs, accession No. AF401310), *Mus musculus* (Mm, #P59708), *Xenopus laevis* (Xl, #AAH57709), *Drosophila melanogaster* (Dm, #Q9VRV7), *Caenorhabditis elegans* (Ce, #AAN73870), *Arabidopsis thaliana* (At, #AAM67034), and *Schizosaccharomyces pombe* (Sp, #CAA18384) were aligned with ClustalX.¹⁶ Secondary structure elements are depicted with green (β -sheet) and yellow (α -helix) boxes. The signature sequences of RNP1 and RNP2 are indicated by the blue lines. The nuclear localization signal sequence is indicated by the green line. The red line indicates the region utilized for structure determination. The red broken line indicates semiconserved residues among species. The arrowheads indicate hydrophobic residues on the β -sheet surface that are buried in the protein fold and conserved in the RRMs.

INTRODUCTION

Pre-mRNA splicing occurs in the nucleus through two successive trans-esterification reactions, which are carried out by the spliceosome,^{1,2} a set of large and highly dynamic macromolecular complexes containing the small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5, and U6.^{3–5} The U2 snRNP is a large protein-RNA complex containing the U2 snRNA, which plays an important role in branch site recognition via base-pairing with a complementary RNA sequence.^{6–9}

The SF3b protein complex, a component of the U2 snRNP,^{10–12} plays a key role in the recruitment and stable binding to the branch site.^{10,13–15} SF3b is composed of seven proteins, SF3b155, SF3b145, SF3b130, SF3b49, SF3b14b, SF3b10, and p14.^{10–12} The p14 protein, which is highly conserved in a wide variety of organisms (see Fig. 1),¹⁷ has been shown to form a stable complex with SF3b155.¹⁴ This protein complex forms the core responsible for branch site recognition; p14 and SF3b155 interact directly with the branch site adenosine and the flanking regions, respectively.^{17,18}

The architectures of isolated SF3b and of SF3b in the U11/U12 di-snRNP, which is a member of the minor spliceosome,¹⁹ have been investigated by cryo-EM experiments.^{20,21} SF3b can reportedly adopt two conformations. In the isolated SF3b complex, the p14 protein is located in the central cavity of the complex, where it is hidden from the solvent. In contrast, upon integration

into the U11/U12 di-snRNP, the structure of SF3b adopts a more open conformation, and p14 is exposed to the solvent for the recognition of pre-mRNA, suggesting that a conformational change of SF3b155 should be necessary for branch site adenosine recognition by p14.^{20,21}

Furthermore, a recent crystallographic study reported the structure of the complex composed of the p14 protein and a peptide corresponding to SF3b155 (373–415).²² In this structure, the p14 protein adopts a canonical RNA recognition motif (RRM) fold (β 1- α 1- β 2- β 3- α 2- β 4), but possesses a peptide-recognition mode that differs from those of other known peptide-binding RRMs.^{23–29} In addition, the N-terminal part of the SF3b155 fragment and the C-terminally extended helix of p14 (α 3) covered the entire β -sheet surface of the RRM fold, and formed an extensive network of hydrophobic and hydrophilic interactions that rigidly hold the C-terminal helices.²² However, no functional involvement of the components in this interaction network has been specified, although such information is essential for precisely understanding the assembly manner of this spliceosomal protein complex.

Consistent with the general knowledge that the two conserved RNP motifs (RNP1 and RNP2) located on β 3 and β 1 are often necessary for RNA recognition by RRMs,³⁰ it was also argued that Tyr22, on the first β -strand (i.e., RNP2 motif), could be involved in the RNA recognition by p14; when the Tyr22 residue was replaced with methionine, the ectopic Met residue (Tyr22Met)

was crosslinked with the target RNA by UV irradiation.²² On the other hand, another group reported a main-chain chemical shift perturbation analysis by NMR, but did not solve the solution structure.³¹ They argued that the residues responsible for RNA binding are located mainly in the β' - β'' hairpin formed by the α 2- β 4 loop and the C-terminal extension α 3, rather than in the RNP motifs.³¹ Thus, the observations for the RNA-binding activity by p14 still remain controversial.

A detailed solution structure of the p14-SF3b155 complex should provide information complementary to that obtained from the crystal structures for the elucidation of the precise interactions between p14 and SF3b155 in solution, as well as the branch site recognition by the complex. Here, we produced a p14-SF3b155 complex suitable for NMR structural determination. Our detailed examination of the solution structure detected a deep, positively charged groove formed by the β 1- α 1 loop and the β' - β'' hairpin. Extensive mutational analyses revealed that the SF3b155 peptide binds to the p14 RRM through three contact points. Furthermore, we investigated the chemical shift perturbations, not only for the main-chain but also for the side-chain resonances, for several p14-SF3b155 complex constructs upon binding to RNA, and identified the C-terminal extension α 3 and the positively charged groove of p14 RRM as the RNA-binding sites. Interestingly, our study clearly shows the independent RNA-binding activities of the positively charged groove and the C-terminal extension. This study provides an important basis for understanding both the protein binding mode of p14 and the branch point recognition mode of the U2 snRNP.

MATERIALS AND METHODS

Preparation of the p14-SF3b 155 complex by reconstitution

The cDNAs encoding full length p14 (NIH accession number AF401310) and the SF3b155 fragments (199–437, 379–423, 391–423, and 391–437) (NIH accession number NM_012433) were cloned into the *Eco* RI/*Sal* I sites of pGEX6P-1 (Amersham Biosciences) and the *Nde* I/*Xho* I sites of pET-15b (Novagen), respectively. In both constructs, a TEV protease cleavage site (cDNA sequence gaaaacctgtattccaggc, corresponding to Glu-Asn-Leu-Tyr-Phe-Gln-Gly) was placed between the tag and protein sequences.

E. coli strain BL21 (DE3) cells were transformed with the recombinant plasmids and grown at 37°C, in 2× YT medium supplemented with 50 mg/L ampicillin for the nonlabeled sample, and in modified minimal medium³² supplemented with 25% ¹⁵N, ¹³C labeled *E. coli*-OD2 (Silantes) and 50 mg/L ampicillin, for the ¹⁵N, ¹³C-labeled samples, respectively. IPTG was added to the culture to a final concentration of 1 mM to induce protein

expression. After 5–6 h of cultivation, the cells were harvested and lysed by sonication in phosphate buffer containing 1 mM DTT, 1 mM PMSE, and a protease inhibitor cocktail for general use (Nacalai Tesque).

To prepare the complex by reconstitution, the solubilized GST-tagged p14 and His-tagged SF3b155 fragments were both applied to a Glutathione Sepharose 4 Fast Flow (Amersham Biosciences) column, in which the complexes were reconstituted. The complex was eluted by the addition of glutathione. The eluted sample was further purified by chromatography on a Ni-NTA SuperFlow (Qiagen) column with an imidazole gradient from 20 to 250 mM, and the tags were removed by overnight TEV protease treatment at room temperature. The tag-free complexes were further purified by gel filtration on a Hi Load 16/60 Superdex 75 column (Amersham Biosciences).

Preparation of the p14-SF3a155 complex by coexpression

The ¹⁵N/¹³C-labeled and unlabeled p14 complex for the structure determination and the mutant protein for the GST pull-down assays were prepared by a coexpression system, using the expression vector pGEX6p-1. In contrast to the system described earlier, a cDNA encoding the SF3b155 fragment (379–424) was cloned into the *Eco* RI/*Sal* I sites with a TEV cleavage site between GST and SF3b155. The cDNA encoding the His-tagged p14 protein, with the TEV cleavage site between the His-tag and p14, was cloned into the downstream site (*Sal* I/*Not* I site) of the pGEX6p-1-SF3b155 plasmid, so that the ribosomal binding site (RBS: gaaggag) was placed between the stop codon of SF3b155 and the start codon of the subsequent p14 cDNA. The complex was coexpressed in *E. coli* strain BL21 (DE3), using pGEX-SF3b155 (379–424)-(RBS)-p14 (8–93). Cells were grown at 37°C in 2× YT media for the nonlabeled complex and in modified minimal medium for the ¹⁵N, ¹³C-labeled complex. IPTG was added to the culture to a final concentration of 1 mM to induce expression of the proteins. After 5–6 h of cultivation at 25°C, the cells were harvested, and the complex was purified as described earlier. In the same way, the coexpression systems for the wild-type and mutant p14 (8–102)-SF3b155 (379–424) complexes and the p14 (8–120)-SF3b155 (379–424) complex were constructed.

Limited proteolysis

The reconstituted p14 (full length)-SF3a155 (199–437) complex (5 mg/mL) was incubated with increasing concentrations of trypsin in 10 mM Tris-HCl (pH 7.0), containing 1 mM TCEP (Tris(2-chloroethyl)phosphate), at 37°C for 40 min. The digested samples were analyzed by MALDI-TOF mass spectrometry (Voyager DE-STR, PerSeptive Biosystems) after calibration of the instrument

with trypsin standards. Assignments of the sequences from MALDI-TOF data were confirmed by amino acid sequencing.

GST pull-down assay

Aliquots of purified 0.25 nmol GST or GST-p14 (full length) were incubated with 25 μ L of Glutathione Sepharose and equimolar amounts of each of the SF3b155 fragments, which were identified by limited proteolysis, for 30 min at 4°C. The resin was washed five times with 500 μ L of phosphate buffer containing 0.05% NP-40. The bound proteins were eluted by the addition of 50 μ L of SDS gel loading buffer and were analyzed by SDS-PAGE. Proteins were visualized by Coomassie Brilliant Blue staining.

Mutational analysis

Point mutations and multiple substitutions were introduced into the GST-SF3b155 fragment (379–424), using previously reported examples.³³ Primers (24–27mer) spanning the site of the desired mutation on both DNA strands were used. Mutations were confirmed by sequencing.

For the detection of complex formation, His-tagged p14 (8–93) and several mutants of GST-SF3b155 (379–424) were coexpressed in 2 \times YT medium, as described earlier. The cells from a 10 mL culture were harvested and lysed by sonication in 1 mL of phosphate buffer, containing 1 mM PMSE, 1 mM DTT, and a protease inhibitor cocktail (Nacalai Tesque). Aliquots (2.5- μ L) volume of the cell lysates were analyzed by SDS-PAGE and stained with Coomassie Brilliant Blue dye, in order to examine the amount of protein produced in the culture. The solubilized samples were bound to 25 μ L of Glutathione Sepharose for 30 min, at 4°C. The proteins bound to the resin were analyzed by SDS-PAGE, as described earlier.

NMR spectroscopy

The NMR samples were concentrated to \sim 150–300 μ M in 20 mM sodium phosphate buffer (pH 6.5), containing 100 mM NaCl and 1 mM d-DTT (in 90% H₂O/10% D₂O), using a Vivaspin 20 mL concentrator (membrane: 3000 MWCO PES, VIVASCIENCE).

NMR experiments were performed at 25 and 35°C on a Bruker 700 MHz spectrometer with a cryo-probe and a Bruker 800 MHz spectrometer (Bruker AV700 and Bruker AV 800). The ¹H, ¹⁵N, and ¹³C chemical shifts were referenced relative to the frequency of the ²H lock resonance of water. Backbone and side-chain assignments were obtained by using a combination of standard triple resonance experiments.^{34,35} 2D [¹H, ¹⁵N]-HSQC and 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra were used for the ¹H, ¹⁵N,

and ¹³C assignments of the protein backbone. Side-chain ¹H and ¹³C assignments of the nonaromatic side-chains, including all prolines, were obtained using 2D [¹H, ¹³C]-HSQC, and 3D HBHA(CO)NH, H(CCCO)NH, (H)CC(CO)NH, HCCH-COSY, HCCH-TOCSY, and (H)CCH-TOCSY spectra. Assignments were checked for consistency with 3D ¹⁵N-edited [¹H, ¹H]-NOESY and ¹³C-edited [¹H, ¹H]-NOESY spectra. The ¹H and ¹³C spin systems of the aromatic rings of Phe, Trp, His, and Tyr were identified using 3D HCCH-COSY and HCCH-TOCSY experiments, and 3D ¹³C-edited [¹H, ¹H]-NOESY was used for the sequence-specific resonance assignment of the aromatic side-chains. NOESY spectra were recorded with mixing times of 80–150 ms. The 2D and 3D spectra were processed using NMRPipe.³⁶ Analyses of the processed data were performed with the programs NMRView³⁷ and KIJIRA.³⁸

For the amide-chemical shift titration experiments, the AdML BPS RNA oligonucleotide (5'-GUGCUGACC CUG-3') (Dharmacon) was dissolved in 20 mM sodium phosphate buffer (pH 6.5) to make a 30 mM solution. 2D [¹H, ¹⁵N] and [¹H, ¹³C]-HSQC spectra were recorded while increasing the concentration of the RNA relative to three different p14-SF3b155 complex solutions (200 μ M), to a final 1:3 ratio of complex/RNA.

The measurements of the nitrogen relaxation times T_1 and T_2 , and the proton-nitrogen heteronuclear NOEs were performed on a Bruker 600 MHz spectrometer with a cryo-probe (Bruker AV 600) at 35°C, using the ¹⁵N, ¹³C-labeled wild-type and mutant p14 RRM- α 3 complexes at a concentration of 200 μ M.³⁹ Eight different values for the relaxation time were recorded for the ¹⁵N T_1 (T_1 delays = 5, 65, 145, 246, 366, 527, 757, and 1148 ms) and ¹⁵N T_2 (T_2 delays = 32, 48, 64, 80, 96, 112, 128, and 144 ms) relaxation experiments. ¹⁵N T_1 and ¹⁵N T_2 values were extracted using a curve-fitting subroutine included in the Sparky program (Goddard TD, and Kneller DG, SPARKY 3, University of California, San Francisco). The proton-nitrogen heteronuclear NOEs values were calculated as the ratio between the cross-peak intensities with (I) and without (I_0) ¹H saturation (I/I_0). The errors were estimated from the root-mean-square of the baseline noise in the two spectra.³⁹

Structure calculations

The 3D structures of the complex were determined by combined automated NOESY cross peak assignment⁴⁰ and structure calculations with torsion angle dynamics⁴¹ implemented in the program CYANA 2.1.⁴² Dihedral angle constraints for ϕ and ψ were obtained from the main-chain and ¹³C ^{β} chemical shift values, using the program TALOS,⁴³ and by analyzing the NOESY spectra. Structure calculations started from 100 randomized conformers and used the standard CYANA simulated annealing schedule⁴¹ with 20,000 torsion angle dynamics steps

per conformer. The 20 conformers with the lowest final CYANA target function values were subjected to restrained energy minimization in a water shell with the program OPALp,⁴⁴ using the AMBER force field.⁴⁵ PROCHECK-NMR⁴⁶ and MOLMOL⁴⁴ were used to validate and to visualize the final structures, respectively.

RESULTS

Identification of the regions of p14 and SF3b155 necessary for stable complex formation

At first, we prepared the recombinant full-length p14 protein using an over-expression system in *E. coli*, and found that although the GST-tagged p14 could be solubilized, the p14 precipitated rapidly upon release from GST, at a concentration of 2.5 mg/mL (data not shown). Therefore, we concluded that free p14 is unstable and has low solubility. On the other hand, when the GST-tagged p14 was mixed with SF3b155 (199–437), the p14 was stable after the release from the GST tag. The NMR spectra suggested that the vast majority of the residues in SF3b155 were flexible and disordered in this complex (data not shown).

Next, we subjected the complex to limited tryptic digestion, and identified R390 of SF3b155 as the main cleavage site by MALDI TOF mass spectrometry and N-terminal amino acid sequencing, leading us to speculate that SF3b155 (199–437) was digested with trypsin into two SF3b155 fragments, 199–390 and 391–437. Using a pull-down assay with GST-tagged p14, we found that SF3b155 (391–437) contained the binding site for p14 [Fig. 2(a,b)]. Furthermore, as it had been reported that SF3b155 (255–424) binds to p14,¹⁸ we investigated the shorter fragment (391–424), and found that it was sufficient for the binding [Fig. 2(b)]. These observations are consistent with other recent analyses.^{22,31,47}

Although we tried to prepare the p14-SF3b155 (391–424) complex by reconstitution, we found that the His-tag of the SF3b155 (391–424) fragment could not be removed efficiently in the sample preparation procedure (see Materials and Methods) (data not shown). Therefore, we decided to extend the N-terminus of the SF3b155 (391–424) fragment. This SF3b155 fragment (391–424) contained no Trp residues, although it has been reported that Trp residues are involved in protein–protein interactions with the RRM domains belonging to the U2AF-homology-motif (UHM) family.^{23–25} Considering the possibility that the Trp residues on the N-terminus of this fragment might be involved in the interaction with p14, we constructed SF3b155 (379–424) containing two Trp residues (Trp386 and Trp388) for the structural study [Fig. 2(c)], and found that the His-tag of the SF3b155 (379–424) fragment could be removed efficiently in the sample preparation procedure.

The ¹⁵N- and ¹³C-labeled fragment SF3b155 (379–424) alone exhibited sharp [¹H,¹⁵N]-HSQC spectra [Fig. 3(a)]. The ¹H chemical shifts of these resonances were observed within a narrow range (7.5–8.5 ppm), indicating that this SF3b155 fragment predominantly adopted a random coil structure. On the other hand, the SF3b155 (379–424) fragment in complex with the nonlabeled p14 protein yielded broad [¹H,¹⁵N]-HSQC spectra [Fig. 3(b)]. At the same time, the ¹H chemical shifts of these resonances were observed within a wide range (7.0–9.5 ppm), indicating that this SF3b155 fragment formed a stable structure only when complexed with p14. Furthermore, the complex of ¹⁵N/¹³C-labeled p14 and unlabeled SF3b155 (379–424) yielded well-dispersed [¹H,¹⁵N]-HSQC spectra [Fig. 3(c)], which indicated that the SF3b155 fragment was required for the stability of the p14 protein. A preliminary main-chain resonance assignment for these complexes suggested that the N- and C-terminal regions (1–15 and 94–125, respectively) of p14 might be flexible and disordered in solution. In contrast, residues 16–93 of p14, the RRM domain (see Fig. 1), exhibited chemical shifts indicative of a rigid conformation (Fig. S1).

Based on these results, we deleted the N- and C-terminal extensions of the p14 RRM to obtain the optimal p14-SF3b155 complex for structure determination. As residues 8–15 of the N-terminal extension of the p14 RRM were found to be necessary to remove the GST tag efficiently in the sample preparation procedure (see Materials and Methods), we deleted only the seven N-terminal residues (1–7) from p14. The amino acid residues in the C-terminal extension (94–102, corresponding to α 3 in the crystal structure) are semiconserved, but those in the nuclear localization signal (103–120, corresponding to α 4 in the crystal structure) are not conserved among species (see Fig. 1).²² Taking into consideration the possibility that the C-terminal extension (94–102) could function in spite of its disordered conformation in solution, we prepared complexes with different lengths of the C-terminal extension of the p14 RRM: p14 (8–93)-SF3b155 (379–424) (“p14 RRM complex” hereafter) and p14 (8–102)-SF3b155 (379–424) (“p14 RRM- α 3 complex” hereafter).

Dynamic properties of the complexes

Since it was difficult to assign the backbone amide resonances using the spectra obtained from the ¹⁵N/¹³C-labeled p14 RRM complex at 25°C, we examined the optimum temperature, and found that the spectra obtained at 35°C were suitable (Fig. S2).

Using the ¹⁵N,¹³C labeling technique, we fully assigned the backbone amide resonance signals of the residues 397–423 of the SF3b155 fragment, except for the prolines, in the p14 RRM complex and the p14 RRM- α 3 complex. We also assigned the backbone amide resonance

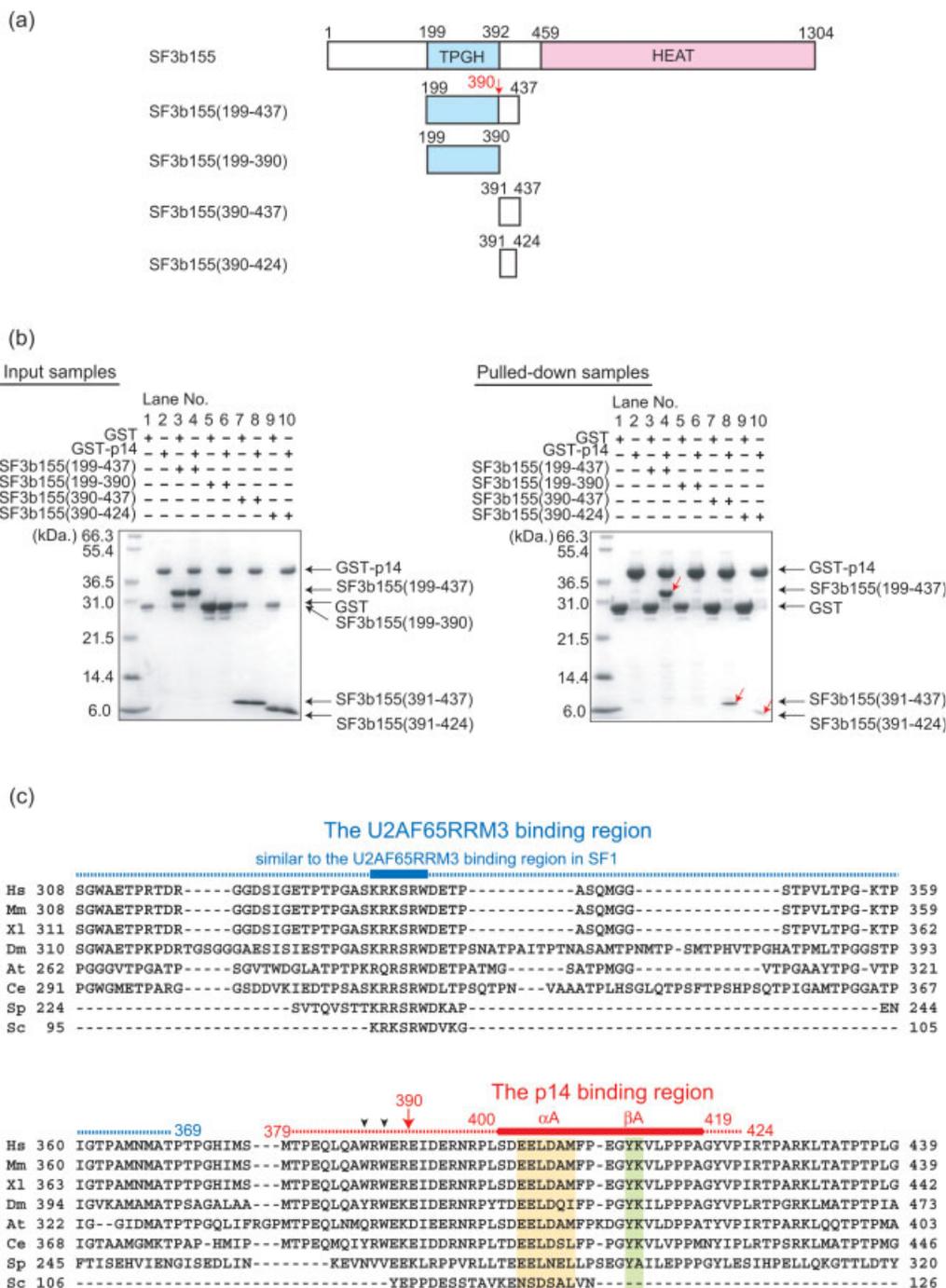


Figure 2

The p14 protein interacts with the SF3b155 fragment (391–424). (a) Schematic diagrams of the SF3b155 domain structure and the domain constructs used in (b). The TPGH repeats (TPGH) and the 22 tandem protein phosphatase 2A-like HEAT repeats (HEAT) are indicated with cyan and pink boxes, respectively. The red arrow indicates the trypsin cleavage site. (b) GST pull-down assays of p14 and the SF3b155 fragments. GST and GST-p14 were each incubated with the SF3b fragments, composed of residues 199–437, 199–390, 391–437, and 391–424 (Input samples, left). After the addition of Glutathione Sepharose, the proteins bound to the resin were analyzed by SDS-PAGE (Pulled-down samples, right). The red arrows indicate the SF3b155 fragments that bound to GST-p14. (c) Sequence alignment of the p14- and U2AF65-binding regions in SF3b155. The sequences of the SF3b155 fragments from *Homo sapiens* (Hs, accession No. NM_012433), *Mus musculus* (Mm, #NP112456), *Xenopus laevis* (Xl, #CAA70201), *Drosophila melanogaster* (Dm, #NP608534), *Arabidopsis thaliana* (At, #BAB09858), *Caenorhabditis elegans* (Ce, #CAA90775), *Schizosaccharomyces pombe* (Sp, #Q10178), and *Saccharomyces cerevisiae* (Sc, #NP014015) were aligned using ClustalX.¹⁶ It was reported that the region composed of residues 267–369 of SF3b155 binds to U2AF65 RRM3,¹⁴ and part of this region is indicated with a blue dashed line. The bold blue line indicates the amino acid sequence that is similar to that of the U2AF65 RRM3 binding site in BBP/SF1. The p14 binding region is indicated with a red dashed line, and the bold red line indicates the region essential for the complex formation. The arrowheads indicate Trp386 and Trp388.

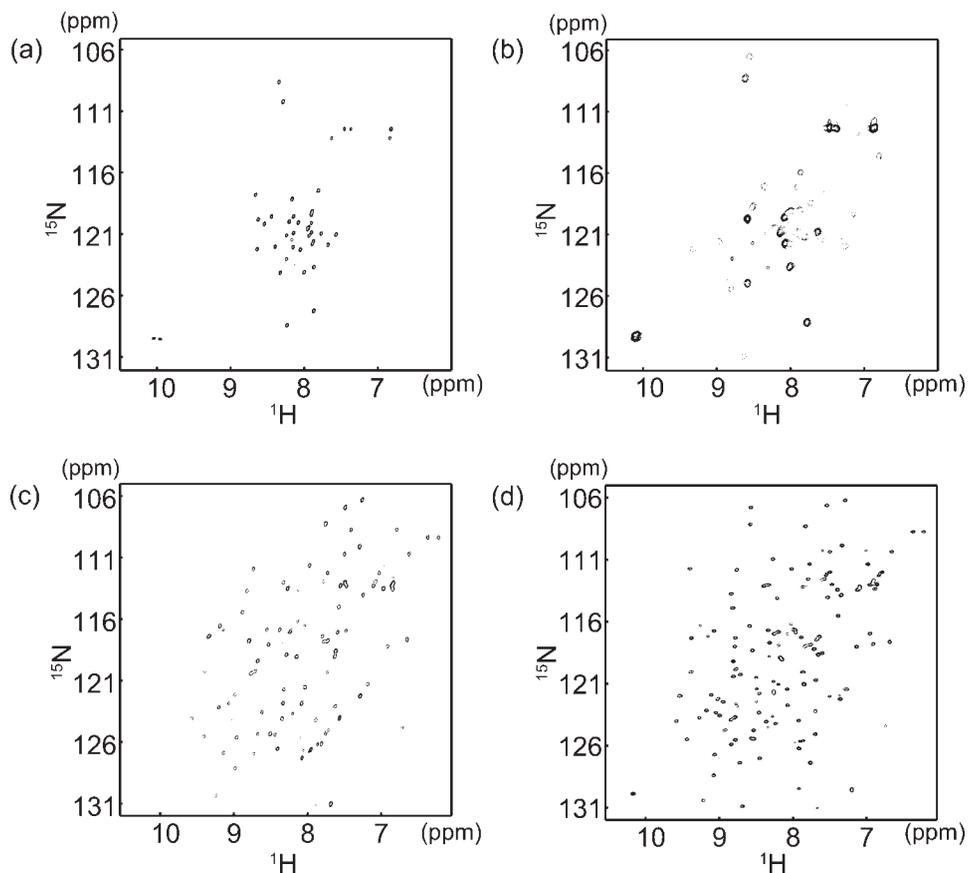


Figure 3

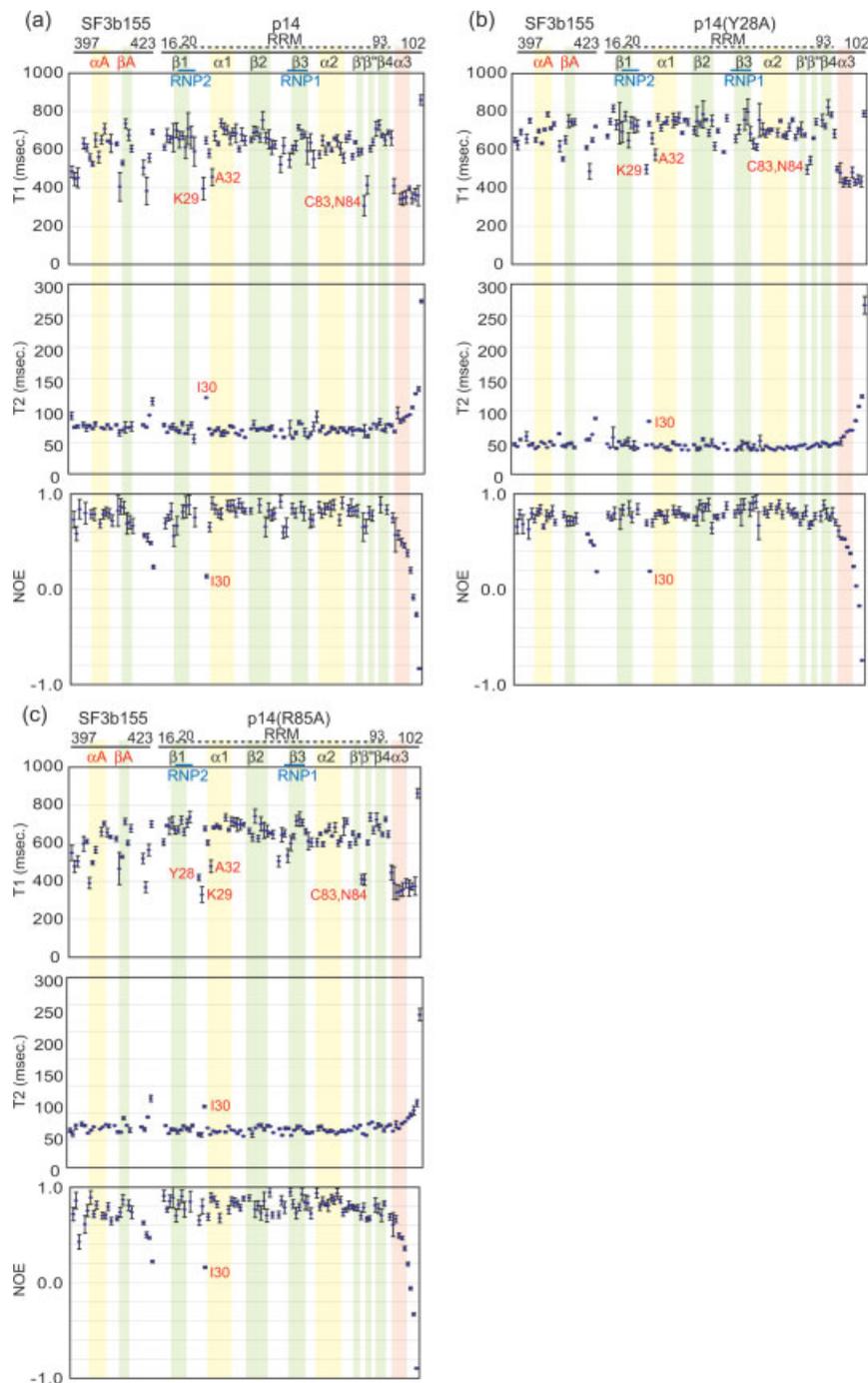
$[^1\text{H},^{15}\text{N}]$ -HSQC spectra of p14 and SF3b155. (a) $^{15}\text{N}/^{13}\text{C}$ -labeled SF3b155 fragment (379–424) alone. (b) $^{15}\text{N}/^{13}\text{C}$ -labeled SF3b155 fragment (379–424) complexed with the unlabeled p14 protein. (c) $^{15}\text{N}/^{13}\text{C}$ -labeled p14 protein complexed with the unlabeled SF3b155 fragment (379–424). (d) $^{15}\text{N}/^{13}\text{C}$ -labeled p14 (8–93)–SF3b155 (379–424) complex. All data were collected at pH 6.5 and at temperatures of 25°C (a, b, and c) and 35°C (d).

signals of residues 16–93 of p14 in the p14 RRM complex and residues 16–102 of p14 in the p14 RRM- α 3 complex, except for the prolines.

We then analyzed the dynamics of the p14 RRM- α 3 complex, by measuring the backbone nitrogen relaxation times T_1 and T_2 , and the proton-nitrogen heteronuclear NOEs [Fig. 4(a)]. The T_1 values for the majority of the C-terminal extension α 3 of p14 (average of 358 ms for residues 94–101, except for residue 102) were significantly (P value of t -test < 0.01) smaller than those for the rest of the p14 protein (average of 628 ms for residues 16–93). The T_2 values were also distinct between the regions of residues 16–93 and 94–102 of p14 (70 and 121 ms on average, respectively) (P value of t -test < 0.01). Furthermore, the heteronuclear NOE values of residues 94–102 (average 0.17) were also significantly (P value of t -test < 0.01) smaller than those of the rest (average 0.79 for residues 16–93). These data were consistent with the above-mentioned preliminary resonance assign-

ment analysis. The C-terminal extension of p14 RRM (94–102) is more flexible than the rest of the complex in solution. Perhaps as a reflection of the disordered nature of the C-terminal extension, the p14 RRM- α 3 complex showed broader and fewer peaks than those of the p14 RRM complex in ^{15}N and ^{13}C -edited NOESY spectra (data not shown), and did not provide sufficient distance restraint information to calculate the 3D structure of the complex (data not shown). In contrast, as shown in Figure 3(d), the p14 RRM complex produced nicely dispersed, sharp peaks with few overlaps (see also Fig. S2). Therefore, we proceeded with the structure determination for the $^{15}\text{N}/^{13}\text{C}$ -labeled complex of p14 (8–93)–SF3b155 (379–424).

Intriguingly, the T_1 , T_2 , and NOE values of the residues in the β 1- α 1 loop (Lys29 and Ile30) and the residue close to the α 1- β 2 loop (Ala32) were distinct from those of the p14 RRM body, indicating that these regions were flexible in solution. In addition, the T_1 values of the resi-

**Figure 4**

Backbone ^{15}N relaxation parameters for individual residues of the p14-SF3b155 complex at 600 MHz ^1H frequency. T_1 (s), T_2 (s), and ^1H - ^{15}N NOEs of the p14 RRM- α 3 complex (a), the p14 RRM- α 3 (Tyr28Ala) complex (b), and the p14 RRM- α 3 (Arg85Ala) complex (c). Secondary structure elements are depicted with green (β -sheet), and yellow and red (α -helix) boxes. Residues that exhibited overlapped resonance peaks and slightly insufficient resonance qualities for the magnetic decay analyses are not shown (The p14 RRM- α 3 complex; T_1 : Tyr28, Arg57, Ile67, and Ala94 of p14. T_2 : Leu404 of SF3b155, and Tyr28, Lys29, and Arg57 of p14. The proton-nitrogen heteronuclear NOEs: Ser400 and Glu403 of SF3b155, and Leu21, Asn25, Tyr28, Lys29, Tyr36, Arg46, Glu47, Arg49, Gly58, Tyr61, Tyr64, Ile67, Cys74, Leu77, Val82, Leu87, Val88, Leu90, and Tyr92 of p14. The p14 RRM- α 3 (Tyr28Ala) complex; T_1 : Asp401 of SF3b155, and Ala28 (mutation point), Asp37, Arg57, Gly58, Tyr61, Asn72, and Leu87 of p14. T_2 : Ala28 (mutation point), Arg58, and Gly58 of p14. The proton-nitrogen heteronuclear NOEs: Glu410 of SF3b155, and Arg19, Tyr22, Asn25, Ala28 (mutation point), Asp37, Arg57, Gly58, Val63, Phe68, Asn72, His76, and Leu87 of p14. The p14 RRM- α 3 (Arg85Ala) complex; T_1 : Leu404 of SF3b155, and Leu26, Tyr36, Ile67 of p14. T_2 : Leu404 of SF3b155, and Glu16 and Tyr36 of p14. The proton-nitrogen heteronuclear NOEs: Leu404 and Lys413 of SF3b155, and Arg24, Tyr36, Asp37, Arg46, Gly58, Ile67, Leu87, and Tyr92 of p14).

Table I

Summary of Conformational Restraints and Structural Statistics for the 20 Energy-Refined Conformers of the p14-SF3b155 Complex Structure

	p14	SF3b155 (400–419)	Complex
NOE upper distance restraints:			
Intraresidual ($ i - j = 0$)	311	102	413
Sequential ($ i - j = 1$)	262	98	360
Medium-range ($1 < i - j < 5$)	150	62	212
Long-range ($ i - j \geq 5$)	378	1	379
Intermolecular			122
Total NOE			1486
Dihedral angle restraints	185		
(ϕ/ψ dihedral angle restraints from TALOS ⁴²)	139		
Distance restraint violations $> 0.2 \text{ \AA}$	0		
Dihedral angle restraint violations $> 5^\circ$	0		
CYANA target function value (\AA^2)	0.06 ± 0.03		
AMBER energies ⁴⁴ (kcal/mol):			
Total	-5190 ± 124		
van der Waals	-351 ± 19		
Electrostatic	-5759 ± 125		
	p14 (20–92)	SF3b155 (400–419)	Complex (20–92; 400–419)
Ramachandran plot statistics ⁴⁵ (%):			
Residues in most favored regions	91.2	87.7	
Residues in additionally allowed regions	8.7	12.3	
Residues in generously allowed regions	0.2	0.0	
Residues in disallowed regions	0.0	0.0	
Average r.m.s.d. from mean coordinates (\AA):			
Backbone	0.47	0.72	0.61
Heavy atoms	0.96	1.14	1.05

dues in the $\alpha 2$ - $\beta 4$ loop (Cys83 and Asn84) were distinct from those of the p14 RRM body, although the T_2 and NOE values of these residues were close to those of the p14 RRM body.

Structure determination

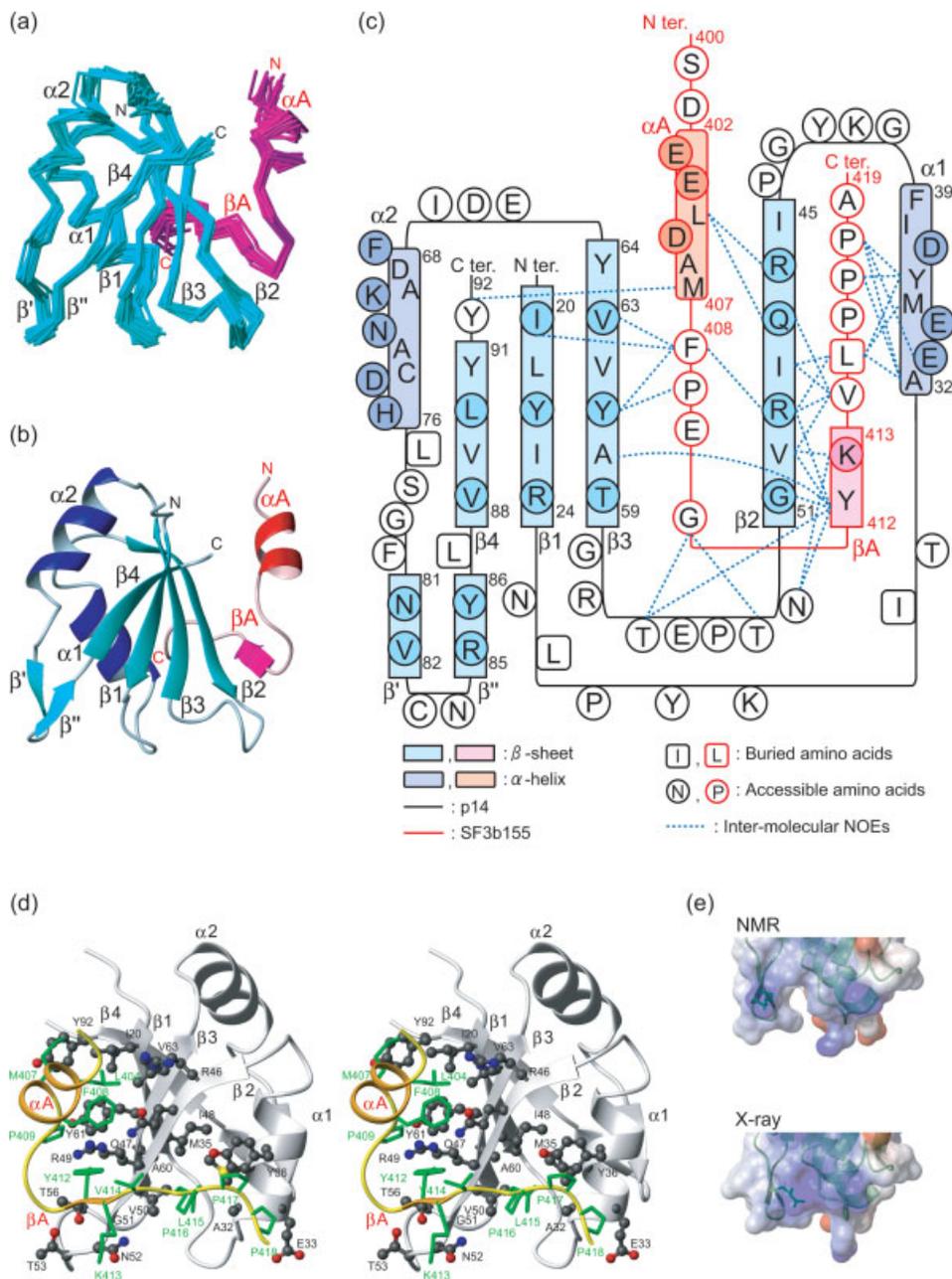
Chemical shift assignments for the p14 RRM complex were performed with standard methods, using a combination of triple resonance NMR experiments (see Materials and Methods). In total, 96% of the main-chain and 93% of the side-chain chemical shifts were assigned in residues 397–423 of the SF3b155 fragment and in residues 16–93 of p14. We solved the structure using NOE distance restraints obtained from ^{15}N - and ^{13}C -edited NOESY spectra.

The structure calculations were performed by CYANA 2.1.⁴² An average of 13 restraints per residue, including 122 intermolecular distance restraints, was utilized for the structure calculation (Table I). An ensemble of 20 NMR conformers and a ribbon representation of the lowest energy structure are shown in Figure 5(a,b), respectively. The complex structure is well defined by the NMR data for p14 (19–93) and SF3b155 (400–419). However, the N-terminal (379–399) and C-terminal

(420–424) regions of the SF3b155 fragment did not show any intermolecular NOEs to p14, and their structures are disordered in solution. Thus, we concluded that SF3b155 (400–419) is more precisely responsible for the complex formation with p14, and that the Trp386 and Trp388 residues do not contribute to the protein–protein interaction between p14 and SF3b155. The intermolecular NOEs between p14 and SF3b155 are summarized in Figure 5(c) and Table I.

Solution structure of the p14-SF3b155 complex

The p14 protein adopts a canonical RRM fold ($\beta\alpha\beta\beta\alpha\beta$) [Fig. 5(b)],^{48,49} consistent with the previous crystallographic study.²² Residues 20–24 ($\beta 1$), 45–51 ($\beta 2$), 59–64 ($\beta 3$), and 88–91 ($\beta 4$) constitute a four-stranded antiparallel β -sheet [Fig. 5(b,c)]. Helix 1 ($\alpha 1$, 32–39) and helix 2 ($\alpha 2$, 68–76) connect $\beta 1$ - $\beta 2$ and $\beta 3$ - $\beta 4$, respectively [Fig. 5(b,c)]. As predicted by a homology-modeling study combined with NMR spectra,³¹ we observed a β -hairpin structure in the loop region between $\alpha 2$ and $\beta 4$ (β' and β'' : 81–86, “ β' - β'' hairpin” hereafter) [Fig. 5(b,c)], which had not been mentioned in the crystallographic study.²² Furthermore, this β -hair-

**Figure 5**

Structure of the p14-SF3b155 complex. (a) Backbone traces of the 20 lowest energy conformers of the p14-SF3b155 complex. The p14 protein (residues 16–93) and the SF3b155 fragment (400–419) are colored cyan and magenta, respectively. (b) Ribbon diagram of the lowest energy structure of the p14-SF3b155 complex. The p14 protein (16–93) is colored cyan and blue, and the SF3b155 fragment (400–419) is magenta and red. The secondary structure elements and the chain termini are labeled. (c) Summary of the secondary structure and the intermolecular NOEs of the p14-SF3b155 complex. The secondary structure elements and the sequence numbering are indicated. (d) Stereo view of a ribbon diagram of the p14 (19–93)–SF3b155 (400–419) complex. The p14 protein is shown in gray, and the SF3b155 fragment is shown in yellow and orange. Residues involved in the complex formation are shown in green (Leu404, Met407, Phe408, Pro409, Tyr412, Lys413, Val414, Leu415, Pro416, Pro417, and Pro418 of SF3b155) and in ball-and-stick models (dark gray, carbon; red, oxygen; blue, nitrogen; yellow, sulfur) (Ile20, Ala32, Glu33, Met35, Tyr36, Arg46, Gln47, Ile48, Arg49, Val50, Gly51, Asn52, Thr53, Thr56, Ala60, Tyr61, Val63, and Tyr92 of p14). (e) Comparison between the solution and crystal (PDB ID 2F9D) structures of the p14 protein. The solution and crystal structures are superimposed on the β 1- α 1 loop, and are represented in green ribbon and electrostatic potential surface models, respectively. The side chain of the residue Arg85 is shown in green.

pin forms a groove together with the β 1- α 1 loop, which is more pronounced in solution than in the crystal structures in two different space groups (PDB IDs 2F9D,

2F9J),²² because the groove in the crystal structures is occupied by the side chain of the positively charged Arg85 residue [Fig. 5(e)].

In the complex, the SF3b155 residues 402–407 adopt an α -helical structure (α A), which is located on β 2 and β 3 of p14 [Fig. 5(b,c)]. The SF3b155 residues 412–413 form a β -strand (β A), which pairs with β 2 of p14 to form an intermolecular β -sheet [Fig. 5(b,c)]. To the best of our knowledge, this protein-protein interaction mode has not been observed in other RRM complexes with proteins.^{23,26–29}

In total, 122 intermolecular NOE signals were observed between p14 and SF3b155 [Fig. 5(c), Table I]. Hydrophobic interactions were found between L404, M407, F408, and Y412 of SF3b155 and the β -sheet surface of p14 [Fig. 5(d)]. V414 and L415 of SF3b155 are well accommodated in the hydrophobic pocket formed by the residues in α 1 and β 2 of p14 [Fig. 5(d)]. Moreover, the C-terminal portion of the SF3b155 fragment contains a stretch of three proline residues (P416, P417, and P418), which were truncated in the previous crystallographic study,²² and one of them (P417) is in the proximity of Tyr36 of p14 [Fig. 5(d)]. Collectively, F408, Y412, V414, and L415 of SF3b155 were found to play important roles in the hydrophobic interaction with p14 [Fig. 5(d)], as indicated by their numerous intermolecular NOEs [Fig. 5(c)]. In addition, electrostatic interactions were observed between p14 and SF3b155. The negatively charged amino acids (E402, E403, and D405) on α A of the SF3b155 fragment were properly located to interact with the positive charges on β 2 of p14 [Fig. 5(c)].

Crucial residues of SF3b155 for complex formation, as revealed by mutational analyses

To clarify the protein-binding mode of the p14 RRM, we performed mutational experiments to identify the residues that are involved in the interactions. We constructed 10 single and 18 combinatorial alanine mutants of SF3b155 to assess the contributions of specific residues to complex formation [Fig. 6(a,b)].

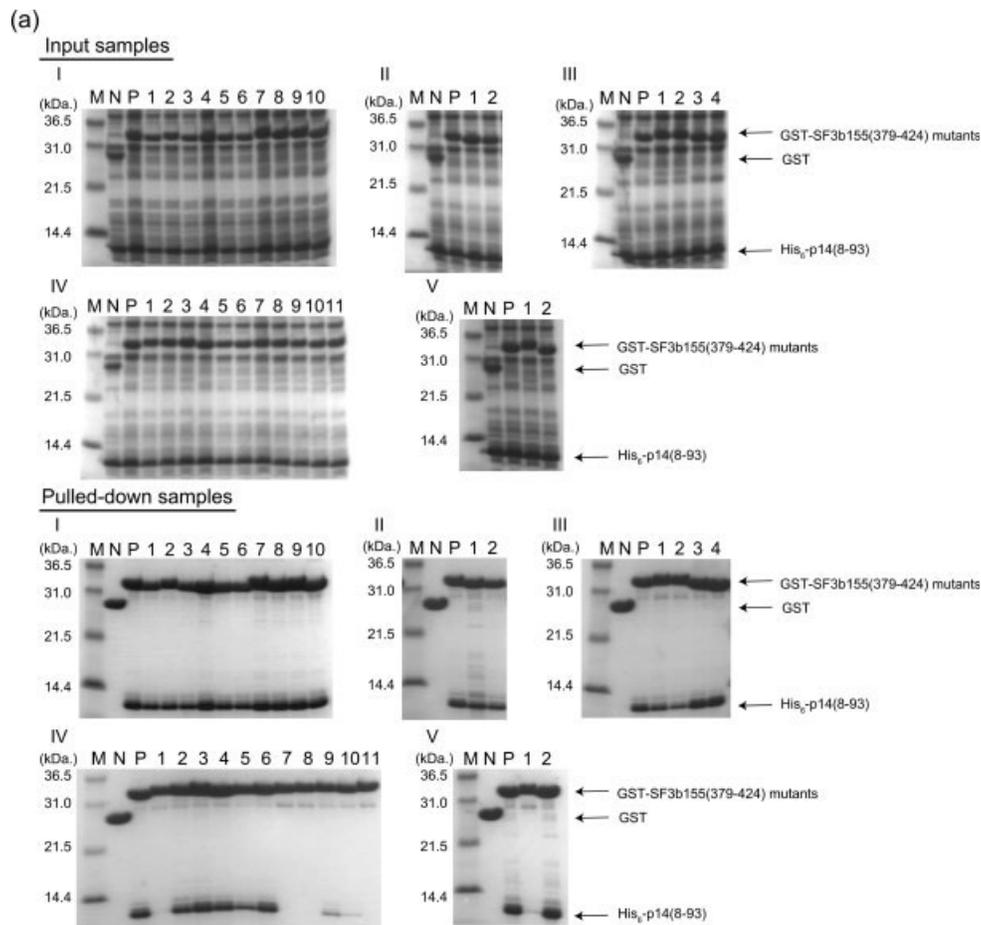
At first, we made a series of single amino acid substitution mutants in the SF3b155 fragment [Fig. 6(a[panel I],b)]. However, none of these point mutations had an obvious effect on the interaction between p14 and SF3b155 [Fig. 6(a[panel I],b)]. Next, we classified the interactions between the SF3b155 fragment and p14 into four categories: electrostatic interactions (D401, E402, E403, and D405), hydrophobic interactions by α A (L404, M407, and F408), the C-terminal proline-rich region (P416, P417, and P418), and the loop and β A (F408, F412, V414, and L415) [Fig. 6(a[panels II–V],b)]. To examine the role of each interaction category, we replaced multiple residues with alanine, and performed GST pull-down assays [Fig. 6(a[panels II–V],b)]. These analyses revealed that the electrostatic interactions [Fig. 6(a) panel II, lanes 1 and 2] and the hydrophobic interactions by the N-terminal α -helix [Fig. 6(a) panel III,

lanes 1 and 2] and the C-terminal proline-rich region [Fig. 6(a) panel III, lanes 3 and 4] had no explicit importance in complex formation [Fig. 6(a[panels II and III],b)]. Furthermore, the combinatorial mutations of the residues involved in the hydrophobic interactions by the α A- β A loop, β A, and the region downstream of β A showed that the complex formation activity disappeared if and only if F408 and Y412 and one or both of the residues V414 and L415 were replaced with alanine [Fig. 6(a[panel IV, lanes 7, 8, and 11],b)]. The simultaneous substitution of the two aromatic residues [Fig. 6(a[panels IV and V],b)] and the substitution of V414 and L415 combined with one of the aromatic residues [Fig. 6(a[panel IV],b)] decreased the complex formation ability. Furthermore, the replacement of Y412 with phenylalanine had no effect on the interaction between p14 and SF3b155 [Fig. 6(a[panel V, lane 2],b)], revealing that the tyrosine OH group does not contribute to the protein-protein interaction. As shown in the upper part of Figure 6(a), the mutations did not affect the amount of protein expression. Collectively, we conclude that SF3b155 binds to p14 through a three-point mounting formed by hydrophobic interactions with two aromatic residues (F408 and Y412) and the hydrophobic residues backing the intermolecular β -sheet (V414 and L415) [see Fig. 5(d)].

Interaction between p14-SF3b155 and single-stranded branch site RNA

A recent extensive UV crosslinking study showed that p14 efficiently binds to the single-stranded branch site RNA, rather than to the heteroduplex of U2 snRNA and the branch site RNA, when it is included both in the SF3b particle and in the recombinant complex with SF3b155.³¹ Therefore, we examined the effect of the presence of the single stranded branch site RNA (AdML branch site sequence: 5'-GUGCUGACCCUG-3'^{15,17,18}) on the chemical shifts of the main-chain amide protons in the p14-SF3b155 complex.

First, we investigated the effect of the presence of RNA on the p14 RRM- α 3 complex, which includes RRM and the semiconserved C-terminal extension α 3 (94–102) of p14 [Fig. 7(b)]. The crosspeaks of the residues located in the β 1- α 1 loop (Asn25, Leu26, and Tyr28) and in the β' - β'' hairpin (Gly79, Asn84, and Leu87) that form the positively charged groove were significantly affected [Fig. 7(b)]. The crosspeak of His76 in the α 2 helix, which is located near the β' - β'' hairpin, was also significantly affected [Fig. 7(b)]. In addition to the crosspeaks of the groove residues, most of the crosspeaks corresponding to the C-terminal extension (Ala94, Asn95, Arg96, Phe98, and Lys100 in p14), as well as those derived from the residues located close to the C-terminal extension (Ile20, Leu90, and Tyr91 in p14 and Phe408 in SF3b155), were significantly affected. On the other hand, the crosspeaks of the RNP aromatic residues (Tyr22 and Tyr61) and the



(b)

SF3b155 (379-424)	Lane No.	400	α A					β A					419	Complex formation										
		S	D	E	E	L	D	A	M	F	P	E	G	Y	K	V	L	P	P	P	A	
Single point mutants	I-1				○																			++
	I-2																							++
	I-3																							++
	I-4																							++
	I-5																							++
	I-6																							++
	I-7																							++
	I-8																							++
	I-9																							++
	I-10																							++
Electrostatic interaction	II-1			○	○	○	○						○											++
	II-2																							++
N-terminal (α -helix)	III-1																							++
	III-2																							++
C-terminal (poly Pro)	III-3																							++
	III-4																							++
Hydrophobic interaction	IV-1, V-1																							±
	IV-2																							++
	IV-3																							++
	IV-4																							++
	IV-5																							++
	IV-6																							++
	IV-7																							-
	IV-8																							-
	IV-9																							±
	IV-10																							±
	IV-11																							-
V-2																							++	

○ : alanine substitution ○(F) : phenylalanine substitution

Figure 6

GST pull-down assays to test for complex formation between p14 and SF3b155 mutants. Bound fractions were analyzed by SDS-PAGE, as indicated in I-V of (a). The mutations and the complex formation efficiencies are summarized in (b). ++, +, ±, and - indicate the efficiencies of the complex formation.

residues in SF3b155, except for Phe408, were only slightly affected at the baseline level [Figs. 7(b) and 8(a,b)]. Canonical RRM motifs recognize RNA molecules on the β -sheet with the RNP aromatic residues stacking onto the RNA bases, and significant chemical shift perturbations of main-chain ^1H - ^{15}N have been observed for the residues in the β -sheet, especially for the RNP aromatic residues.^{50–52} In contrast, our data strongly suggest that in the case of the p14-SF3b155 complex, the aromatic residues on RNP1 and RNP2 are less likely to form stacking interactions with the RNA bases, and that the positively charged groove and the C-terminal extension play more important roles for the RNA recognition.

Second, in order to evaluate the contribution of the C-terminal extension to RNA binding, we investigated how RNA binding affects the p14 RRM complex lacking the C-terminal extension. As shown in Figure 7(a), while the crosspeaks of the RNP aromatic residues (Tyr22 and Tyr61) and the SF3b155 residues were only negligibly

affected [Figs. 7(a) and 8(a,b)], those of the residues in the β 1- α 1 loop (Asn25, Leu26, Tyr28, and Lys29) and in the β' - β'' hairpin (Gly79 and Asn84), which together form the positively charged groove, were affected significantly by RNA binding [Figs. 7(a) and 9(a)]. Note that, all of the crosspeaks, which were affected in the p14 RRM complex, also exhibited chemical shift change in the p14 RRM- α 3 complex, except for that of Lys 29 [Fig. 7(a,b)]. These observations strongly suggest that the RNA-binding activity of the groove is independent of the C-terminal extension.

Furthermore, we performed site-directed mutagenesis to examine whether the RNA-binding activity of the C-terminal extension depends on the positively charged groove. It is well known that aromatic amino acids are often necessary for RNA recognition (e.g., the RNP aromatic residues in canonical RRMs).³⁰ Thus, we speculated that an aromatic amino acid residue in the positively charged groove of the p14 RRM complex could play a critical role in the RNA binding. Accordingly, we focused on the only aromatic residue located inside the groove, Tyr28, which showed a significant chemical shift perturbation [Figs. 7(a,b) and 8(c)]. We substituted Tyr28 with Ala (p14 RRM- α 3 Tyr28Ala complex). The chemical shift values in the mutant protein were not affected by the mutation, except for those of Lys29 and Arg57, which are in the proximity of Tyr28 [Fig. 9(a,d)]. We also analyzed the dynamics of the p14 RRM- α 3

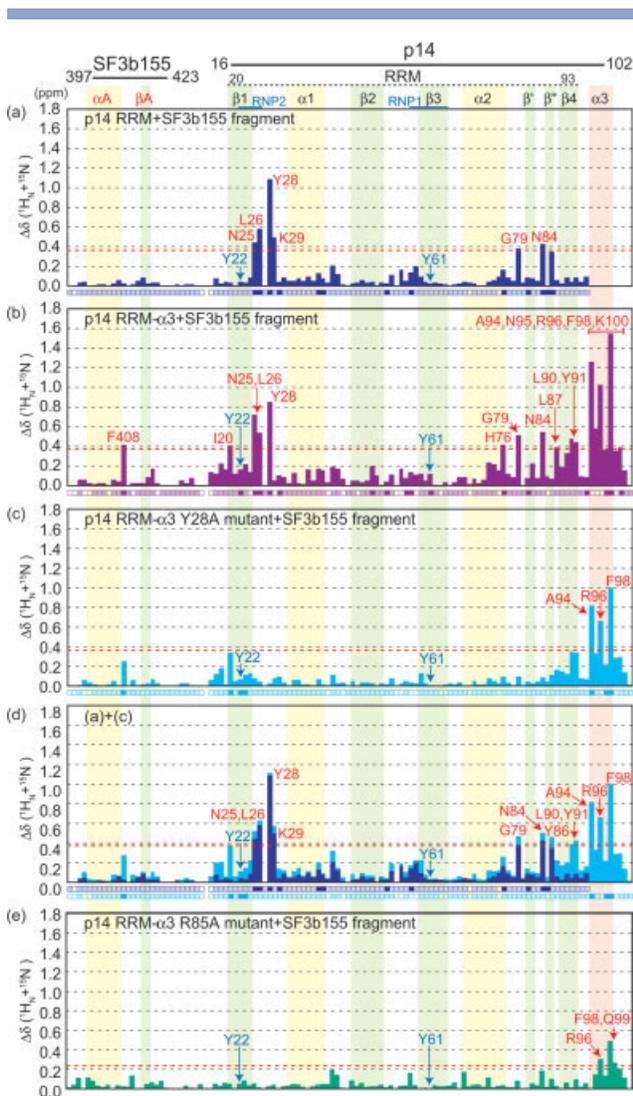
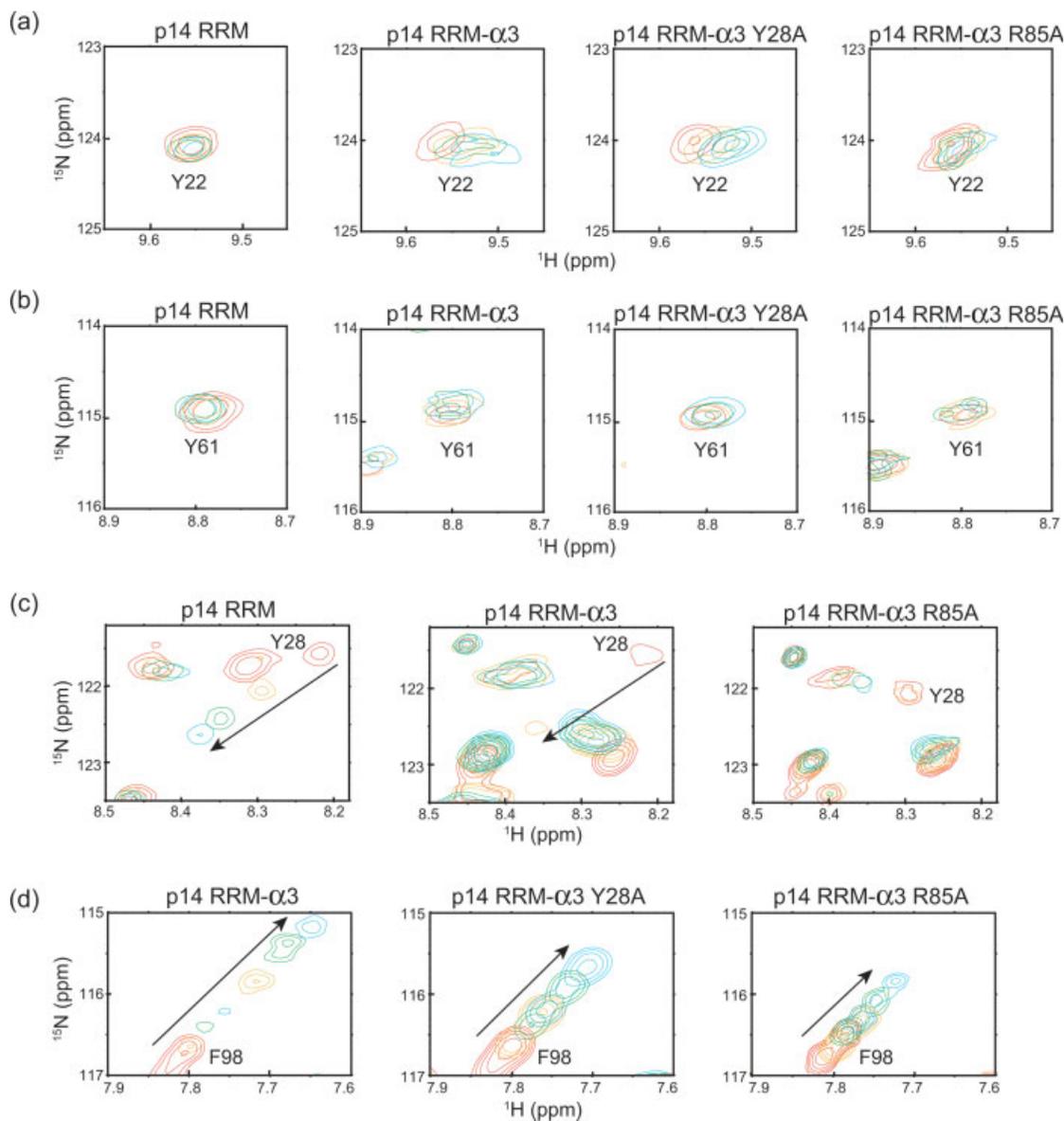


Figure 7

NMR chemical shift perturbation of p14-SF3b155 complexes upon RNA binding. (a) The p14 RRM complex. (b) The wild-type p14 RRM- α 3 complex. (c) The p14 RRM- α 3 Tyr28Ala complex. (d) The sum of the perturbation values in (a) and (c). (e) The p14 RRM- α 3 Arg85Ala complex. The RNA was added to the protein solutions at a 1:3 ratio of complex/RNA. The absolute value of the chemical shift change $\Delta\delta$ ($^{15}\text{N} + ^1\text{H}_\text{N}$), which combines the backbone amide proton and nitrogen chemical shift perturbation values, is shown for each residue. $\Delta\delta$ ($^{15}\text{N} + ^1\text{H}_\text{N}$) was calculated as follows: $\Delta\delta$ ($^{15}\text{N} + ^1\text{H}_\text{N}$) = $\sqrt{(\delta_{15\text{N}})^2 + (\delta_{1\text{H}})^2}$. The baseline of the amide perturbation was defined as follows: the perturbation values from the three constructs (a–c) were merged, and the average of the smallest 95% was defined as the baseline (i.e., 0.088 ppm). The perturbation values greater than the baseline plus three times the standard deviation of the baseline (3×0.097 ppm) were defined as significant perturbations (i.e., the significant level is 0.38 ppm). As the p14 RRM- α 3 Arg85Ala complex (e) showed small chemical shift changes, in contrast to the results for the other three complexes (a–c), the significant levels of this complex were independently calculated, as described above (i.e., 0.26 ppm). Blue letters indicate the two conserved aromatic amino acids, Tyr22 and Tyr61, on β 1 and β 3. Red letters indicate amino acid residues that showed significant chemical shift changes. In (b) and (e), because the chemical shift resonances of Tyr28 at the 1:2 and 1:3 ratios of the complex/RNA disappeared, the absolute values of the chemical shift change $\Delta\delta$ of Tyr28 were calculated from the data at the 1:1 ratio of the complex/RNA. The boxes under the histogram represent side-chain chemical shift assignments in the [^1H , ^{13}C]-HSQC spectrum, and the clear boxes represent residues with no assigned side-chain atoms. Light-colored boxes represent residues with at least one assigned side-chain atom and no side-chain atom showing a significant perturbation. Strong-colored boxes represent residues for which at least one side-chain atom showed a significant perturbation. The threshold level of 0.075 ppm for significant side-chain perturbations was defined in the same way as that for the main chain perturbations. We could assign 59, 40, and 59% of the side-chain atoms in the p14 RRM complex, the p14 RRM- α 3 complex, and the p14 RRM- α 3 Tyr28Ala complex, respectively.

**Figure 8**

(a–d) Close-up views of the ^1H - ^{15}N HSQC spectrum of the p14 RRM complex, the p14 RRM- α 3 complex, the p14 RRM- α 3 Tyr28Ala complex, and the p14 RRM- α 3 Arg85Ala complex, showing selected amide shift changes in the absence (red) and presence (orange, green, cyan; ratio complex:RNA = 1:1, 1:2, 1:3) of the AdML branch site RNA (sequence 5'-GUGCUGACCCUG-3'). In (c), middle and right panels, the chemical shift resonances of Tyr28 at the 1:2 and 1:3 ratios of complex/RNA disappeared.

Tyr28Ala complex by measuring the nitrogen relaxation times T_1 and T_2 , and the proton-nitrogen heteronuclear NOEs, and confirmed that these values were consistent with those of the p14 RRM- α 3 complex [Fig. 4(b)]. These results indicated that the Tyr28Ala mutation caused no significant change in the p14-SF3b155 structure. Strikingly, when the Tyr28Ala mutant complex was titrated with the RNA, no residues in the groove showed chemical shift changes, in sharp contrast to the result for

the p14 RRM- α 3 complex [Fig. 7(b,c)]. On the other hand, the chemical shift perturbations of the residues in the C-terminal extension were essentially conserved (Ala94, Arg96, and Phe98) [Figs. 7(c) and 8(d)].

In addition, to examine whether the β' - β'' hairpin, which forms the positively charged groove with the β 1- α 1 loop, contributes to the interaction with RNA, we focused on the positively charged residue located inside the groove, Arg85 [Fig. 9(c)]. We substituted Arg85 with

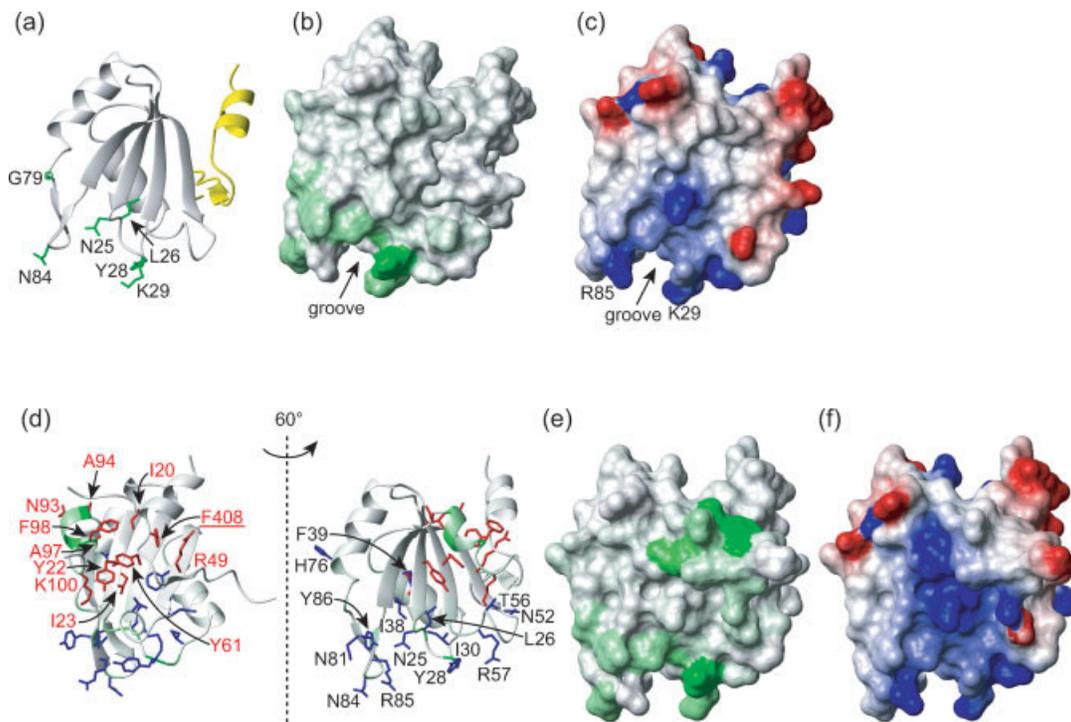


Figure 9

The positively charged groove and the C-terminal extension of the p14-SF3b155 complex interact with RNA. (a) Ribbon representation of the p14 RRM complex. The p14 protein is shown in gray, and the SF3b155 fragment is shown in yellow. Amino acid residues with significant (see Fig. 7) ^{15}N and $^1\text{H}^{\text{N}}$ chemical shift changes upon RNA binding are represented in green and labeled. (b) Surface representation of the p14 RRM complex. Distinct ^{15}N and $^1\text{H}^{\text{N}}$ chemical shift changes are mapped onto the surface representation. Residues are colored based on the magnitude of the chemical shift changes upon RNA binding, ranging from white (change insignificant or not measured) to green (largest chemical shift change). (c) Electrostatic potential surface of the p14 RRM complex. Regions of negative and positive potential are colored red and blue, respectively. The positively charged residues in the groove are labeled. (d) Ribbon representation of the p14 RRM- α 3 complex. The C-terminal extension was imported to the solution structure from the crystal structure.²² Backbone ^{15}N and $^1\text{H}^{\text{N}}$ chemical shift changes are mapped onto the ribbon representation, as described in (b). Significant chemical shift changes of $^1\text{H}^{\text{N}}$ - ^{13}C groups upon RNA binding are represented in blue for residues of the positively charged groove and in red for residues of the C-terminal extension. (e) Surface representation of the p14 RRM- α 3 complex. Distinct ^{15}N and $^1\text{H}^{\text{N}}$ chemical shift changes are mapped onto the surface representation, as described in (b). (f) Electrostatic potential surface of the p14 RRM- α 3 complex, as described in (c).

Ala (p14 RRM- α 3 Arg85Ala complex). The chemical shift values in the mutant protein were not affected by the mutation, except for those of Asn25, Tyr28, and Val82-Leu87, which are in the proximity of Arg85 [Fig. 5(c) and 9(d)]. We also analyzed the dynamics of the p14 RRM- α 3 Arg85Ala complex by measuring the nitrogen relaxation times T_1 and T_2 , and the proton-nitrogen heteronuclear NOEs, and confirmed that these values were consistent with those of the p14 RRM- α 3 complex [Fig. 4(c)]. These results indicated that the Arg85Ala mutation caused no significant change in the p14-SF3b155 structure. After binding to RNA, the crosspeaks of the residues in the positively charged groove and the RNP aromatic residues (Tyr22 and Tyr61) were only slightly affected at the baseline level [Figs. 7(e) and 8(a–c)]. On the other hand, the crosspeaks of the residues in the C-terminal extension were affected by RNA (Arg96, Phe98, and Gln99) [Figs. 7(e) and 8(d)], although this mutant complex displayed small chemical shift changes, in

contrast to the results for the other three complexes (see Fig. 7).

Consequently, these results clearly show that the RNA-binding activity of the positively charged groove requires the aromatic residue Tyr28 and the positively charged residue Arg85, and is not based solely on electrostatic interactions. Taken together, these data suggest that the RNA-binding activities of the groove and the C-terminal extension are essentially independent of each other. As shown in Figure 7(d), the perturbation profile from the wild-type p14 RRM- α 3 complex can almost be reconstructed by the sum of those from the p14-RRM complex and the p14 RRM- α 3 Tyr28Ala mutant [Fig. 7(a,c)].

Finally, we assessed the effect of the RNA on the main- and side-chain ^1H - ^{13}C group chemical shifts of the p14-SF3b155 complex, using [^1H , ^{13}C]-HSQC spectra. As the p14 RRM- α 3 Arg85Ala complex showed small chemical shift changes [Fig. 7(e)], in contrast to the results for the other three complexes [Fig. 7(a–c)], we compared the

three complexes. Although the quality of the NMR spectrum was severely impaired upon RNA addition, we could assign the resonances originating from the side-chain atoms of almost all of the residues (see the legend of Fig. 7) in the three constructs, as shown in Figures 7(a–c).

In agreement with the data from the main-chain ^1H - ^{15}N chemical shift perturbation experiments, the residues whose crosspeaks were affected in the p14 RRM complex upon RNA binding were located mainly in the groove [Asn25, Leu26, Tyr28, Ile30, Asn81, Asn84, Arg85, and Tyr86; Fig. 7(a)]. In addition, the crosspeaks of Asn52 and Arg57 in the β 2- β 3 loop, Ile38 and Phe39 in the α 1-helix, and His76 in the α 2-helix were affected. In the p14 RRM- α 3 complex, on the other hand, the crosspeaks of the side-chains of the residues located in or proximal to the C-terminal extension (Phe408 in SF3b155; Tyr22, Tyr61, Asn93, Ala97, Phe98, and Lys100 in p14) were affected in addition to those in or near the positively charged groove (Asn25, Leu26, Tyr28, Ile30, His76, and Asn81) [Fig. 7(b)], consistent with the main chain data. For five residues in p14 (Asn52, Thr56, and Arg57 in the β 2- β 3 loop, Arg49 adjacent to Phe408 of SF3b155, and Tyr61), RNA binding caused the chemical shift change for side-chain resonances, but not for main-chain resonances. In the Tyr28Ala mutant, the crosspeaks for the side-chains of the residues in the groove were hardly affected by RNA binding, in contrast to those in or proximal to the C-terminal extension (Phe408 in SF3b155; Ile20, Tyr22, Ile23, Ala94, Ala97, and Phe98 in p14) [Fig. 7(c)], including the RNP aromatic residue (Tyr22) [Fig. 7(c)].

These results are summarized on the model structure of the p14 RRM- α 3 complex in Figure 9(d,e). The C-terminal extension was imported to the solution structure from the crystal structure,²² and was predicted to be located on the β -sheet. A large, positively charged patch was predicted to exist between the groove and the C-terminal extension [Fig. 9(f)]. The residues with main-chains and side-chains whose crosspeaks were affected by RNA titration are located within this patch. Moreover, the RNP aromatic residues (Tyr22 and Tyr61) are predicted to be located between the groove and the C-terminal extension. Additionally, the chemical shift perturbation of His76 in the α 2 helix, which is located opposite to the β -sheet, suggested that the RNA could pass through the groove. These findings suggest that the p14 RRM- α 3 complex provides a continuous RNA-binding surface in a positively charged patch.

DISCUSSION

Protein-protein interactions on the p14 RRM

Our mutational analyses revealed that the SF3b155 fragment (400–419) contacts three points on the β -sheet

surface of p14 RRM (20–92), using Phe408, Tyr412, Val414, and Leu415 [Fig. 6]. As shown in Figure 5(c), the hydrophobic residues on one of the edges of the β -sheet (Ile20 and Val63) of p14 make direct contacts with Phe408 of SF3b155. On the other hand, the small residues on the other edge (Gly51 and Thr59) form a pocket to accommodate the aromatic ring of Tyr412 [Fig. 5(c,d)].

To determine whether these residues are conserved among RRM, we performed a multiple sequence alignment of 109 RRM from the Pfam database (see Materials and Methods)⁵³ and calculated the frequency of appearance of amino acids at each position in the RRM [Fig. 10(a)]. As shown in Figure 10(a), the buried hydrophobic residues (Leu21, Ile23, Ile48, Val50, Val62, and Val89 in human p14) are highly conserved. The aromatic residues in the RNP motifs (Tyr22 and Tyr61 in p14) are also conserved among the RRM, and among the p14 proteins from different species (see Fig. 1). On the other hand, the positions corresponding to Ile20 and Val63 in p14 were predominantly occupied by neutral or charged amino acids [Fig. 10(a)]. Moreover, the positions of Gly51 and Thr59 are generally occupied by aliphatic and aromatic residues, respectively [Fig. 10(a)]. These results indicate that the residues of p14 involved in the critical interactions with SF3b155 are the exceptions among the RRM. The noncanonical composition of amino acids on the surface of the β -sheet is consistent with the peculiar features of the p14 RRM that bind protein, rather than RNA.

RNA recognition by the p14 protein and by other RRM

There have been two controversial arguments about the amino acid residues in the p14-SF3b155 complex that are responsible for the RNA-binding activity. One described the importance of the RNP aromatic residue, Tyr22, on the basis of the UV crosslinking assay for the Tyr22Met mutant.²² However, a mutant with Tyr22 replaced with methionine could bind to RNA, indicating that the stacking interaction mediated by the aromatic moiety of Tyr22 is dispensable, although the aromatic residue can contact the RNA.²² On the other hand, a previous NMR perturbation experiment indicated that the β' - β'' hairpin and the C-terminal extension α 3 are more important than the aromatic residues on RNP motifs (Tyr22),³¹ although the previous NMR study did not describe the solution structure of the p14-SF3b155 complex.

In the present study, we solved the solution structure of the p14-SF3b155 complex with high quality NMR spectra obtained at a functional temperature for human proteins (35°C). This allowed us to assess the contributions of the relevant amino acid residues in RNA binding by NMR perturbation in combination with mutational analyses. As shown in Figure 7, our results revealed that

[Fig. 10(b)].^{51,54,55} In contrast to the β 2- β 3 loop, which is not involved in the groove of p14, the β 1- α 1 loop and the β' - β'' hairpin adopted similar conformations to those from the other noncanonical RRM; the conformation of the β 1- α 1 loop is quite similar to that from Fox-1, and the conformation of the β' - β'' hairpin resembles those from Fox-1 and hnRNP F qRRM2 [Fig. 10(b)]. Figure 10(c) shows the considerable similarity of the detailed structures of the p14 and Fox-1 RRM. In Fox-1, two RNA bases stack on each side of Phe126, which corresponds to Tyr28 in p14, and are surrounded by Arg127, Arg153, and Arg184 [Fig. 10(c)].⁵¹ The distances between the aromatic ring of Phe126 (C^δ and/or C^ϵ) and the positively charged N^ϵ atoms of Arg127, Arg153, and Arg184 are 6–11 Å. It should be noted that the counterparts of Fox-1 Phe126 and Arg184 in p14, Tyr28 and Arg85, respectively, were also shown to be essential for the RNA-binding activity of the groove. Moreover, the corresponding residues in p14, Tyr28, Lys29, Arg57, and Arg85, have a similar spatial arrangement [Fig. 10(c)]. The β 2- β 3 loop structure of p14 is different from that of Fox-1 [Fig. 10(c)], presumably because the β 2- β 3 loop of p14 is involved in the p14-SF3b155 complex formation. However, the distances between the aromatic ring of Tyr28 (C^δ and/or C^ϵ) and the positively charged N^ζ/N^ϵ atoms of Lys29, Arg57, and Arg85 vary between 4 and 13 Å. Accordingly, these residues are flexible, suggesting that they could accommodate two RNA bases. Therefore, one may speculate that the RNA recognition mode of the groove of p14 is similar to that of Fox-1. On the other hand, Fox-1 also recognizes RNA bases through the canonical RNP aromatic residues,⁵¹ which is not the case for p14. These comparisons suggest that the RNA recognition mode of p14 is novel among the RRM with no structural homology. Furthermore, the positively charged groove and the C-terminal extension of the p14 protein interact with RNA in an independent manner, which may contribute to the dynamic spliceosome assembly of the U2 snRNP.

A more complete understanding of the RNA recognition mode of p14-SF3b155 would be obtained from the structure of the p14-SF3b155 complex bound to the target RNA. The present study provides fundamental information toward further structural studies of the spliceosome.

CONCLUSION

The p14 protein is a key component of the U2 snRNP, which plays an important role in mRNA splicing by recognizing the branch site adenosine.¹⁷ Although the crystal structure of p14-SF3b155 is available,²² precise understanding of the complex formation and the RNA-binding mode by p14 and SF3b155 required detailed functional analyses of the residues involved in the interaction. Here

we identified the fragments of p14 and SF3b155 that are essential for their stable interaction, and solved the tertiary structure of this minimal complex in solution. In combination with extensive mutational analyses, we then identified three contact points for the peptide binding. Furthermore, we investigated the chemical shift perturbations, not only for the main-chain but also for the side-chain resonances, for several p14-SF3b155 complex constructs upon RNA binding. Our results revealed the noncanonical RNA recognition mode of p14 RRM and strongly suggested that the positively charged groove and the C-terminal extension of the p14 RRM (α 3) independently contribute to RNA binding. Collectively, these findings clearly revealed the functionally validated roles of the residues of p14 and SF3b155 in the complex formation and the RNA-binding manner.

Accession codes

The atomic coordinates for the ensemble of 20 energy-refined NMR conformers that represent the solution structure of the p14-SF3b155 complex have been deposited in the Protein Data Bank, with the accession code 2FHO.

ACKNOWLEDGMENTS

We are grateful to Dr. R. Reed for kindly providing the SF3b155 cDNA clone. We are grateful to Dr. T. Kigawa for help with the primary protein expression in the *E. coli* cell-free system, and to Dr. S. Unzai and Mr. T. Tomizawa for help with the primary characterization of the purified proteins. We are grateful to Dr. S. Koshiba and Dr. N. Tochio for help with the NMR measurements. We thank Dr. N. Kobayashi, Dr. F. He, Dr. T. Nagata, and Ms. S. Suzuki for help with the NMR data analysis, structure calculations, and structure refinement, and to Dr. T. Kaminishi for help with the structure analysis of the U2 snRNP. We are grateful to Dr. K. Kurimoto for helpful discussions about the experiments and the manuscript preparation. We would like to thank Ms. A. Ishii, Ms. K. Yajima, and Ms. T. Nakayama for help with the preparation of the manuscript.

REFERENCES

1. Staley JP, Guthrie C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 1998;92:315–326.
2. Burge CB, Tuschl T, Sharp PA. Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RR, Cech TR, Atkins JF, editors. *The RNA World II*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1999. pp 525–560.
3. Black DL, Steitz JA. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell* 1986;46:697–704.
4. Bindereif A, Green MR. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J* 1987;6:2415–2424.
5. Tam WY, Steitz JA. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem Sci* 1997;22:132–137.

6. Zhuang Y, Weiner AM. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 1986;46:827–835.
7. Seraphin B, Kretzner L, Rosbash M. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J* 1988;7:2533–2538.
8. Michaud S, Reed R. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev* 1991;5:2534–2546.
9. Nilsen TW. RNA-RNA interactions in nuclear pre-mRNA splicing. In: Simons R, Grunberg-Manago M, editors. *RNA structure and function*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1997. pp 279–307.
10. Gozani O, Feld R, Reed R. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev* 1996;10:233–243.
11. Das BK, Xia L, Palandjian L, Gozani O, Chyung Y, Reed R. Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. *Mol Cell Biol* 1999;19:6796–6802.
12. Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M, Luhrmann R. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *EMBO J* 2002;21:4978–4988.
13. Champion-Arnaud P, Reed R. The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev* 1994;8:1974–1983.
14. Gozani O, Potashkin J, Reed R. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 1998;18:4752–4760.
15. MacMillan AM, Query CC, Allerson CR, Chen S, Verdine GL, Sharp PA. Dynamic association of proteins with the pre-mRNA branch region. *Genes Dev* 1994;8:3008–3020.
16. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882.
17. Query CC, Strobel SA, Sharp PA. Three recognition events at the branch-site adenine. *EMBO J* 1996;15:1392–1402.
18. Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Luhrmann R, Query CC. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* 2001;20:4536–4546.
19. Tarn WY, Steitz JA. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* 1996;84:801–811.
20. Golas MM, Sander B, Will CL, Luhrmann R, Stark H. Molecular architecture of the multiprotein splicing factor SF3b. *Science* 2003;300:980–984.
21. Golas MM, Sander B, Will CL, Luhrmann R, Stark H. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell* 2005;17:869–883.
22. Schellenberg MJ, Edwards RA, Ritchie DB, Kent OA, Golas MM, Stark H, Luhrmann R, Glover JN, MacMillan AM. Crystal structure of a core spliceosomal protein interface. *Proc Natl Acad Sci USA* 2006;103:1266–1271.
23. Kielkopf CL, Lucke S, Green MR. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 2004;18:1513–1526.
24. Kielkopf CL, Rodionova NA, Green MR, Burley SK. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* 2001;106:595–605.
25. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* 2003;11:965–976.
26. Fribourg S, Gatfield D, Izaurralde E, Conti E. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* 2003;10:433–439.
27. Kadlec J, Izaurralde E, Cusack S. The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. *Nat Struct Mol Biol* 2004;11:330–337.
28. Price SR, Evans PR, Nagai K. Crystal structure of the spliceosomal U2B⁺-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 1998;394:645–650.
29. Mazza C, Ohno M, Segref A, Mattaj JW, Cusack S. Crystal structure of the human nuclear cap binding complex. *Mol Cell* 2001;8:383–396.
30. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 1994;372:432–438.
31. Spadaccini R, Reidt U, Dybkov O, Will C, Frank R, Stier G, Corsini L, Wahl MC, Luhrmann R, Sattler M. Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. *RNA* 2006;12:410–425.
32. Kuwasako K, He F, Inoue M, Tanaka A, Sugano S, Güntert P, Muto Y, Yokoyama S. Solution structures of the SURP domains and the subunit-assembly mechanism within the splicing factor SF3a complex in 17S U2 snRNP. *Structure* 2006;14:1677–1689.
33. Ito W, Ishiguro H, Kurosawa Y. A general method for introducing a series of mutations into cloned DNA using the polymerase chain reaction. *Gene* 1991;102:67–70.
34. Bax A. Multidimensional nuclear magnetic resonance methods for protein studies. *Curr Opin Struct Biol* 1994;4:738–744.
35. Kay LE. NMR methods for the study of protein structure and dynamics. *Biochem Cell Biol* 1997;75:1–15.
36. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 1995;6:277–293.
37. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 2004;278:313–352.
38. Kobayashi N, Iwahara J, Koshiba S, Tomizawa T, Tochio N, Güntert P, Kigawa T, Yokoyama S. KUIJIRA, a package of integrated modules for systematic and interactive analysis of NMR data directed to high-throughput NMR structure studies. *J Biomol NMR* 2007;39:31–52.
39. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation. *Biochemistry* 1994;33:5984–6003.
40. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 2002;319:209–227.
41. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 1997;273:283–298.
42. Güntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 2004;278:353–378.
43. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 1999;13:289–302.
44. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996;14:29–32, 51–55.
45. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995;117:5179–5197.
46. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 1996;8:477–486.

47. Cass DM, Berglund JA. The SF3b155 N-terminal domain is a scaffold important for splicing. *Biochemistry* 2006;45:10092–10101.
48. Burd CG, Dreyfuss G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* 1994;13:1197–1204.
49. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* 1994;265:615–621.
50. Howe PW, Allain FH, Varani G, Neuhaus D. Determination of the NMR structure of the complex between U1A protein and its RNA polyadenylation inhibition element. *J Biomol NMR* 1998;11:59–84.
51. Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, Allain FH. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* 2006;25:163–173.
52. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, Lian LY, Wilson SA, Allain FH. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J* 2006;25:5126–5137.
53. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. Pfam: clans, web tools and services. *Nucleic Acids Res* 2006;34:D247–D251.
54. Volpon L, D’Orso I, Young CR, Frasch AC, Gehring K. NMR structural study of TcUBP1, a single RRM domain protein from *Trypanosoma cruzi*: contribution of a beta hairpin to RNA binding. *Biochemistry* 2005;44:3708–3717.
55. Dominguez C, Allain FH. 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* 2006;34:3634–3645.