Structure *14*, 1677–1689, November 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.09.009

Solution Structures of the SURP Domains and the Subunit-Assembly Mechanism within the Splicing Factor SF3a Complex in 17S U2 snRNP

Kanako Kuwasako,[1,5] Fahu He,[1,5] Makoto Inoue,[1] Akiko Tanaka,[1] Sumio Sugano,[3] Peter Güntert,[1,2] Yutaka Muto,[1,*] and Shigeyuki Yokoyama[1,4,*]

[1] Protein Research Group
[2] Tatsuo Miyazawa Memorial Program
RIKEN Genomic Sciences Center
Yokohama 230-0045
Japan
[3] Laboratory of Functional Genomics
Department of Medical Genome Sciences
Graduate School of Frontier Sciences
The University of Tokyo
4-6-1 Shirokanedai, Minato-ku
Tokyo, 108-8639
Japan
[4] Department of Biophysics and Biochemistry
Graduate School of Sciences
The University of Tokyo
Tokyo 113-0033
Japan

Summary

The SF3a complex, consisting of SF3a60, SF3a66, and SF3a120, in 17S U2 snRNP is crucial to spliceosomal assembly. SF3a120 contains two tandem SURP domains (SURP1 and SURP2), and SURP2 is responsible for binding to SF3a60. We found that the SURP2 fragment forms a stable complex with an SF3a60 fragment (residues 71–107) and solved its structure by NMR spectroscopy. SURP2 exhibits a fold of the α1-α2-3_{10}-α3 topology, and the SF3a60 fragment forms an amphipathic α helix intimately contacting α1 of SURP2. We also solved the SURP1 structure, which has the same fold as SURP2. The protein-binding interface of SURP2 is quite similar to the corresponding surface of SURP1, except for two amino acid residues. One of them, Leu169, is characteristic of SF3a120 SURP2 among SURP domains. Mutagenesis showed that this single Leu residue is the critical determinant for complex formation, which reveals the protein recognition mechanism in the subunit assembly.

Introduction

mRNA splicing is an essential event in the posttranscriptional process of gene expression and is thought to be regulated by hundreds of splicing-associated proteins (Dziembowski et al., 2004; Hastings and Krainer, 2001; Jurica and Moore, 2003; Nilsen, 2003; Varani and Nagai, 1998; Will et al., 2002; Zhu and Krainer, 2000). Splicing occurs through two successive *trans*-esterification reactions. In step 1, the conserved adenosine in the branch site cleaves the 5′ splice site by a nucleophilic attack, and in step 2, the 5′ splice site attacks the 3′ splice site to release the intron (Query et al., 1994). These sequential reactions are carried out by the spliceosome, a large and highly dynamic macromolecular complex containing a series of small nuclear ribonucleoproteins, the U1, U2, U4, U5, and U6 snRNPs (Jurica and Moore, 2003; Nilsen, 2003; Staley and Guthrie, 1998). Each of the snRNPs contains a small nuclear RNA (U snRNA) and protein components (Jurica and Moore, 2003; Krämer, 1996).

The U2 snRNP, an essential component of the spliceosome, binds to the pre-mRNA branch site via base-pairing with the complementary RNA sequence of the U2 snRNA and interacts with the U1 snRNP that binds to the 5′ splice site (Michaud and Reed, 1991; Séraphin et al., 1988; Zhuang and Weiner, 1986). This U1 snRNP•U2 snRNP•pre-mRNA ternary complex recruits the U4/U6•U5 tri-snRNPs and forms the functionally active spliceosome, which catalyzes the two *trans*-esterification reactions (Jurica and Moore, 2003; Nilsen, 2003). The U2 snRNP is a large protein-RNA complex containing the core 12S U2 snRNP particle (the U2 snRNA, the Sm core proteins, and the U2A′ and U2B″ complex) (Caspary and Séraphin, 1998; Kambach et al., 1999; Lührmann et al., 1990; Price et al., 1998; Raker et al., 1996; Scherly et al., 1990; Tang et al., 1996) and the SF3a and SF3b protein complexes (Brosi et al., 1993; Das et al., 1999; Gozani et al., 1996; Will et al., 2002). The core 12S U2 snRNP particle binds to SF3b to form the pre-mature 15S U2 snRNP (Figure 2A) (Nesic and Krämer, 2001). Furthermore, the pre-mature 15S U2 snRNP binds to SF3a to form the functionally mature 17S U2 snRNP (Figure 2A) (Nesic and Krämer, 2001). Upon binding of the 17S U2 snRNP to the pre-mRNA, all of the SF3a components interact with a region ~25 nucleotides upstream of the branch site and are necessary for the successive spliceosomal assembly (Gozani et al., 1996). Thus, the SF3a complex is a crucial component for the formation of the functionally active 17S U2 snRNP.

The SF3a complex is composed of three proteins: SF3a60, SF3a66, and SF3a120 (Chiara et al., 1994; Krämer et al., 1994, 1995; Legrain and Chapon, 1993; Nesic and Krämer, 2001; Wiest et al., 1996). The SF3a60 and SF3a66 proteins each contain a zinc-finger domain (Figure 2B) (Nesic and Krämer, 2001). SF3a120 contains two tandem SURP domains (referred to as SURP1 and SURP2, hereafter) and a ubiquitin domain (Figure 2B) (Krämer et al., 1995). There are no other known motifs among the components of the SF3a complex. Far-western blotting analyses indicated that SF3a60 and SF3a66 interact with SF3a120, but not with each other (Nesic and Krämer, 2001). Namely, the N-terminal region (1–216) of SF3a66, excluding the zinc-finger domain, binds to a segment (residues 243–372) of SF3a120; none of the known domains mediates this interaction. On the other hand, a segment (residues 35–107) of SF3a60 binds to the SF3a120 SURP2. Therefore, SF3a120 is a pivotal molecule for the formation of the ternary complex.

The SURP domain, also known as the SWAP domain, was first found as two repeated sequence motifs in

*Correspondence: yokoyama@biochem.s.u-tokyo.ac.jp (S.Y.), ymuto@gsc.riken.jp (Y.M.)
[5] These authors contributed equally to this work.

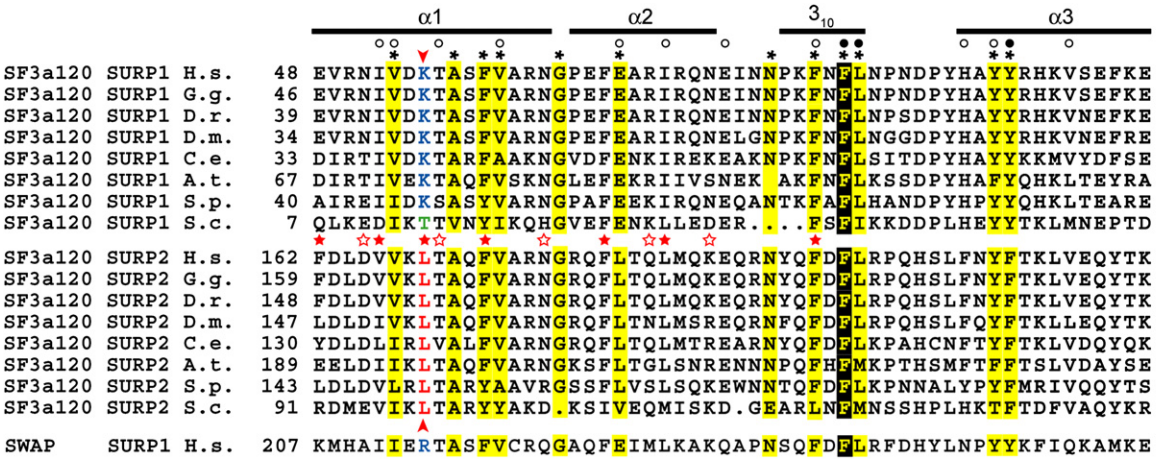

Figure 1. Sequence Alignment of SURP1 and SURP2 in SF3a120 and SURP1 in Human SWAP

Yellow backgrounds and asterisks indicate highly conserved residues among the SURP domains, and a black background with yellow letters indicates the conserved phenylalanine among all of the SURP domains, as shown in Figure 8B. The critical position for complex formation is indicated by a red arrowhead, and the colors reflect the amino acid type: red, blue, and green letters represent hydrophobic (protein-binding type), charged, and neutral (S, T, and Q) amino acids, respectively. The hydrophobic-core-forming residues are indicated by filled circles, and the residues that interact with the core residues are indicated by open circles. Red stars and red open stars indicate the hydrophobic- and hydrophilic-complex-forming residues in SURP2 of SF3a120, respectively.

a splicing regulator, Suppressor-of-White-Apricot from *Drosophila melanogaster* (Denhez and Lafyatis, 1994; Spikes et al., 1994). The exclusive appearance of SURP domains in splicing-associated proteins indicates that this motif emerged in concordance with the appearance of pre-mRNA splicing (Denhez and Lafyatis, 1994; Krämer et al., 1995; Sampson and Hewitt, 2003; Will et al., 2002). In addition, several recent lines of evidence indicated that the SURP-containing proteins play important roles in spliceosomal assembly (Nesic and Krämer, 2001; Sarkissian et al., 1996; Utans and Krämer, 1990; Will et al., 2002). Despite the functional importance of SURP-containing proteins, nothing is known about the properties of the SURP domains, except for the SF3a120 SURP2.

To elucidate the structural basis for the specific interaction between the SF3a120 SURP2 and SF3a60, we determined the solution structure of human SF3a120 SURP2 in complex with the fragment bearing residues 71–107 of SF3a60 by heteronuclear NMR spectroscopy. We also determined the solution structure of human SF3a120 SURP1 and compared the two SURP domains. Furthermore, we identified a single Leu residue (Leu169) of SURP2 that is responsible for the protein-binding activity, by structural and mutational analyses. As far as we know, this study is the first report of the SURP domain structure and its protein-recognition mode, as well as the first structural description of the subunit interactions in the SF3a complex.

Results

Identification of a Soluble Form of the SURP Domains from Human SF3a120

Based on the sequence alignment, we predicted that the two SURP domains of human SF3a120 are residues 48–110 (SURP1) and 158–217 (SURP2), respectively (Figures 1 and 2B). For structural studies on the two SURP domains of SF3a120, we constructed expression vectors for recombinant GST- or His-tagged SF3a120 fragments, spanning residues 48–110, 48–157, 48–217, 111–157, 111–217, and 158–217 (Figure 2C). Among them, the two SF3a120 fragments that contained SURP1 (48–110 and 48–157) were soluble. Therefore, we concluded that the SURP1 domain alone (48–110) (see Figure 2C) was suitable for the structural determination by NMR (see also Experimental Procedures).

On the other hand, two SF3a120 fragments (111–217 and 157–217) containing SURP2 were not expressed as His-tagged proteins in *E. coli* cells and were insoluble when expressed as GST-tagged proteins (Figure 2C). As described above, the region containing SURP2 in SF3a120 interacts with the SF3a60 fragment (35–107) (Nesic and Krämer, 2001). We speculated that direct binding of the SURP2 domain to a specific region within the SF3a60 fragment (35–107) is necessary for the soluble and stable expression. Similarly to SURP2, the SF3a60 fragment (1–107) alone was not soluble when expressed in *E. coli* cells (Figure 2D). Therefore, we established an *E. coli* coexpression system of the GST-tagged human SF3a60 (1–107) protein with several His-tagged SF3a120 fragments (48–217, 111–157, 111–217, and 158–217, respectively) (Figure 2E), to examine whether they formed a complex and remained soluble. From the four coexpression systems, two SF3a120 fragments (48–217 and 111–217) that contain SURP2 (158–217) with N-terminal extensions (48–157 or 111–157) were obtained as soluble forms in complex with the SF3a60 fragment (Figure 2E). Furthermore, we examined the complex formation activity for several lengths of SF3a120 and SF3a60 fragments and found that the combination of SF3a120 (134–217) and SF3a60 (71–107) was the minimum size complex of the fragments examined (Figure 2E) and that it was suitable for the structure analysis.

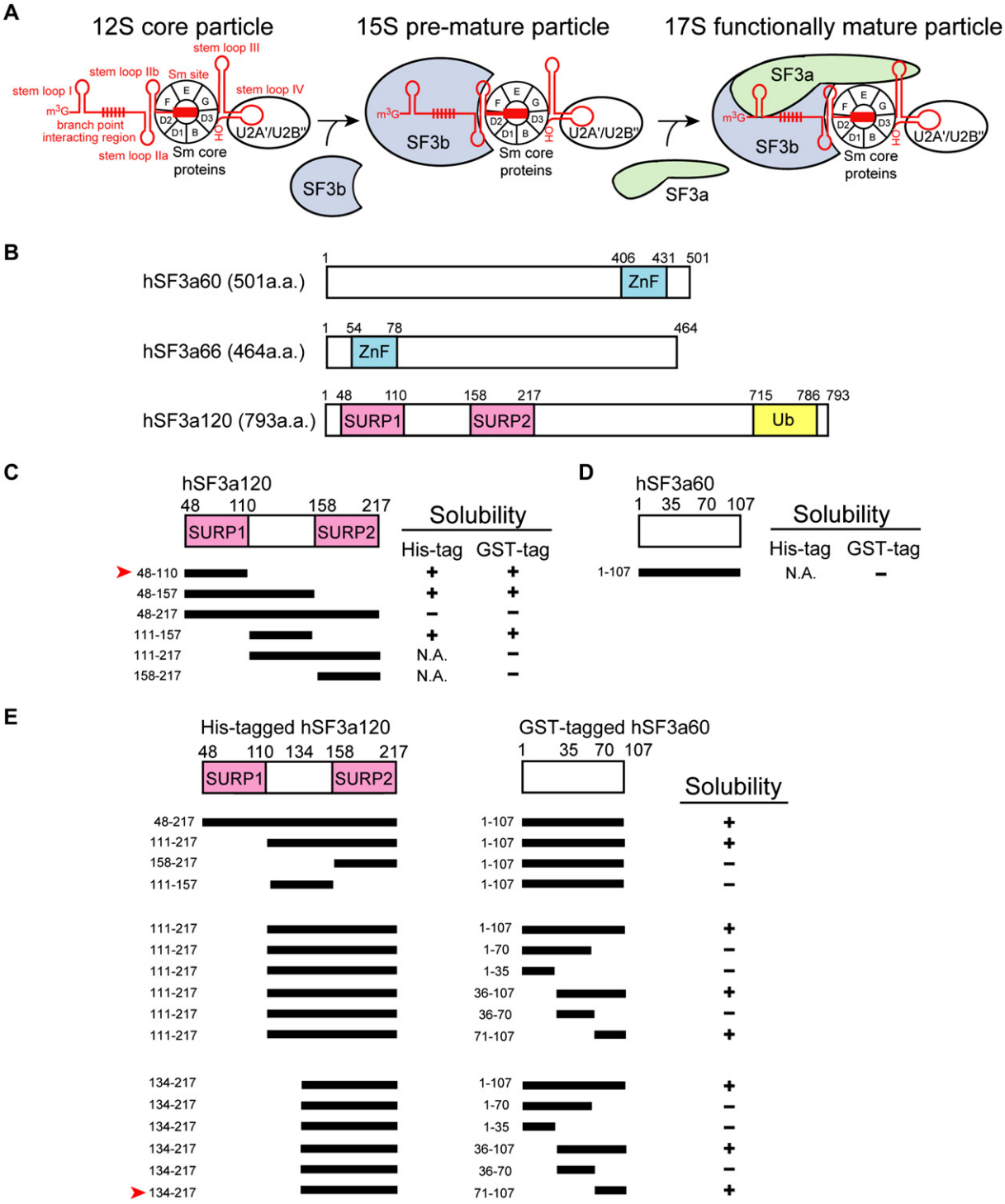

Figure 2. The cDNA Constructs Providing Soluble Recombinant Proteins

(A) Schematic diagram of the maturation process of U2 snRNP. Red lines indicate the U2 snRNA.

(B) Schematic diagrams of the domain structures of human SF3a60, SF3a66, and SF3a120. SURP, ubiquitin, and zinc-finger domains are represented with pink, yellow, and blue boxes, respectively.

(C) Solubility of each of the SF3a120 fragments expressed alone in *E. coli* cells. The red arrowhead indicates that this fragment was used for the solution structure analysis of SURP1. The solubility of the recombinant proteins is represented in (C), (D), and (E) as follows: plus and minus indicate soluble and insoluble proteins, respectively. N.A. indicates that the fragments were not expressed.

(D) Solubility of each of the SF3a60 fragments expressed alone in *E. coli* cells.

(E) Solubility of the His-tagged SF3a120 fragments coexpressed with various lengths of the GST-tagged SF3a60 fragments in *E. coli* cells. The red arrowhead indicates that this combination of fragments was used for the solution structure analysis of the SURP2-SF3a60 complex.

Table 1. Summary of Conformational Constraints and Structural Statistics for the 20 Energy-Refined Conformers of the SURP1 Domain and SURP2–SF3a60 Complex Structures

	SURP1	SURP2-SF3a60
NMR Distance and Dihedral Angle Constraints		
Distance constraints		
Total NOE	1867	2571
Intraresidue	421	696
Interresidue		
Sequential (|i−j| = 1)	518	658
Medium range (1 < |i−j| < 4)	545	672
Long range (|i−j| ≥ 5)	383	545
Intermolecular	—	203
Hydrogen bonds	0	6
ϕ/ψ dihedral angle constraints (TALOS[a])	79	118
CYANA target function (Å^2)	0.12 ± 0.02	0.17 ± 0.05
Structure Statistics		
Residual NOE violations		
Number > 0.10 Å	2	2
Maximum (Å)	0.11	0.11
Residual dihedral angle violations		
Number > 2.5°	0	0
Maximum (°)	0.51	1.36
AMBER energies[b] (kcal/mol)		
Total	−3117 ± 49	−4721 ± 85
van der Waals	−199 ± 7	−287 ± 16
Electrostatic	−3386 ± 48	−5329 ± 99
Ramachandran plot statistics[c,d] (%)		
Residues in most favored regions	96.2	95.4
Residues in additionally allowed regions	3.6	4.6
Residues in generously allowed regions	0.2	0.0
Residues in disallowed regions	0.0	0.0
Average rmsd from mean coordinates[c] (Å)		
Backbone	0.23	0.34
Heavy atoms	0.60	0.78

[a] Cornilescu et al. (1999).
[b] Cornell et al. (1995).
[c] For the structured regions comprising residues 48–103 of the SURP1 domain and residues 160–214 and 80p–96p of the SURP2-SF3a60 complex.
[d] Laskowski et al. (1996).

Solution Structure of the SURP1 Domain in SF3a120

Using standard multidimensional heteronuclear NMR spectroscopy, we assigned the main-chain and side-chain resonances (see Experimental Procedures); 95.3% (48–110) of the main chain and 98.4% (48–110) of the side chains were successfully assigned. We determined the solution structure of SURP1 on the basis of the 1867 ^{1}H-^{1}H distance constraints from nuclear Overhauser effect spectroscopy (NOESY) and 79 torsion angle restraints (Table 1). Out of the 100 independently calculated structures, the 20 lowest-energy conformers were used for further analyses.

The overall structure of the SURP1 domain adopts a compact fold, comprising three main α helices and one short 3_{10}-helix, with an α1-α2-3_{10}-α3 topology (Figure 3A). The three main α helices, α1, α2, and α3, are formed by residues Glu48-Asn63, Glu66-Asn74, and His91-Glu103, respectively (Figures 1 and 3B). The 3_{10} helix is formed by Pro79-Leu84 (Figures 1 and 3B). The α1 and α3 helices are located close together in a parallel manner and face the α2 helix across the 3_{10} helix in an antiparallel orientation (Figure 3B).

The 3_{10} helix serves as a scaffold for spatially positioning the α helices in the compact conformation, with the hydrophobic amino acids (Phe81, Phe83, and Leu84) playing important roles (Figure 4A): the Phe83 and Leu84 residues in the 3_{10} helix and the Tyr94 residue in the α3 helix form the hydrophobic core (Figure 4A). Residues in the α1 (Ile52, Thr56, and Val60), α2 (Glu68 and Ile71), α3 (His91 and Val98), and 3_{10} helices (Phe81) contact the core residues (Figure 4A). Furthermore, Phe81, in the 3_{10} helix, anchors Ile71 and Glu75 in the α2 helix. In addition, intimate contacts are observed in the α1 and α3 helices; namely, Tyr93 and Tyr94 of the α3 helix embrace Val53 of the α1 helix (Figure 4A). Collectively, these interactions among the helices reinforce their spatial relationships and maintain the compact folding of the SURP domain.

The Complex Structure of SURP2 and the Peptide from the SF3a60 Subunit

We solved the structure of the SF3a120 fragment (134–217) containing SURP2 complexed with SF3a60 (71–107) by heteronuclear multidimensional NMR (see Experimental Procedures). As a result, 96.5% (134–217) of the main chain and 97.6% (134–217) of the side chains of SF3a120 and 94.7% (71–107) of the main chain and 97.4% (71–107) of the side chains of SF3a60 were successfully assigned. From the complex, 2571 NOESY-based distance constraints and 118 torsion angle restraints were obtained (Table 1). In addition, 203 intermolecular distance constraints were included in the structure calculation. Out of the 100 independently calculated structures, the 20 lowest-energy conformers were further analyzed. Even though the N-terminal extension (134–157) of SURP2 is necessary to produce a soluble complex as described above, this region provided no distance constraints with the SF3a60 fragment and does not contribute to complex formation.

The SURP domain in complex with SF3a60 (71–107) adopts essentially the same tertiary structure as that of the SURP1 domain (Figures 3C and 3D). The α1, α2, α3, and 3_{10} helices are formed by Ala161-Asn177, Arg179-Lys188, Leu204-Tyr215, and Tyr193-Leu198, respectively (Figures 1 and 3D). The core-forming hydrophobic residues (Phe197 and Leu198 in the 3_{10} helix and Phe208 in the α3 helix) adopt a similar spatial relationship as that in SURP1 (Figure 4B). The peptide (71–107) from SF3a60 forms an amphipathic α helix (denoted as the αA helix hereafter), composed of Glu80p-Arg96p (the residue numbers followed by "p [peptide]" are those from the SF3a60 peptide), upon binding to the SURP2 domain (Figure 3D). The αA helix of SF3a60 mainly interacts with the α1 helix of SF3a120 and is located in an antiparallel orientation relative to the α1 helix of SURP2 (Figure 3D). The αA helix in SF3a60 also contacts the α2 and 3_{10} helices in SURP2.

The detailed interactions between the SURP2 domain and the SF3a60 fragment are as follows (Figure 5A). The hydrophobic amino acids of the SF3a60 fragment (Phe81p, Phe84p, Leu88p, and Ile91p) are aligned on helix A and interact with the SURP2 domain. The anchoring

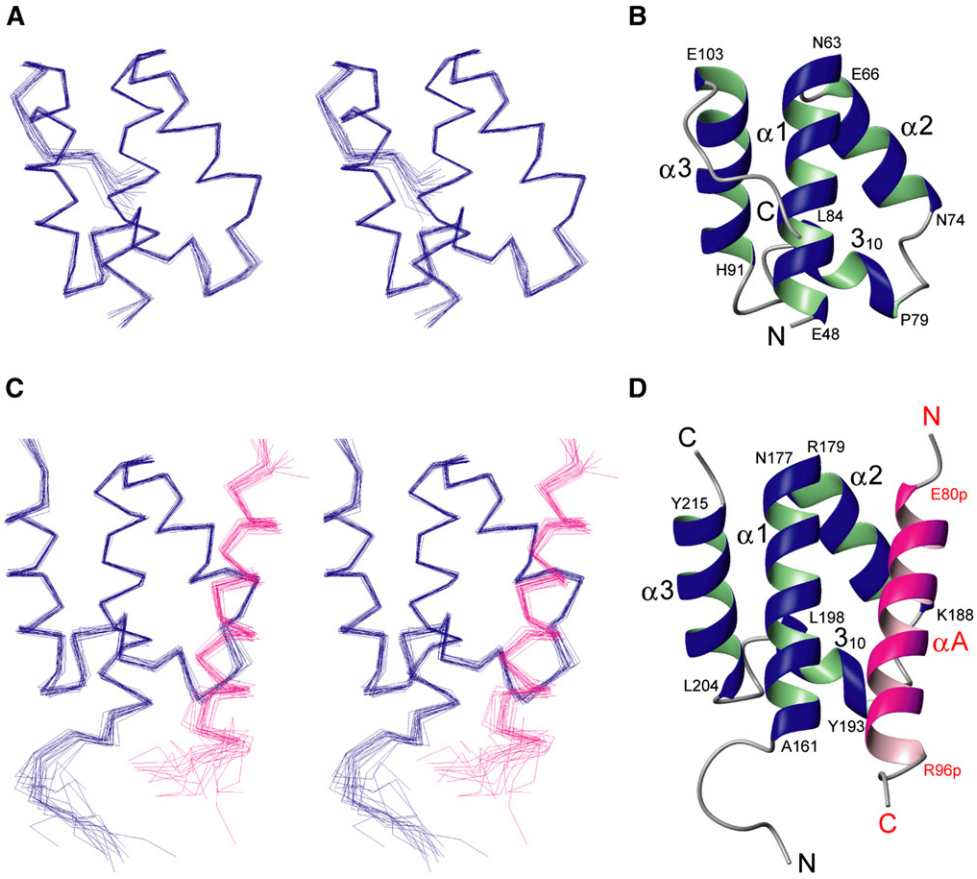

Figure 3. Overall Structures of the Human SF3a120 SURP1 Domain and the SURP2 Domain in Complex with the SF3a60 Fragment

(A) The superimposed, 20 lowest-energy conformers of the SURP1 domain (48–110). Blue lines represent Cα traces.

(B) Ribbon representation of SF3a120 SURP1.

(C) The superimposed, 20 lowest-energy conformers of the SURP2 domain (153–217) in complex with SF3a60 (78–100). The blue and pink lines represent Cα traces of SF3a120 SURP2 and SF3a60, respectively.

(D) Ribbon representation of SF3a120 SURP2 in complex with SF3a60. The blue and pink helices represent SF3a120 SURP2 and SF3a60, respectively.

point of the peptide is Phe81p, located near the turn of α1 and α2. The aromatic side chain of Phe81p fits well into the pocket composed of the side chains of Phe181, Gln184, Leu185, and Lys188, which are all located on the α2 helix. Val166, Leu169, Thr170, and Phe173 on α1, Phe181 and Leu185 on α2, and Phe195 on the 3_{10} helix form a shallow groove for the recognition of Phe84p and Leu88p of helix A. Furthermore, the side chain of Ile91p on helix A also fits well in the pocket composed of Phe162, Asp165, Val166, and Leu169 on α1. In summary, each hydrophobic amino acid on helix αA of the SF3a60 fragment intimately contacts a corresponding hydrophobic amino acid on SURP2, thus providing a zipper-like binding interaction (Phe81p-Phe181, Phe84p-Phe173, Leu88p-Leu169, and Ile91p-Phe162, respectively) (Figures 5A and 5B). Additionally, the hydrophilic residues of SF3a120 α1 (Asn177, Thr170, and Asp165) and SF3a60 αA (Gln80p, Gln83p, Tyr85p, Arg87p, Lys92p, and His95p) are involved in the intermolecular interactions of these two closely aligned α helices (Figures 5A and 5B).

A comparison of the location of the hydrophobic residues on the surface of SURP2 with that of SURP1 revealed a large hydrophobic patch on the surface of SURP2 (Figure 6B), while SURP1 has only a smaller hydrophobic area (Figure 6A). The network of these hydrophobic interactions fixes the SF3a60 αA helix tightly on the surface of SURP2.

Mutational Analyses Revealed the Crucial Residue of the SURP2 Domain for Complex Formation

SF3a60 bound to the surface of SURP2 composed of the α1, α2, and 3_{10} helices. Interestingly, as shown in Figure 1, the majority of the residues in SURP2 that interact with the SF3a60 fragment are conserved or semiconserved in SURP1 (Figures 1 and 5B), except for three hydrophilic residues, Asp165, Gln184, and Lys188, and two hydrophobic amino acid residues, Phe162 and Leu169. Phe162 and Leu169 have more intimate contacts with the SF3a60 peptide than Asp165, Gln184, and Lys188 (Figure 5A). Furthermore, the corresponding positions for Phe162 and Leu169 of SURP2 are occupied by amino acids with very different properties in SURP1 (charged amino acids: Glu48 and Lys55, respectively) (Figure 1), while the counterparts for Asp165, Gln184, and Lys188 of SURP2 have relatively similar properties in SURP1 (hydrophilic amino acids: Asn51, Arg70, and Asn74, respectively) (Figure 1). Importantly, Phe162

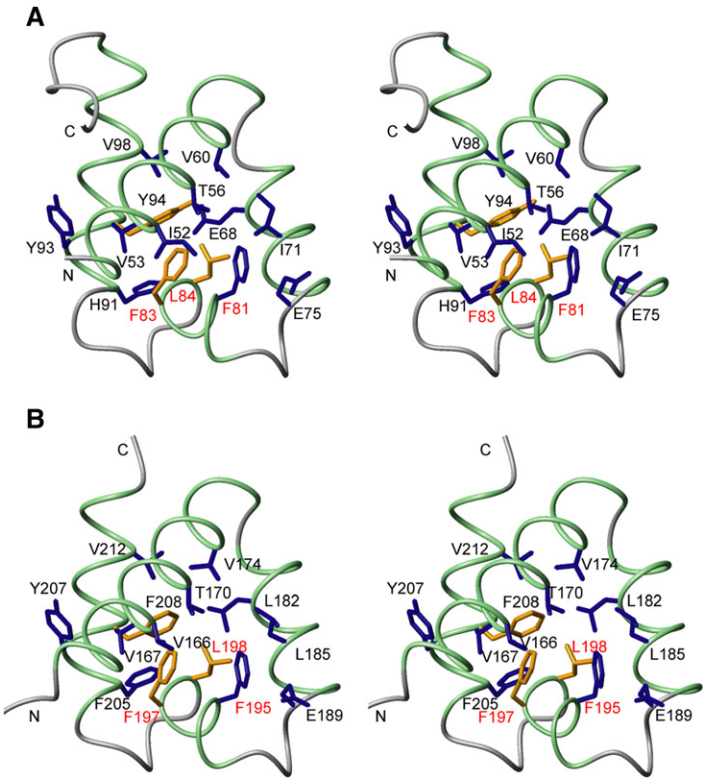

Figure 4. Hydrophobic-Core Formation of the SURP Domains of SF3a120

(A) Ribbon representation of the hydrophobic core of SF3a120 SURP1. The α helices are colored green. Side chains of the core-forming residues (Phe83, Leu84, and Tyr94) are represented in orange, and those of the core-associated residues are represented in blue. Red letters indicate residues on the 3_{10} helix.

(B) Hydrophobic core of SF3a120 SURP2. The representation code is the same as in (A).

and Leu169 of SURP2 are involved in the formation of a wider hydrophobic interface for SF3a60, as compared to the corresponding region in SURP1 (Figures 6A and 6B). Given the above results, these residues were further examined to identify their roles in domain solubility and complex formation.

We swapped these residues between SURP1 and SURP2 by site-directed mutagenesis. Glu48 and Lys55 of SURP1 were swapped with Phe and Leu, which are the corresponding residues in SURP2 (F162 and L169), respectively. The *E. coli* cells expressing the GST-tagged wild-type or mutant SURP1 domain were lysed by sonication, and the supernatant and pellet fractions were analyzed by 15%–25% polyacrylamide-gel electrophoresis (PAGE) (see Experimental Procedures). As shown in Figure 7A, when one of the charged residues in SURP1, Glu48, was replaced with phenylalanine (E48F mutant), the insoluble fraction of SURP1 increased. The mutation of another charged residue (Lys55; the K55L mutant) caused a severe deficiency in solubility, and the double mutation (E48F/K55L) made SURP1 completely insoluble. The *E. coli* cells were then transformed with plasmids encoding the His-tagged SF3a60 fragment (71–107) and the GST-tagged wild-type or mutant SURP1 domains. The cell lysates were analyzed by 15%–25% PAGE and were purified with Glutathione Sepharose (Figures 7C and 7E). As described above, the wild-type SURP1 was soluble (Figure 7C) and exhibited no affinity for the SF3a60 fragment (Figure 7E). The E48F mutant of SURP1 coexpressed with SF3a60 was predominantly soluble (Figure 7C) and also exhibited no affinity for the SF3a60 fragment, as in the case of the wild-type SURP1 (Figure 7E). Surprisingly, the K55L mutant of SURP1, which was insoluble when expressed alone in *E. coli* cells (Figure 7A), became soluble with the coexpression system (Figure 7C), and a significant amount of the stained band corresponding to the SF3a60 fragment was observed by the GST pull-down assay (Figure 7E). Similarly, the double mutant E48F/K55L showed SF3a60-binding ability (Figure 7E). These facts clearly indicate that the single K55L mutation provides SURP1 with the ability to form a complex with the SF3a60 fragment.

On the other hand, when both of the hydrophobic residues (Phe162 and Leu169) in the SURP2 domain were replaced with the corresponding charged residues of SURP1 (F162E/L169K), a portion of the expressed mutant SURP2 domain was soluble (Figure 7B), although both of the single mutants (F162E and L169K) had almost the same solubility as that of the wild-type under those conditions (Figure 7B). Furthermore, the soluble fraction of the L169K mutant of SURP2 lost its affinity for the SF3a60 fragment (71–107), while the F162E mutant of SURP2 showed normal complex-formation ability (Figure 7F). The L169K mutant of SURP2 was partially digested by endogenous bacterial proteases (Figures 7D and 7F), suggesting the importance of complex formation with SF3a60 for the stability of SURP2. Moreover, the double mutant (F162E/L169K), which was more soluble than the single L169K mutant, also showed no affinity for the SF3a60 fragment (71–107), similar to the L169K mutant (Figures 7D and 7F). These results indicate that Leu169, but not Phe162, is indispensable for SURP2 to form the complex with SF3a60.

Collectively, the mutational analyses performed on SURP1 and SURP2 clearly showed that the positions that correspond to Lys55 in SURP1 and Leu169 in SURP2 are critical for the complex formation. It is a remarkable fact that the exchange of just a single amino

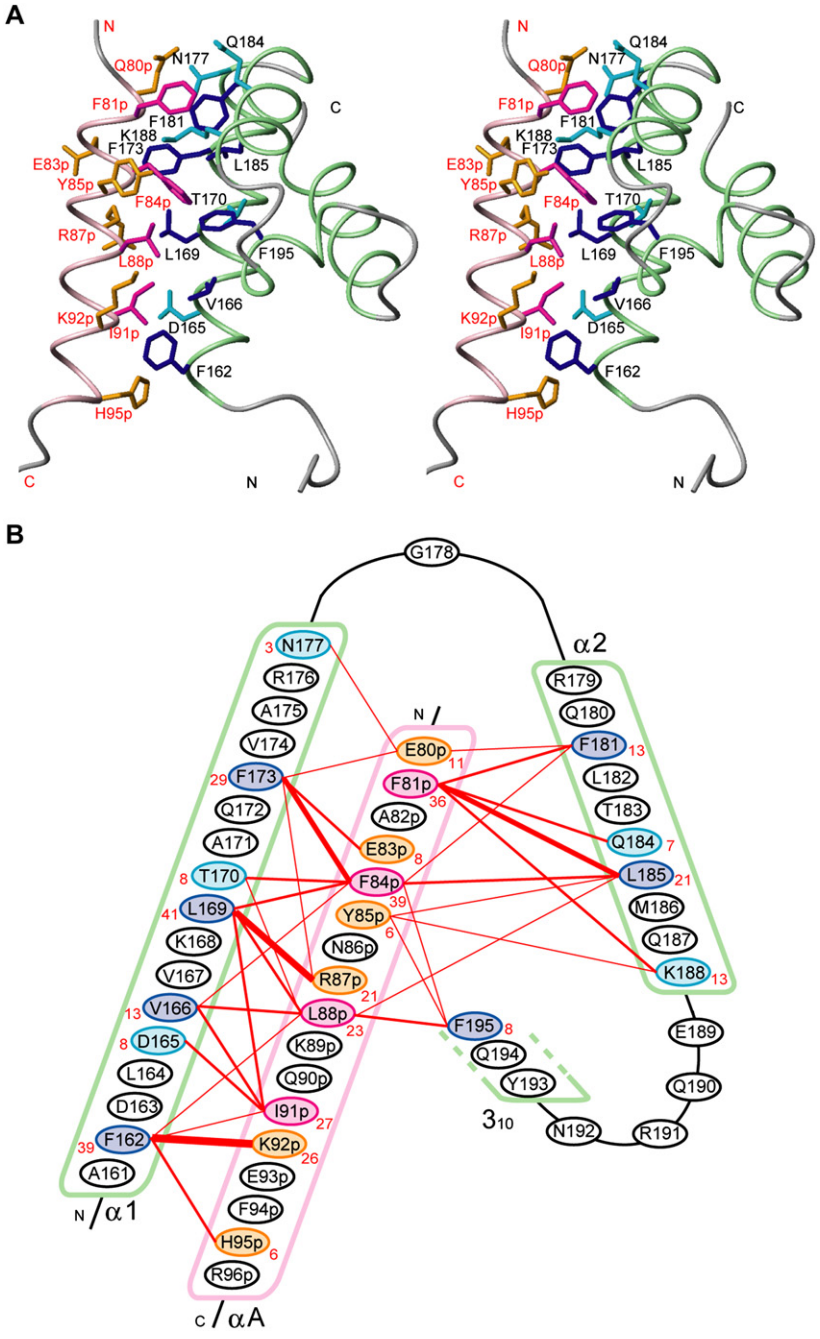

Figure 5. Complex Formation between SF3a120 SURP2 and SF3a60

(A) Ribbon representation of the complex forming interactions. The α helices of SURP2 and SF3a60 are colored green and pink, respectively. The side chains of the hydrophobic residues in SF3a120 and SF3a60 are represented in blue and magenta, respectively. The side chains of the non-hydrophobic residues in SF3a120 and SF3a60 are represented with cyan and orange sticks, respectively. Note that the hydrophobic residues in both of the proteins are positioned along the α helices, forming a zipper-like structure.

(B) Schematic diagram of the intermolecular NOEs between SF3a120 SURP2 and SF3a60. The numbers of intermolecular NOEs are indicated by the red numbers and are represented schematically by the line widths.

acid completely swapped the complex formation ability as well as the solubility between SURP1 and SURP2.

Discussion

The SF3a120 protein contains two tandem SURP domains. The SF3a120 SURP2, but not SURP1, is responsible for the assembly of the SF3a complex through binding to SF3a60. To reveal the structural basis for the specific interaction mediated by SURP2, we determined the solution structure of human SF3a120 SURP1 and that of SURP2 in complex with the SF3a60 fragment (71–107) by heteronuclear NMR spectroscopy. The SURP domain adopts a fold with an α1-α2-3_{10}-α3 topology. Searches in the Protein Data Bank (PDB) using the Dali and CE programs revealed no entries with structural similarity to

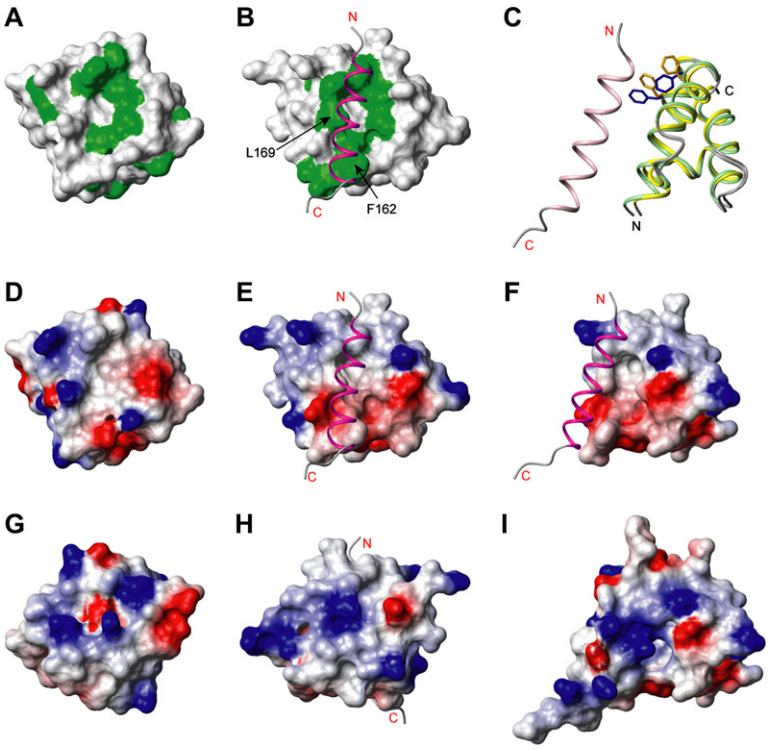

Figure 6. Surface Properties of the SF3a120 SURP1 and SURP2 Domains

(A and B) The locations of the hydrophobic residues on the surfaces of SURP1 (48–103) (A) and SURP2 (160–217) (B). The hydrophobic residues (Ala, Ile, Leu, Met, Phe, Pro, Trp, and Val) are colored green. The α helix of the SF3a60 fragment (78p–100p) is represented by a magenta ribbon.

(C) Comparison of the aromatic residues on the SF3a60-binding surface of the SF3a120 SURP2 domain. The yellow, green, and pink ribbons represent the main chains of SURP1 and SURP2 of SF3a120 and the SF3a60 fragment, respectively. The side chains of the two aromatic residues that undergo a conformational change between the SURP1 (Phe59 and Phe67) and SURP2 (Phe173 and Phe181) domains are shown in orange and blue.

(D–I) The electrostatic surface potentials of SURP1 (D and G) and SURP2 (E, F, H, and I). The SF3a60-binding side of SURP2 (E) and its counterpart in SURP1 (D), with the opposite sides shown in (H) and (G), respectively. The positive and negative charges are colored blue and red, respectively. The main chain of the SF3a60 peptide is represented by a magenta coil. (F and I) The electrostatic surface potentials of the SURP2 domain in complex with the SF3a60 peptide. The viewpoint is rotated by 45° along the y axis as compared to that of (E).

the SURP domain (Holm and Sander, 1995, 1996; Shindyalov and Bourne, 1998). Furthermore, the folding topology of the SURP domain is totally different from that of any other mRNA associated domain composed of only α helices, such as PWI, TPR and the Prp18 functional domain (Fukuhara et al., 2005; Jiang et al., 2000; Szymczyna et al., 2003). A precise comparison between SURP1 and SURP2 revealed the peculiarity of SURP2. The SURP2 holds SF3a60 tightly on its hydrophobic surface. A single Leu residue (Leu169) of SURP2 is the critical determinant for complex formation. These interactions fix the spatial relationship between the SF3a120 and SF3a60 subunits and may contribute to the assembly of the SF3a complex.

To gain further insight into the SURP domains, first, we searched for SURP-containing proteins in the SMART database (Letunic et al., 2006) and identified six proteins in the human genome, SF3a120, SR140, SWAP, CHERP, SF4, and SFRS14, which are all splicing-associated proteins (Figure 8A). In addition, as listed in Figure 8A, we searched for the corresponding ortholog of each protein in other species for which all of the genomic information is available. Figure 8A indicates that the number of (putative) SURP-containing proteins in the genomes increases according to the evolutional process; especially, it varies greatly among animals. Among the SURP-containing proteins, only SF3a120 is found in all species. SF3a120 is a constitutive splicing factor, in contrast to the other proteins, which are regulatory factors. These results suggest that SF3a120 is the most ancestral protein and that the SURP domains of SF3a120 diversified during evolution, perhaps along with the increased complexity of gene regulatory systems, including splicing. The consensus residues among all of the SURP domains listed in Figure 8A were analyzed with profile hidden Markov models (Schuster-Bockler et al., 2004), and 12 highly conserved residues were found among 56 amino acid residues (Figure 8B). These consensus residues are conserved in both of the SURP domains in SF3a120, with a few exceptions (Figure 1). They play an important role in the formation of the ternary structure of the SURP domain (Figure 1).

The differences in the properties of the two SURP domains from SF3a120 are considerable: SURP1 is soluble and has no protein binding activity, while SURP2 is insoluble by itself and is solubilized upon binding to the SF3a60 peptide. However, unexpectedly, a comparison between the structures of SURP1 and the SURP2-SF3a60 complex revealed that the composition of amino acids constituting the protein-binding surface of SURP2 was quite similar to the counterpart in SURP1, with few exceptions (Figures 1, 6A, and 6B). Strikingly, the protein-binding activity for SF3a60 could be swapped completely between the two SURP domains by the substitution of a single key amino acid residue (Lys55 in SURP1 and Leu169 in SURP2). Detailed comparisons between the features of the SF3a60-binding site on SURP2 and those of the corresponding region in SURP1 revealed differences in the conformations of the conserved aromatic residues (Figure 6C), in addition to the change of the key amino acid. While the main-chain structures are essentially identical between the two SURP domains, the side chains of two aromatic residues in SURP2 (Phe173 and Phe181) are properly oriented for the formation of the hydrophobic patch that captures the SF3a60 peptide. These two aromatic amino acids are conserved in SURP1. However, their conformations are quite different from those in SURP2 (Figure 6C): the χ1 angles of Phe173

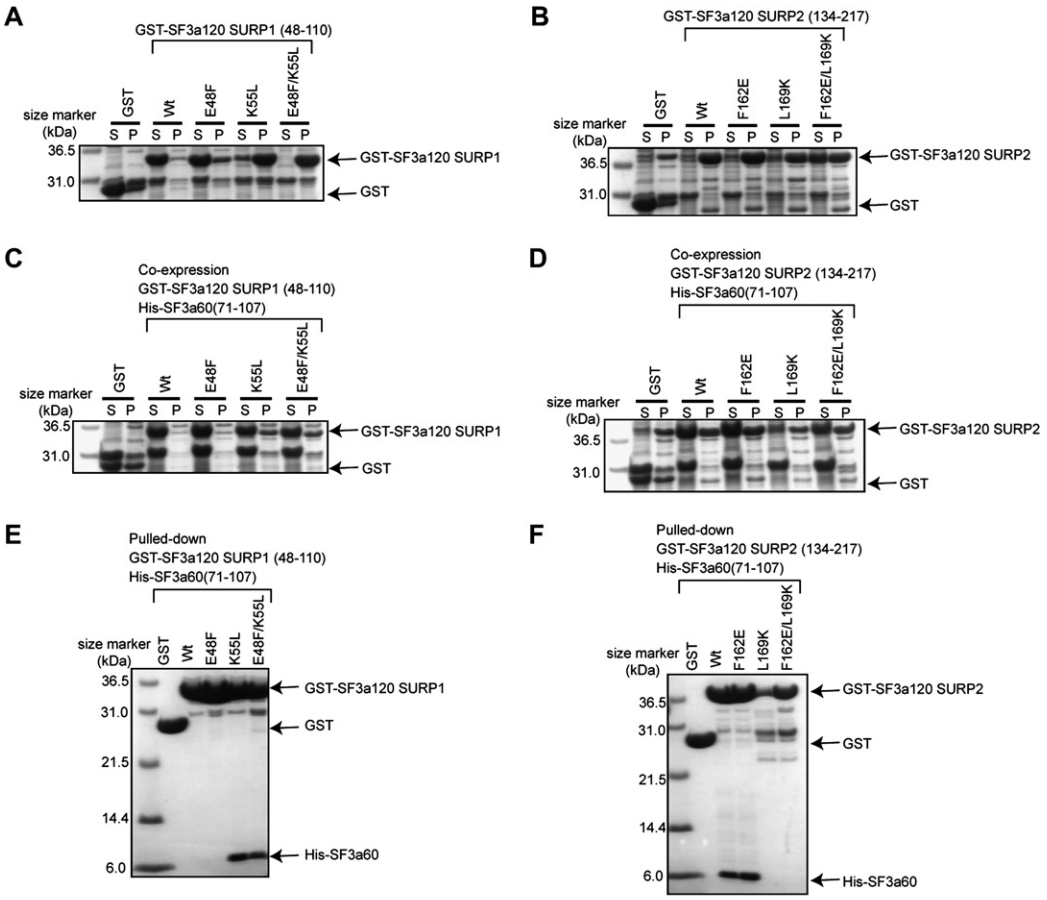

Figure 7. A Single Amino-Acid Replacement in the SF3a60-Binding Surface Swaps the Properties of the SURP Domains

(A) Effect on the solubility of the SURP1 domain by the mutations of Glu48 and Lys55. The crude extracts of the *E. coli* cells that expressed the indicated mutant or wild-type SURP1 domains were analyzed by SDS-PAGE and stained with Coomassie brilliant blue dye. The soluble (S) and precipitated (P) fractions are shown side by side.

(B) Effect on the solubility of the SURP2 domain by the mutations of Phe162 and Leu169. Note that the *E. coli* cells expressing the SURP1 and SURP2 domains alone were analyzed in (A) and (B), respectively.

(C) Effect on the solubility of the SURP1 domain in the presence of the SF3a60 peptide by the mutations of Glu48 and Lys55. The crude extracts of the *E. coli* cells that coexpressed the wild-type or mutant SURP1 domain and the wild-type SF3a60 peptide were analyzed by SDS-PAGE.

(D) Effect on the solubility of the SURP2 domain in the presence of the SF3a60 peptide by the mutations of Phe162 and Leu169.

(E) GST pull-down assay of SURP1 and SF3a60. The cell lysates of *E. coli* coexpressing the GST-tagged wild-type/mutant SF3a120 SURP1 and the His-tagged SF3a60 were mixed with Glutathione Sepharose, and the protein fractions bound to the resin were analyzed by SDS-PAGE.

(F) GST pull-down assay of SURP2 and SF3a60. The cell lysates of *E. coli* coexpressing the GST-tagged wild-type/mutant SF3a120 SURP2 and the His-tagged SF3a60 were mixed with Glutathione Sepharose, and the protein fractions bound to the resin were analyzed by SDS-PAGE.

and Phe181 in SURP2 are approximately 180° and −60°, respectively, while those of Phe59 and Phe67 in SURP1 are approximately −60° and 180°, respectively. The SURP1 domain of SF3a120 also has the potential to bind with the SF3a60 peptide, as a conformational change of the conserved aromatic amino acid residues could generate the binding surface. Our mutational experiments clearly showed that the SURP1 domain gained the protein-binding function through the substitution of just one crucial residue. There has been no report about the protein-protein interactions mediated by the SURP1 domain of SF3a120. However, it is probable that the Lys55 residue of the SURP1 domain contacts other molecules, like proteins or nucleic acids, which have aromatic and acidic properties on the surface, thus forming the cation-π and electrostatic interactions, respectively.

The difference in the properties of the two SURP domains from SF3a120 may reflect their molecular evolution. We performed a phylogenic analysis and found that the SURP domains are divided into two major subgroups (termed subgroups 1 and 2 hereafter) (Figure 8C). The sequence identities among the human SURP domains within subgroups 1 and 2 are 38% ± 2% and 36%, respectively, whereas that between the two subgroups is 24% ± 6%. The SURP1 and SURP2 domains from SF3a120 belong to subgroups 1 and 2, respectively. Intriguingly, the SURP domains that appear only in animals (CHERP, SF4, and SFRS14) and the SURP domain of SR140 are included in subgroup 1 (Figure 8C). It seems reasonable to

A

Protein name	SF3a120	SR140	SWAP	CHERP	SF4	SFRS14
Number of SURP domains	2	1	2	1	2	2
Organism species	Accession No. (Uniprot sequence identifier or Ensemble accession number)					
Homo sapiens	Q15459	O15042	Q12872	Q8IWX8	Q8IWZ8	Q8IX01
Gallus gallus	Q5ZM84	UPI0000448A93	UPI0000449208	UPI00003ACA9E	UPI0000449A7D	UPI0000449A50
Danio rerio	Q90X41	UPI0000545688	Q7ZUT4*	Q8AX67	ENSDARP00000064894	-
Drosophila melanogaster	Q9VEP9	Q7K0X5	P12297	Q8I9J8	-	-
Caenorhabditis elegans	Q10577	Q95YF1	Q10580	-	-	-
Arabidopsis thaliana	Q8RXF1	Q9C5J3	Q9FLQ1	-	-	-
Schizosaccharomyces pombe	O13900	-	-	-	-	-
Saccharomyces cerevisiae	P32524	-	-	-	-	-

*:SWAP of *Danio rerio* does not contain SURP2.

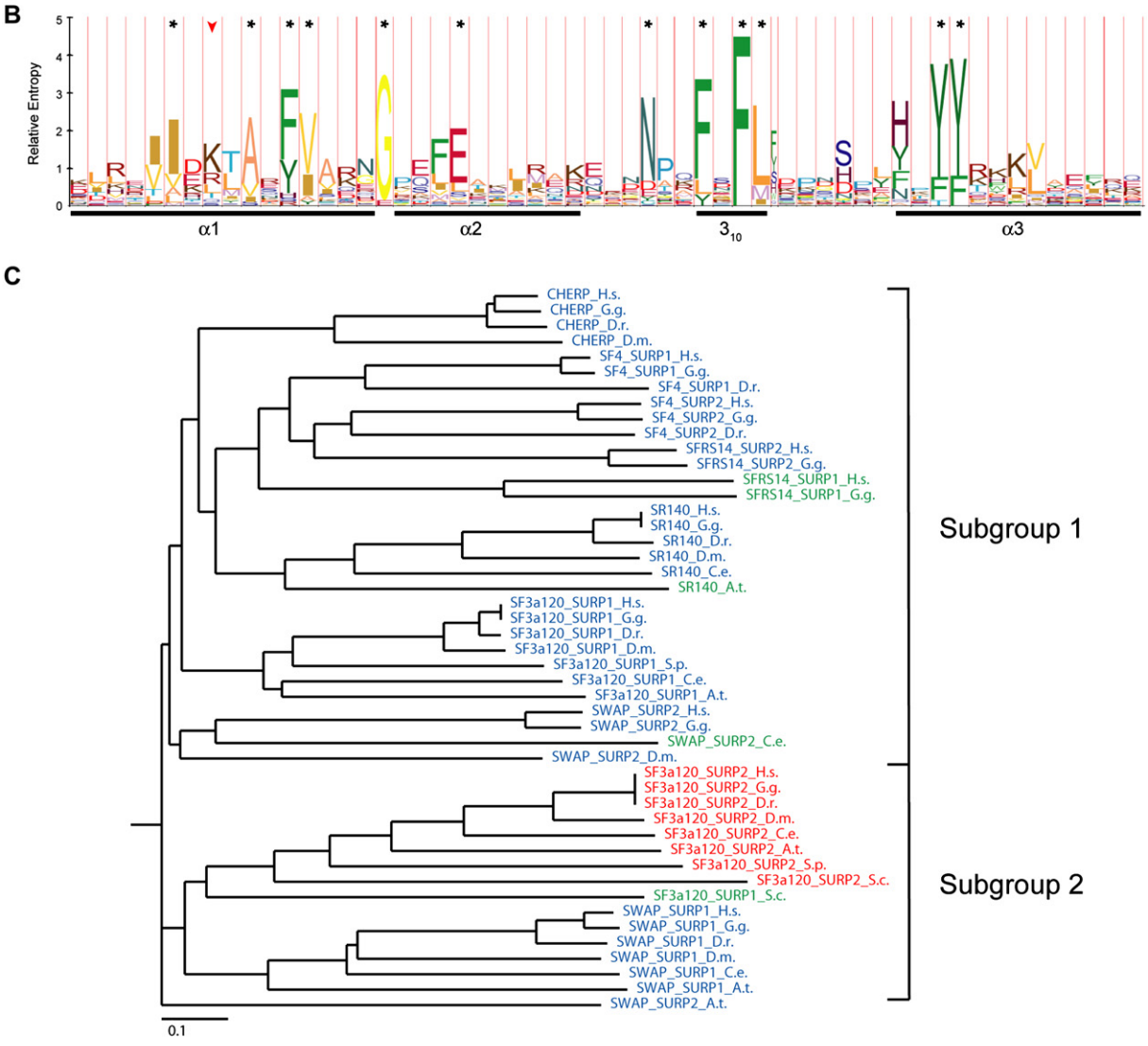

Figure 8. Sequence Profiling Analyses of SURP Domains

(A) SURP-containing proteins in eukaryotes.

(B) HMM-logo format representation of the amino acid conservation in the SURP domains. The sequence information of all of the SURP domains shown in (A) was analyzed by the HMMER program (Schuster-Bockler et al., 2004). Asterisks indicate highly conserved residues among the SURP domains. The red arrowhead indicates the critical position for the complex formation.

(C) Phylogenic tree of the SURP domains. All of the SURP domains shown in (A) are included. The colors of the letters reflect the amino acid type of the critical residues for the complex formation (Leu169 in the SURP2 domain on human SF3a120); red, blue, and green letters represent hydrophobic (protein-binding type), charged, and neutral (S, T, and Q) amino acids in the positions, respectively.

speculate that these SURP domains were generated from SURP1 of SF3a120 during evolution. Actually, the key amino-acid residues (corresponding to Lys55 in SURP1 and Leu169 in SURP2) on α1 of these SURP domains are hydrophilic: in many cases the key position is occupied by a positively charged amino acid and even in a few exceptions, a Gln, Ser, or Thr residue is located there. These hydrophilic residues probably enhance the solubility and stability of the SURP domains in subgroup 1, and as a result, the SURP domains with these residues diversified in the genomes during animal evolution. The SURP1 domain of SWAP is closer to SURP2 of

SF3a120, as compared to SURP1 of SF3a120, in the phylogenic tree (Figure 8C). However, it also has a charged amino acid residue (Arg214) at the critical position (Figure 1). This suggests that the SURP1 domain of SWAP may resemble the L169R mutant of SURP2 in SF3a120 and that it has intermediate properties between those of SURP1 and SURP2 from SF3a120. Consequently, the critical Leu residue is restricted to the SF3a120 SURP2. Consistent with this fact, only the SURP2 domain plays an important role in the SF3a complex assembly.

The SURP-containing proteins are found exclusively in splicing factors. This provides a hint about their common functions: they could serve as RNA-binding domains. A UV cross-linking analysis revealed that SF3a120 binds to the upstream region of the branch point of pre-mRNA, but not to U2 snRNA (Gozani et al., 1996). Moreover, one of the sterile α motif (SAM) domains (Ponting, 1995), which had originally been described as a protein-binding domain, has a positively charged patch on its surface and binds to a stem-loop RNA (Oberstrass et al., 2006). As shown in Figures 6D–6I, the SURP1 and SURP2 domains of SF3a120 have positive charges on their surfaces. Furthermore, a new positively charged patch is formed in the groove between SURP2 and helix A of SF3a60 (Figures 6F and 6I). These findings raise the possibility that although the SURP2 domain has a peptide-binding function, the SURP domains of SF3a120 may also have an RNA-binding activity through the positively charged patch. Therefore, we examined whether the SF3a120 fragment (48–217), containing both of the SURP domains complexed with the SF3a60 peptide (71–107), binds to the RNA 12–25 nucleotides upstream from the branch point of the pre-mRNA, with the sequence GGG UUU CCU UGA AG, by using an ultrasensitive isothermal titration calorimeter (VP-ITC, Microcal). However, the protein complex showed no binding to the RNA (data not shown). It is probable that the SURP domains in SF3a120 bind to structured RNA, like the SAM domain.

In the present study, we revealed the structural basis for the specific interaction between SF3a60 and SF3a120 mediated by the SURP domain. To gain further insight into the molecular assembly of the SF3a complex and the U2 snRNP in the spliceosome, more structural information is necessary. Work is now in progress to analyze the SF3a-RNA interaction, which will reveal the common functions of the SURP domains.

Experimental Procedures

Expression of Recombinant Proteins

The cDNAs encoding the fragments of human SF3a60 (Accession number: Q12874) (1–107, 1–70, 1–35, 36–107, 36–70, and 71–107) and the cDNAs encoding the fragments of human SF3a120 (Accession number: Q15459) (48–217, 48–157, 48–110, 111–157, 111–217, 134–217, and 158–217) were cloned in the plasmid vector pGEX6P-1 (GE Healthcare) and/or the plasmid vector pET-15b (Novagen). In both constructs, the TEV protease cleavage site (Glu-Asn-Leu-Tyr-Phe-Gln-Gly, encoded by gaaaacctgtatttccagggc) was placed between the tag and protein sequences. The TEV protease cleaves the amide bond of the Gln-Gly dipeptide, resulting in an additional Gly residue at the N terminus of each construct.

For the coexpression system of SF3a60 and SF3a120, the cDNA encoding each SF3a120 fragment was cloned into the pGEX6p-1 plasmid, with the TEV cleavage site between GST and SF3a120. The cDNA encoding each His-tagged SF3a60 fragment, with the TEV cleavage site between the His-tag and SF3a60, was cloned into the downstream region of the aforementioned pGEX6p-1-SF3a120 plasmid. The ribosomal binding site (RBS: gaaggag) was placed between the stop codon of SF3a120 and the start codon of the downstream SF3a60 cDNA. In the same way, the coexpression vectors for the GST-tagged SF3a60 and the His-tagged SF3a120 were constructed.

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. IPTG was added to the culture to a final concentration of 1 mM, to induce protein expression. After 4–5 hr of cultivation, the cells were harvested and were lysed by sonication in a phosphate buffer containing 1 mM DTT, 1 mM PMSF, and protease inhibitor cocktail for general use (Nacalai Tesque). The lysates were centrifuged at 17,000 × g for 10 min at 4°C to separate the supernatant and pellet. The protein solubility was assessed by SDS polyacrylamide gel electrophoresis analyses of the lysates.

Preparation of the SF3a120 SURP1 Domain for Structure Determination

To obtain the ^{15}N,^{13}C-labeled recombinant His-tagged SURP1 domain of SF3a120, *E. coli* strain BL21 (DE3) cells, transformed with the pET15b-SF3a120 SURP1 (48–110), were cultivated at 37°C as described above, except for the use of modified minimum medium (15 g/l Na_2HPO_4, 6 g/l KH_2PO_4, 5 g/l NaCl, 1 g/l ^{15}N-NH_4Cl, 4 g/l ^{13}C-glucose, 0.12 g/l $MgSO_4$, 11.88 mg/l $FeSO_4$, 2 mg/l thiamine hydrochloride, 2 mg/l D-biotin, 2.65 mg/l $CaCl_2$, 57.15 mg/l citric acid) supplemented with 50 mg/l ampicillin. The harvested culture was lysed as described above. The lysate was applied to a Ni-NTA SuperFlow (Qiagen) column with an imidazole gradient from 20 mM to 250 mM, and the tag was removed by an incubation with TEV protease overnight at room temperature. The tag-free SURP1 was further purified by RESOURCE S column chromatography (GE Healthcare).

Preparation of the SF3a60-SF3a120 SURP2 Domain Complex for Structure Determination

For the structure determination of the SF3a120 SURP2-SF3a60 complex, we prepared the ^{15}N,^{13}C -labeled recombinant His-tagged SF3a120 SURP2 (134–217) and the GST-tagged SF3a60 (71–107) proteins by using the coexpression system. *E. coli* strain BL21 (DE3) cells, transformed with the coexpression plasmid vector pGEX6p-1-SF3a60 (71–107)-His(6)-SF3a120 SURP2 (134–217), were cultivated, harvested, and lysed in the same way as in the preparation of the ^{15}N,^{13}C-labeled SURP1 domain. The lysate was applied to a Glutathione Sepharose 4 Fast Flow (GE Healthcare) column and was eluted by 10 mM reduced glutathione. The eluate was further purified by chromatography on Ni-NTA SuperFlow with an imidazole gradient from 20 mM to 250 mM, and the His and GST tags were removed by an incubation with TEV-protease overnight at room temperature. The tag-free complex was further purified by RESOURCE S column chromatography.

NMR Spectroscopy

The protein samples for NMR measurements were concentrated to approximately 2.0 mM, in 20 mM sodium phosphate buffer (pH 6.0 and 7.0) containing 1 mM 1,4-*DL*-dithiothreitol-d_{10} (*d*-DTT) dissolved in 90% ^1H_2O/10% ^2H_2O, by using a Vivaspin 20 ml concentrator (membrane: 3000 MWCO PES, Vivascience).

NMR experiments were performed at 5°C for the SURP1 domain and 35°C for the SURP2 domain on Bruker 700 MHz and 800 MHz spectrometers (Bruker AV700 and Bruker AV 800), respectively. ^1H, ^{15}N, and ^{13}C chemical shifts were referenced relative to the frequency of the ^2H lock resonance of water. Main-chain and side-chain assignments were obtained from a combination of standard triple resonance experiments (Bax and Saxena, 1994; Kay, 1997). 2D [^1H, ^{15}N]-HSQC and 3D [^1H, ^{15}N, ^{13}C]-HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH spectra were used for the ^1H, ^{15}N, and ^{13}C assignments of the main chain. Side-chain ^1H and ^{13}C assignments of the nonaromatic residues were obtained with 2D [^1H, ^{13}C]-HSQC, 3D HBHA(CO)NH, H(CCCO)NH, (H)CC(CO)NH, HCCH-COSY, HCCH-TOCSY, (H)CCH-TOCSY spectra. Assignments were confirmed with the 3D ^{15}N-edited [^1H, ^1H]-NOESY and ^{13}C-edited [^1H, ^1H]-NOESY spectra. The ^1H and ^{13}C spin systems of the aromatic rings of Phe, Trp, His, and Tyr were identified by

3D HCCH-COSY and HCCH-TOCSY experiments. Then, 3D ^{13}C-edited [^{1}H, ^{1}H]-NOESY was used for the establishment of the sequence-specific resonance assignments of aromatic side chains. All NOESY spectra were recorded with mixing times of 80–150 ms. The 2D and 3D spectra were processed by using NMRPipe (Delaglio et al., 1995). Analyses of the processed data were performed with the programs NMRView (Johnson, 2004) and KUJIRA (N. Kobayashi, personal communication).

Structure Calculations

The three-dimensional structures of the protein and the protein-peptide complex were determined by combined automated NOESY cross-peak assignment (Herrmann et al., 2002) and structure calculations with torsion-angle dynamics (Güntert et al., 1997) implemented in the program CYANA 2.1 (Güntert, 2004). Dihedral angle constraints for ϕ and ψ were obtained from the main chain and $^{13}C^{\beta}$ chemical-shift values with the program TALOS (Cornilescu et al., 1999) and by analyzing the NOESY spectra. Structure calculations started from 100 randomized conformers and used the standard CYANA simulated annealing schedule (Güntert et al., 1997). 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 (Koradi et al., 2000) by using the AMBER force field (Cornell et al., 1995). PROCHECK-NMR (Laskowski et al., 1996) and MOLMOL (Koradi et al., 1996) were used to validate and to visualize the final structures, respectively.

Mutational Analysis

Point mutations were introduced into GST-tagged SURP1 (48–110) and SURP2 (134–217) from SF3a120 by PCR, with 28- to 30-mer primers spanning the site of the desired mutation, as described (Ito et al., 1991). Mutations were confirmed by sequencing. The solubility of the mutant proteins was assessed as described above. To examine the complex-formation ability of these mutant proteins, His-tagged SF3a60 (71–107) and the GST-tagged SF3a120 mutants were coexpressed in *E. coli* cells. A 200 μl volume of the cell lysate was mixed with a 5 μl volume of the Glutathione Sepharose resin for 30 min at 4°C, and then the resin was washed five times with PBS containing 0.05% Nonidet P-40 (Nacalai Tesque). The proteins bound to the resin were analyzed by SDS-PAGE.

Informatics Analyses to Search for SURP Domains

We identified nine SURP-containing proteins in the human genome, accessed in the SMART database (Letunic et al., 2006), and removed the redundancy of the entries based on the cut-off value = 100% homology. As a result, six proteins were identified as SURP-containing proteins. The orthologs of the human SURP-containing proteins were obtained from the entries in the SMART database, as described above. Their amino acid sequences were aligned and used to create a phylogenic tree with the ClustalX software (Jeanmougin et al., 1998). The consensus residues among all of the SURP domains were analyzed with the profile hidden Markov models (Schuster-Bockler et al., 2004).

Acknowledgments

We are grateful to Dr. N. Kobayashi for help with the NMR data analysis and structure calculations. We would like to thank Mr. T. Tomizawa for help with the VP-ITC operation and Mr. N. Tochio for software implementation. We would like to acknowledge Dr. K. Kurimoto for discussions and critical reading of the manuscript. We would like to thank Ms. A. Ishii, Ms. K. Yajima, and Ms. T. Nakayama for help with the preparation of the manuscript. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI) of the National Project on Protein Structural and Functional Analyses, the Ministry of Education, Culture, Sports, Science and Technology of Japan. This work was also supported by a grant from the Human Frontier Science Program (H.F.S.P.) to the Muto research group.

Received: August 10, 2006
Revised: September 4, 2006
Accepted: September 9, 2006
Published: November 14, 2006

References

Bax, W.A., and Saxena, P.R. (1994). The current endothelin receptor classification: time for reconsideration? Trends Pharmacol. Sci. *15*, 379–386.

Brosi, R., Groning, K., Behrens, S.E., Lührmann, R., and Krämer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science *262*, 102–105.

Caspary, F., and Séraphin, B. (1998). The yeast U2A′/U2B″ complex is required for pre-spliceosome formation. EMBO J. *17*, 6348–6358.

Chiara, M.D., Champion-Arnaud, P., Buvoli, M., Nadal-Ginard, B., and Reed, R. (1994). Specific protein-protein interactions between the essential mammalian spliceosome-associated proteins SAP 61 and SAP 114. Proc. Natl. Acad. Sci. USA *91*, 6403–6407.

Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Cajdwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Das, B.K., Xia, L., Palandjian, L., Gozani, O., Chyung, Y., and Reed, R. (1999). Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol. Cell. Biol. *19*, 6796–6802.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Denhez, F., and Lafyatis, R. (1994). Conservation of regulated alternative splicing and identification of functional domains in vertebrate homologs to the *Drosophila* splicing regulator, suppressor-of-white-apricot. J. Biol. Chem. *269*, 16170–16179.

Dziembowski, A., Ventura, A.P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprevote, O., and Séraphin, B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO J. *23*, 4847–4856.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol. Cell *17*, 537–547.

Gozani, O., Feld, R., and Reed, R. (1996). 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. *10*, 233–243.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. *278*, 353–378.

Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. *273*, 283–298.

Hastings, M.L., and Krainer, A.R. (2001). Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. *13*, 302–309.

Herrmann, T., Güntert, P., and Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. *319*, 209–227.

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.

Holm, L., and Sander, C. (1996). The FSSP database: fold classification based on structure-structure alignment of proteins. Nucleic Acids Res. *24*, 206–209.

Ito, W., Ishiguro, H., and Kurosawa, Y. (1991). A general method for introducing a series of mutations into cloned DNA using the polymerase chain reaction. Gene *102*, 67–70.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. Trends Biochem. Sci. *23*, 403–405.

Jiang, J., Horowitz, D.S., and Xu, R.M. (2000). Crystal structure of the functional domain of the splicing factor Prp18. Proc. Natl. Acad. Sci. USA *97*, 3022–3027.

Johnson, B.A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol. Biol. *278*, 313–352.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol. Cell *12*, 5–14.

Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Lührmann, R., Li, J., and Nagai, K. (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. Cell *96*, 375–387.

Kay, L.E. (1997). NMR methods for the study of protein structure and dynamics. Biochem. Cell Biol. *75*, 1–15.

Koradi, R., Billeter, M., and Güntert, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. Comput. Phys. Commun. *124*, 139–147.

Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. *14*, 51–55, 29–32.

Krämer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. Annu. Rev. Biochem. *65*, 367–409.

Krämer, A., Legrain, P., Mulhauser, F., Groning, K., Brosi, R., and Bilbe, G. (1994). Splicing factor SF3a60 is the mammalian homologue of PRP9 of *S. cerevisiae*: the conserved zinc finger-like motif is functionally exchangeable in vivo. Nucleic Acids Res. *22*, 5223–5228.

Krämer, A., Mulhauser, F., Wersig, C., Groning, K., and Bilbe, G. (1995). Mammalian splicing factor SF3a120 represents a new member of the SURP family of proteins and is homologous to the essential splicing factor PRP21p of *Saccharomyces cerevisiae*. RNA *1*, 260–272.

Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*, 477–486.

Legrain, P., and Chapon, C. (1993). Interaction between PRP11 and SPP91 yeast splicing factors and characterization of a PRP9-PRP11-SPP91 complex. Science *262*, 108–110.

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

Lührmann, R., Kastner, B., and Bach, M. (1990). Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. Biochim. Biophys. Acta *1087*, 265–292.

Michaud, S., and Reed, R. (1991). An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. Genes Dev. *5*, 2534–2546.

Nesic, D., and Krämer, A. (2001). Domains in human splicing factors SF3a60 and SF3a66 required for binding to SF3a120, assembly of the 17S U2 snRNP, and prespliceosome formation. Mol. Cell. Biol. *21*, 6406–6417.

Nilsen, T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? Bioessays *25*, 1147–1149.

Oberstrass, F.C., Lee, A., Stefl, R., Janis, M., Chanfreau, G., and Allain, F.H. (2006). Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. Nat. Struct. Mol. Biol. *13*, 160–167.

Ponting, C.P. (1995). SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. Protein Sci. *4*, 1928–1930.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645–650.

Query, C.C., Moore, M.J., and Sharp, P.A. (1994). Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev. *8*, 587–597.

Raker, V.A., Plessel, G., and Lührmann, R. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. EMBO J. *15*, 2256–2269.

Sampson, N.D., and Hewitt, J.E. (2003). SF4 and SFRS14, two related putative splicing factors on human chromosome 19p13.11. Gene *305*, 91–100.

Sarkissian, M., Winne, A., and Lafyatis, R. (1996). The mammalian homolog of suppressor-of-white-apricot regulates alternative mRNA splicing of CD45 exon 4 and fibronectin IIICS. J. Biol. Chem. *271*, 31106–31114.

Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J., and Mattaj, I.W. (1990). Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. Nature *345*, 502–506.

Schuster-Bockler, B., Schultz, J., and Rahmann, S. (2004). HMM Logos for visualization of protein families. BMC Bioinformatics *5*, 7.

Séraphin, B., Kretzner, L., and Rosbash, M. (1988). 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. *7*, 2533–2538.

Shindyalov, I.N., and Bourne, P.E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng. *11*, 739–747.

Spikes, D.A., Krämer, J., Bingham, P.M., and Van Doren, K. (1994). SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a highly conserved sequence motif with the prp21 family of constitutive splicing proteins. Nucleic Acids Res. *22*, 4510–4519.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315–326.

Szymczyna, B.R., Bowman, J., McCracken, S., Pineda-Lucena, A., Lu, Y., Cox, B., Lambermon, M., Graveley, B.R., Arrowsmith, C.H., and Blencowe, B.J. (2003). Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes Dev. *17*, 461–475.

Tang, J., Abovich, N., and Rosbash, M. (1996). Identification and characterization of a yeast gene encoding the U2 small nuclear ribonucleoprotein particle B″ protein. Mol. Cell. Biol. *16*, 2787–2795.

Utans, U., and Krämer, A. (1990). Splicing factor SF4 is dispensable for the assembly of a functional splicing complex and participates in the subsequent steps of the splicing reaction. EMBO J. *9*, 4119–4126.

Varani, G., and Nagai, K. (1998). RNA recognition by RNP proteins during RNA processing. Annu. Rev. Biophys. Biomol. Struct. *27*, 407–445.

Wiest, D.K., O'Day, C.L., and Abelson, J. (1996). In vitro studies of the Prp9.Prp11.Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. J. Biol. Chem. *271*, 33268–33276.

Will, C.L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. *21*, 4978–4988.

Zhu, J., and Krainer, A.R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev. *14*, 3166–3178.

Zhuang, Y., and Weiner, A.M. (1986). A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell *46*, 827–835.

Accession Numbers

The atomic coordinates for the ensemble of 20 NMR structures of SURP1 and the SURP2-SF3a60 complex have been deposited in the Protein Data Bank (Accession codes 2DT6 and 2DT7, respectively).