BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 4571–4582 4571
DOI: 10.1021/bi100374d

Coevolution of *Drosophila snf* Protein and Its snRNA Targets†

Sandra G. Williams and Kathleen B. Hall*

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110

Received March 10, 2010; Revised Manuscript Received May 4, 2010

ABSTRACT: SNF is a protein that is found in the U1 and U2 snRNPs (small nuclear ribonucleoproteins) of *Drosophila*. Its mammalian counterparts are two homologous proteins, U1A and U2B″. In vivo, these proteins segregate to the U1 and U2 snRNPs, respectively, where they bind distinct RNA hairpins. The RNA binding properties and mechanism of U1A have been studied extensively, but much less is known about SNF and U2B″ binding to their RNA targets. By comparing thermodynamic aspects of SNF−RNA interactions with those of U1A−RNA interactions, we find that SNF binds its RNA targets in a manner that is distinct from that of U1A. In vitro, SNF is able to bind both *Drosophila* U1 stem−loop II and U2 stem−loop IV with high affinity, although it binds stem−loop II more tightly than it binds stem−loop IV. Intriguingly, SNF is unable to bind human U2 stem−loop IV, which suggests that both the protein and RNAs have coevolved to interact with each other such that a single protein can bind RNAs that are more commonly bound by two distinct proteins.

RNA recognition motifs (RRMs)[1] are the most abundant RNA binding element in eukaryotic proteins. To date, more than 6000 distinct RRMs have been identified, and RRMs are found in ∼2% of all human genes (*1*). RRMs are involved in processes that include splicing, translation, and RNA trafficking. They are often part of a protein containing multiple functional domains, and in this context, they play key roles in linking RNA binding with RNA regulation. In spite of the high degree of structural similarity between RRMs, they are highly variable with respect to the RNA sequences and structures that they bind, the affinity they have for RNA, and the mechanisms by which they interact with their RNA targets. It is not possible to predict ab initio what RNA sequence an RRM will bind, how tightly it will bind it, and the mechanism by which it will recognize its RNA. It is clear that as yet, we have a very poor understanding of how RRMs work.

SNF as a Member of the snrpA/snrpB2 Family. SNF (Sans fille) is a member of the *snrpA*/*snrpB2* protein family. Members of this family are found in all eukaryotes, and in general, proteins in this family contain two RRMs. In most eukaryotes, these proteins are found in pairs; U1A segregates to the U1 snRNP and U2B″ to the U2 snRNP, where they bind highly related but distinct RNA targets within the U1 and U2 snRNA, respectively. In a minority of organisms, neither U1A nor U2B″ is present. Instead, a single protein, SNF, seems to replace both U1A and U2B″ and is found in both the U1 and U2 snRNPs.

The function of these proteins in splicing remains unclear, although U1A and U2B″ proteins are extremely conserved among eukaryotes. U1A binds with high affinity and specificity to stem−loop II (SLII) of U1 snRNA, while U2B″ binding to its RNA target, SLIV in U2 snRNA, is less well characterized. The U1 snRNP can be functionally reconstituted without the U1A protein (*2*); analogous experiments with U2B″ are not available, although it appears to remain in the spliceosome with the U2 snRNP throughout the splicing cycle (*3*). In *Caenorhabditis elegans*, the U1A and U2B″ proteins are redundant: in the absence of either one, the remaining protein is incorporated into both snRNPs (*4*), suggesting that each protein can bind both RNA targets. In *Drosophila*, SNF is found in both snRNPs, where it is assumed to bind to either SLII or SLIV depending on its snRNP context.

A *snf* knockout is embryonic lethal to the fly, but the reason for this phenotype remains unclear. Mutational experiments have shown that in *Drosophila*, SNF is not required for recognition of the 5′ splice site by the U1 snRNP (*5*) or for U2 snRNP activity (*6*). Salz (*6*) has suggested that rather than playing a primary role in splicing events, SNF interacts with modulators to modify splice site selections, thus controlling U1 (and possibly U2) snRNPs at critical times in response to cues from other proteins. If this is SNF's true role in the spliceosome, it indicates that inherent to SNF's function is a flexibility of interactions. Part of this flexibility includes an ability to bind different RNA targets.

SNF and Its RRMs. SNF is a 216-amino acid protein with two RRMs connected by a flexible linker of approximately 35 amino acids. As with other RRMs, the two RRMs of SNF adopt an α/β-sandwich global fold with two α-helices and an antiparallel, four-stranded β-sheet. The RRM domains of SNF are homologous to those of U1A and U2B″ (Figure 1a,b), and although much work has been done to characterize human U1A, much less has been done to understand how human U2B″ binds RNA.

By analogy to human U1A, the N-terminal RRM (RRM1) is likely to be the workhorse of SNF. The canonical site of RNA binding in RRMs is on the surface of the β-sheet, where an RRM

†This work was supported in part by National Institutes of Health Grant F31-GM089576-01 (S.G.W.).

*To whom correspondence should be addressed. Phone: (314) 362-4196. Fax: (314) 362-7183. E-mail: kathleenhal@gmail.com.

[1]Abbreviations: SNF, Sans fille; RRM, RNA recognition motif; snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein particle; SLII, stem−loop II; SLIV, stem−loop IV; dSL, *Drosophila* stem−loop; hSL, human stem−loop; FL, full-length; 6-FAM, 6-carboxyfluorescein; CD, circular dichroism; MRE, mean residue ellipticity; PDB, Protein Data Bank.

Published on Web 05/10/2010

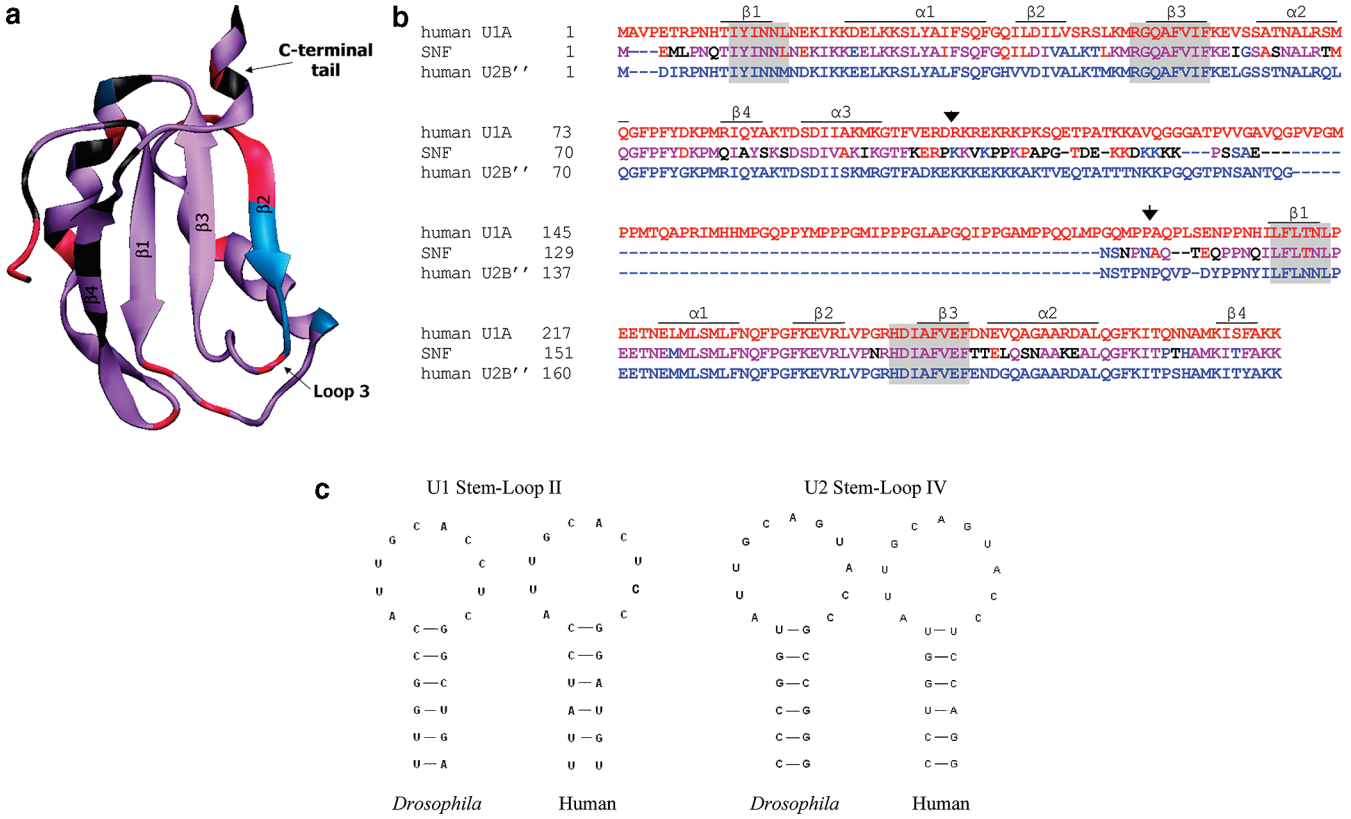

FIGURE 1: Cartoon of a crystal structure of U1A RRM1 (PDB entry 1URN) (*23*). With respect to color, purple residues are identical in SNF, U1A, and U2B″, red residues are shared by U1A and SNF, blue residues are shared by U2B″ and SNF, and black residues are unique to SNF. The C-terminal tail, loop 3, and β-strands are labeled. RNP1 is located on β_3, and RNP2 is located on β_1. (b) Amino acid sequences of human U1A, U2B″, and *Drosophila snf* proteins. Human U1A residues are colored red, and U2B″ residues are colored blue. For the SNF sequence, residues are colored as in panel a. Sequences highlighted in gray correspond to the RNP motifs. The first arrow indicates the end of the RRM1 construct, and the second arrow indicates the beginning of the RRM2 construct. (c) RNAs recognized by human U1A, U2B″, and *Drosophila* SNF. SLII of the U1 snRNA and SLIV of the U2 snRNA are recognized by U1A and U2B″, respectively. SNF recognizes both stem−loop structures.

recognizes between two and eight nucleotides of single-stranded RNA (*7*). SNF RRM1 contains several highly conserved motifs that are critical for RNA binding in canonical RRMs. These are the "RNP motifs" located on β_1 and β_3, which contain solvent-exposed aromatic amino acids that stack with RNA bases. In SNF, both RNPs of RRM1 are identical to those of both U2B″ and U1A, with the exception of the last residue of RNP2 (Figure 1b). The loop that connects β_2 and β_3 (loop 3) is critical for RNA recognition, and in SNF, loop 3 appears to be a chimera of its human counterparts (Figure 1a). A noncanonical feature of the *snrpA*/*snrpB2* family of proteins is the C-terminal tail of RRM1, which appears to contain a third α-helix and which contributes directly to RNA recognition.

The RNA binding partners of U1A, U2B″, and SNF are U1 snRNA stem-loop II (SLII) and U2 snRNA stem-loop IV (SLIV) (Figure 1c). Although these RNAs are quite similar to each other, human U1A binds to SLII with subnanomolar affinity but is unable to bind SLIV. Knowledge of U2B″ binding its target RNA is much more limited, and accounts differ with respect to its affinity and specificity for SLII and SLIV RNAs. *Drosophila* U1 SLII and U2 SLIV are also quite similar, and SNF must be able to recognize both RNAs.

The U1A−SLII interface is characterized by a number of interactions. The RNP motifs are involved in nonspecific stacking interactions, as well as specific hydrogen bonding to the 5′ half of the RNA loop. Loop 3 protrudes through the RNA loop, disrupting base stacking as well as forming specific contacts with SLII. Loop 3 contacts the RNA bases as well as other parts of the protein, and the hydrogen bonding network presumably holds both the loop and RNA in place, opening up the RNA loop so that it is presented as a single strand. Additionally, the conserved TDS sequence mediates interactions between the C-terminal tail of RRM1 (from the end of β_4 through the beginning of the third helix) and the CAC nucleotides at the top of the RNA loop. Conspicuously absent are interactions between U1A and the last three nucleotides (UCC) at the 3′ end of the RNA loop.

From a cocrystal complex of human U2B″ RRM1 with SLIV (*8*), it appears that most protein−RNA interactions are shared with those observed between U1A and SLII. However, this cocrystal structure demonstrated a number of additional interaction sites. Predominant among these are interactions between loop 3 of the protein and the UACC nucleotides at the 3′ end of the SLIV loop. The complex was crystallized in the presence of human U2A′, a member of the leucine-rich repeat protein family. There has been debate over whether this protein is required for U2B″ to bind SLIV. *Drosophila* also contains a U2A′ protein, which appears to associate with SNF in the U2 snRNP (*9*). Whether this protein is required for SLIV binding by SNF has also been unclear.

For 15 years, we have known that the *snrpA*/*snrpB2* system functions differently in *Drosophila* than it does in humans. However, the wealth of sequencing information that is currently available provides the ability to undertake a much more comprehensive phylogenetic comparison than simply aligning *Drosophila* SNF, human U1A, and human U2B″. Two fundamental questions emerge from a phylogenetic analysis of this family. First,

how do the binding mechanisms across the *snrpA*/*snrpB2* family of proteins differ? Answering this question should greatly inform our understanding of RNA−protein interactions and, more specifically, how binding specificity can be achieved by RRMs. The second question is whether the different binding mechanisms result in similar biological consequences, regardless of the organism. Given the high degree of evolutionary conservation in this system, the presumption is that the final biological results of binding are similar; that is, Nature has solved one problem in multiple ways.

Although a sequence analysis of these proteins has proven to be informative, it provides no mechanistic information about protein−RNA interactions. In this work, we study the thermodynamics of SNF−RNA interactions. The dual ability of SNF to bind both U1 SLII and U2 SLIV allows us to probe whether the protein binds the two RNAs differently and, if so, what constitutes the basis of these differences. From a thermodynamic perspective, we show that the RNA binding properties of SNF are distinct from those of U1A. Furthermore, the binding mechanism changes depending on the RNA target of SNF. SNF's ability to bind *Drosophila* SLIV along with its inability to bind human SLIV indicates that evolutionary adaptation has occurred in both the protein and RNA.

MATERIALS AND METHODS

Cladogram. The original curated phylogenetic tree for the *snrpA*/*snrpB2* protein family was obtained from TreeFam (www.treefam.org; accession number TF313834). The complete trees are available in Figure 1 of the Supporting Information, as well as online. These data were pruned and redrawn using the treedrawing software Archaeopteryx (available at http://www.phylosoft.org/archaeopteryx/).

RNA Synthesis. RNA stem−loop structures for nitrocellulose filter binding assays were enzymatically synthesized with T7 RNA polymerase from DNA oligonucleotides, as described previously (*10*, *11*). DNA was obtained from IDT (Integrated DNA Technologies). The RNA was internally labeled with [α-^{32}P]UTP and [α-^{32}P]CTP. Labeled RNAs were gel-purified before being used in binding assays. The RNA product for dU1 SLII was 5′-GGCUUGGCCAUUGCACCUCGGCUGAGCC. The RNA product for dU2 SLIV was 5′-GGCCGGUAUUGCAGUACCGCCGGGUCC. The RNA product for hU1 SLII was 5′-GGAGACCAUUGCACUCCGGUUUCC. The RNA product for hU2 SLIV was 5′-GGUGGUAUUGCAGUACCUCCACC. Regions corresponding to the RNA loops are underlined. Experiments for nonspecific binding were conducted using a 25-nucleotide RNA of random sequence.

For fluorescence binding experiments, 6-carboxyfluorescein (6-FAM) was used to label dSLIV. The labeled RNA was obtained from IDT. The RNA sequence was 5′-6-FAM-GGGCCCGGUAUUGCAGUACCGCCGGGUCC.

Protein Purification. All protein constructs were isolated from *Escherichia coli* BL 21 Codon-Plus RP cells (Stratagene) transformed with a plasmid carrying the protein of interest under control of a TAC promoter. The *E. coli* strain contained extra copies of the *argU* and *proL* genes, which encode tRNAs that recognize codons that are more prevalent in the *Drosophila* genome than in the *E. coli* genome. The SNF RRM1 construct is composed of residues 1−102. The SNF RRM2 construct is composed of residues 134−216, as well as an additional N-terminal methionine. Cells were grown in LB medium at 37 °C and induced in midlog phase with 1 mM IPTG. FL SNF and SNF RRM1 were grown for an additional 6 h at 30 °C, whereas SNF RRM2 was grown for an additional 4 h at 37 °C. The cells were pelleted and stored at −70 °C until lysis. Containers used in the purification were acid-washed before use, and all solutions were filtered through a 0.45 μm cellulose nitrate (CN) filter (Nalgene) to remove RNases.

Two different methods of purification were used, depending on the protein construct. For FL SNF and SNF RRM1, cells were resuspended on ice in 30 mM sodium acetate (pH 5.3), 200 mM NaCl, 2 mM EDTA, and 8.5% sucrose. A protease inhibitor cocktail (Sigma), PMSF, and DNase II were added prior to lysis. The suspension was passed through a French press four times. The lysate was spun down at 4 °C and 45000*g* in an ultracentrifuge. The supernatant was passed through a 0.22 μm cellulose acetate filter and loaded directly onto an SP Sepharose column pre-equilibrated in 50 mM Tris (pH 7.5). FL SNF was eluted using a gradient running from 275 to 360 mM NaCl. RRM1 was eluted with a gradient running from 100 to 350 mM NaCl. Immediately after elution, EDTA and PMSF were added to protein-containing fractions to final concentrations of 5 mM and 20 μg/mL, respectively. This was done to minimize protein degradation by trace proteases, which was otherwise substantial.

The estimated pI for SNF RRM2 was 6.1, whereas FL SNF and SNF RRM1 are basic. Because purification was achieved via ion exchange chromatography, SNF RRM2 had to be purified differently. Cells were resuspended in 20 mM Tris-HCl (pH 7.5), 20 mM NaCl, and 2 mM EDTA. A protease inhibitor cocktail (Sigma), PMSF, and DNase II were added prior to lysis. The suspension was passed through a French press four times. The lysate was centrifuged at 25000*g* for 40 min. The supernatant was fractionated with 30% followed by 65% ammonium sulfate. The 65% fractionation was centrifuged at 25000*g* for 45 min, and the pellet was resuspended in 50 mM Tris-HCl (pH 7.5). This was dialyzed overnight against 50 mM Tris-HCl. After centrifugation, the supernatant was loaded onto a CM column pre-equilibrated with 50 mM Tris-HCl (pH 7.5). The column was run to baseline, and SNF RRM2 was eluted (no gradient was used). Fractions containing SNF RRM2 were collected and loaded onto a Q-Sepharose column pre-equilibrated with 20 mM Tris-HCl (pH 7.5). Fractions were eluted with a gradient run from 0 to 300 mM NaCl.

After purification, fractions were concentrated using Vivaspin concentrators, and the protein was exchanged into 10 mM sodium cacodylate and 50 mM KCl (pH 7) for storage. The concentration of all proteins was calculated spectrophotometrically. For FL SNF and SNF RRM1, $\varepsilon_{280} = 5120\ M^{-1}\ cm^{-1}$ (both proteins contain four tyrosine residues). For RRM2, the protein concentration was calculated using an ε_{260} of $1152\ M^{-1}\ cm^{-1}$ as RRM2 contains eight phenylalanine residues and an $\varepsilon_{260,Phe}$ of $144\ M^{-1}\ cm^{-1}$. For all three constructs, the protein absorption spectra from 220 to 360 nm were identical regardless of whether the spectra were recorded in denaturing conditions (8 M urea) or not (0 M urea).

Circular Dichroism Spectra and Unfolding Experiments. All CD spectra were recorded using a Jasco J715 instrument. Recordings were taken at room temperature. The buffer contained 50 mM KCl and 10 mM sodium cacodylate (pH 7) and a protein concentration of 20 μM. For denaturation studies, the concentration of urea was varied and the mean residue ellipticity (MRE) at 221 nm was followed as a function of the urea concentration. Chemical denaturation using guanidine chloride

was also performed, but for SNF RRM1, this resulted in a folded baseline that was too small to be used for successful fitting of the data. Data were fit in Scientist (Micromath) using the linear extrapolation method (*12*) for which the equation

$$y = \{(y_D + m_D C) \times \exp[(mC - \Delta G^\circ_{D,H_2O})/RT] + y_N + m_N C)\} / \{1 + \exp[(mC - \Delta G^\circ_{D,H_2O})/RT]\}$$

was used. y is the observed MRE; C is the concentration of urea, and m is the slope of the unfolding transition. y_N and y_D are the intercepts of the native and denatured baselines, respectively, and m_N and m_D are the slopes of the native and denatured baselines, respectively. $\Delta G^\circ_{D,H_2O}$ is the standard unfolding free energy in the absence of denaturant, and $\Delta G_D = \Delta G^\circ_{D,H_2O} - mC$.

For refolding experiments, a solution of >10 M urea was added to proteins such that the final urea concentration was 8 M (9 M was used for SNF RRM2). The proteins were allowed to equilibrate for at least 6 h. The unfolded protein samples were then diluted into lower urea concentrations and allowed to equilibrate overnight, before CD spectra were recorded. For plotting the refolding data, MRE values for SNF RRM1 were normalized by a factor of 1.05, and MRE values for FL SNF were normalized by a factor 1.1 to account for differences in protein concentration between the stocks used for unfolding and refolding experiments.

Filter Binding Assays. Nitrocellulose filter binding assays were used to determine standard binding free energies for binding of RNA to different SNF constructs, as described previously (*13*, *14*). A constant, picomolar concentration of RNA and variable protein concentrations were used. BSA (Roche) was added to a final concentration of 40 μg/mL. All experiments described in the text were conducted at pH 7. However, additional experiments with the pH varying between 6 and 8 show that binding is pH-independent within this range. Solution conditions were otherwise variable and are indicated in the text and figures. All experiments were performed in duplicate, and binding curves were fit to a standard Langmuir isotherm using Scientist (Micromath).

RNA−Protein Binding Assessed by Fluorescence. Fluorescence experiments were conducted using an SLM 8000 instrument. A circulating water bath was used to control the cuvette temperature. Reaction buffers contained 20 μg/mL BSA and 10 mM potassium phosphate (pH 8). Variable amounts of KCl and $MgCl_2$ were used, as described in the text. A fixed RNA concentration of 10 nM 6-FAM-dSLIV was used in all experiments.

Acid-washed cuvettes were blocked for 1 h with buffer. The excitation and emission wavelengths were set to 490 and 520 nm, respectively, and polarizers were set at the magic angle for fluorescence intensity measurements. For protein titrations, fluorescence anisotropy and fluorescence intensity were recorded as functions of the total protein concentration. While both measurements yielded similar binding isotherms, fluorescence intensity measurements provided data sets that were less noisy and were therefore used for further analysis. The fluorescence intensity of the dye was significantly enhanced when protein bound the RNA. Binding curves were fit to a standard Langmuir isotherm using Scientist.

RESULTS

Evolution of snrpA/snrpB2 Proteins. *snf* proteins, found in a variety of flies, are evolutionary oddballs in that they are chimeric proteins that can and do function in the place of two separate proteins, U1A and U2B″. Phylogenetic analysis of the *snrpA/snrpB2* family of proteins, which contains U1A, U2B″, and SNF, reveals that most organisms have two separate proteins that function like mammalian U1A and U2B″ in terms of their specificity for snRNAs. This is depicted in a cladogram of representatives of this protein family in different eukaryotes (Figure 2). The nodes of the cladogram are colored so that inferred gene duplication events are red and speciation events are blue. As is evident from the tree, within a phylum, U1A and U2B″ are more closely related than U1A is to other U1A proteins outside of the phylum. A larger tree (which includes evolutionary distance information) is included in Figure 1 of the Supporting Information.

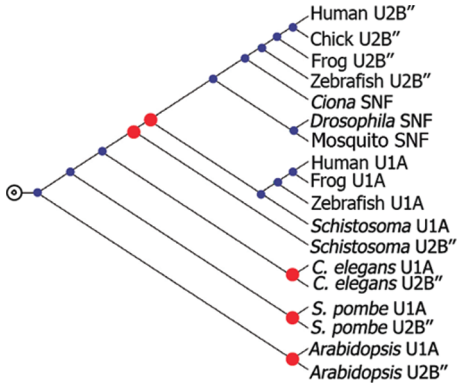

FIGURE 2: Cladogram of representatives of the *snrpA/snrpB2* protein family. The original phylogenetic tree was obtained from TreeFam, and the data were pruned and redrawn using Archaeopteryx. Gene duplication nodes are colored red, and speciation nodes are colored blue. The concentric circles indicate the common ancestor.

As drawn, the cladogram in Figure 2 suggests that the current U1A, U2B″, and *snf* proteins arose from at least five gene duplication events (three of which occurred in animals, along with separate duplication events in plants and fungi). Additionally, at least two gene loss events would have had to occur. This evolutionary history is quite complex, and it is important to remember that this history is inferred from the relatedness of the gene sequences. As an alternative to multiple gene duplications, it was proposed (*15*) that a single gene duplication event, presumably occurring in a much earlier common ancestor, could be responsible for the current collection of U1A and U2B″ genes if these genes were coevolving within an organism. This would account for the relatedness of U1A and U2B″ within phyla.

Regardless of when the gene duplication event(s) took place and what the relevant selection pressures have been for these proteins, there has been significant sequence divergence within this protein family. The proteins, however, are only part of the story: the RNA targets have also diverged. While SLII and SLIV sequences are nearly invariant in vertebrates, the sequences vary significantly in other phyla (Figure 3b). This begs the question of whether the proteins are recognizing their target RNAs through a common mechanism or whether these proteins have evolved different ways of identifying their target RNAs. To begin to address this question, we chose to compare protein sites that were identified as important for interactions between U1A or U2B″ and RNA. The analyzed protein sequences included the third α-helix and its preceding TDS sequence, the RNP motifs, and loop 3. Figure 3a shows sequence alignments of these regions from U1A and U2B″ in various organisms (full sequence alignments are included in Figures 2 and 3 of the Supporting Information). The sequences were aligned and numbered according to the

a

		RNP2	K23	β_2	L3	RNP1	TDS	α_3
U1A	Vertebrates	IYINNL	K	ILDILV	SRXXKM	RGQAFVIF	TDS	DIIAKMK
	Plants	IYINNL	I/L	ILEXLA	FKTXKH	KGQAWVVF	TKS	DXXAKAD
	C. elegans	IYINNL	I	IVSVMC	FRTLKM	RGQAHVIF	EDS	DVVAQEK
	S. mansoni	IYINNM	C	ILDIIT	SRTLKM	RGQAFVVF	VDS	DVIAKRK
	S. pombe	LYIRNI	L	VIDVQA	RKTLRM	RGQAFVVF	SKS	DIIVQRE
SNF	*Drosophila*	IYINNL	K	ILDIVA	LKTLKM	RGQAFVIF	SDS	DIVAKIK
	Ciona	IYANNL	K	ILDIVA	MKTLKM	RGQAFVIF	KDS	DVIAKMK
U2B″	Vertebrates	IYINNM	K	VVDIVA	LKTMKM	RGQXFVIF	TDS	DIISKMR
	Plants	IYIXNX	K	IXDXXA	LKTPKL	RGQXWVXF	XKS	DCXAKXE
	C. elegans	IYVNNL	K	IIQLMS	FRKEKM	RGQAHIVF	EDS	DVISRAK
	S. mansoni	LYVNNL	K	LIDIIA	MKTMKM	RGQAFIIY	KDS	QQIICLK
	S. pombe	LYVNNL	K	VVDIVA	LKTPKM	RGQAHVVF	SKS	KIIERIV

b

	U1 SLII RNAs	U2 SLIV RNAs
Vertebrates	UXAXCCAUUGCACUCC GGAUGU	CCXGGUAUUGCAGUACCUCCXGG
Plants	ACCUCCAUUGCACAUAXXGGAGGG	CCAUGCGUUGCACUACUGXACGG
Drosophila	UUGGCCAUUGCACCUC GGCUGA	CCCGGUAUUGCAGUACCGCCGGG
Ciona	CAGGCCAUUGCACAUC GGCUUA	CCUGGUAUUGCAGUACCUCCAGG
C. elegans	CUACCCAUUGCACUUUU GGUGCG	CCUGGCGUUGCACUGCUGCCGGG
S. mansoni	CUCACCAUUGCACUUC GGUGGG	CCUCGUAUUGCAUUGCCUCGAGG
S. pombe	CUUGGCAUUGCACUGA GCCCUG	CUUGCUAUUGCACUUACUGGCAA

FIGURE 3: U1A, U2B″, SLII, and SLIV sequence alignments from representative eukaryotes. (a) Sequence alignments of RNP motifs, loop 3 (L3), and additional regions in RRM1 that have been implicated in protein−RNA interactions. (b) Sequence alignments of U1 SLII and U2 SLIV RNAs in multiple organisms. The loop sequences are underlined. The vertebrate sequences analyzed included human, rat, mouse, frog, and zebrafish proteins and RNAs. The representative plants were potato, soybean, and *Arabidopsis thaliana*. An X indicates imperfect sequence conservation among the representative species. Complete sequences are available in Figures 2−4 of the Supporting Information.

corresponding residues of U1A. Comparison of the sequences shows an intricate pattern of both conservation and divergence. Five residues were universally conserved in all species analyzed: Y13 and N16 in β_1, K20 in loop 1, Q54 in β_3, and K80 in loop 5. These include residues from both RNPs (Y13, N16, and Q54). A number of regions diverged considerably between species but showed remarkable similarity between U1A and U2B″ within the same species. These include the TDS sequence (residues 89−91), RNP1 (especially at the variable position 55), and the second half of the third α-helix. These are sites in which there appears to have been significant evolutionary pressure for the two proteins to coevolve.

Finally, there are a number of regions that are highly conserved in either U1A or U2B″ but that are much less conserved in the partner protein. For example, the third α-helix and both RNPs are much more variable in U2B″ proteins, while they tend to be conserved among U1A proteins. In contrast, K23 and the VALKT sequence of loop 3 are much more highly conserved in U2B″ proteins. Presumably, these regions are important for specifying the preferred RNA target and recognize characteristics of a given RNA stem−loop found in all phyla. As such, U1A or U2B″ appear to be coevolving with their RNA targets in these regions.

To further define the role of coevolution in the system, we were curious to see if specific amino acid variations tracked with changes in the RNA sequences. Therefore, we used information from available cocrystal structures of RNA-bound U1A and U2B″ to identify interaction partners. Although there were some suggestive variations in which nucleotide changes seemed to track with amino acid changes, these were few and subtle. The hydrogen bonding patterns of human U1A and U2B″ binding to their RNA targets are both complex and distinct. Given that proteins from other phyla are even more divergent, it appears that the specific interactions that characterize the binding interfaces are not identical. This may account for the poor correlation between nucleotide and amino acid variation.

With SNF, we can ask how a single protein is able to recognize two different RNAs with high affinity and specificity. Although we hypothesize that protein dynamics and subtleties of protein structure likely contribute to SNF's ability to bind two RNAs, our initial experiments have focused on comparing the thermodynamics of SNF and U1A. The thermodynamic differences between the proteins undoubtedly reflect the mechanistic differences of RNA recognition and protein function.

Protein Structure and Stability. (*i*) *Structure*. Structural differences between U1A and SNF certainly have the potential to influence RNA binding and are therefore important to consider when trying to understand the different RNA binding properties of SNF compared with those of U1A. Although we assume that the structures of the two RRMs are comparable to those of U1A, we are currently characterizing them further by NMR (data not shown). The NMR spectra we have obtained are well-dispersed and show the protein to be folded under our solution conditions. Like U1A, SNF is composed of two RRMs, each of which is approximately 90 amino acids long. One noticeable difference between SNF and U1A is that the linker connecting these two RRMs is much longer in U1A. Given that in both systems our NMR data show that the two domains tumble independently, the consequences of this difference are unclear.

To determine whether the secondary structures of U1A and SNF were similar, we measured CD spectra of our different SNF constructs. Far-UV CD spectra for the full-length (FL), first-domain (RRM1), and second-domain (RRM2) constructs of SNF are compared in Figure 4a. These spectra are qualitatively similar to those recorded for U1A protein constructs (*16*, *17*). All constructs yield spectra consistent with folded proteins that contain significant β-sheet character. As with U1A, RRM1 displays substantially greater helical character than RRM2, as indicated by a more negative mean residue ellipticity at 220 nm. This is consistent with the presence of an additional helix as well as with the longer helix lengths found in RRM1. Although the magnitude of the MRE for RRM2 at 221 nm is small, this domain is certainly folded, as indicated by the well-dispersed NMR spectrum and the clear transition observed in CD spectra upon titration with urea or guanidine. The molar ellipticities of

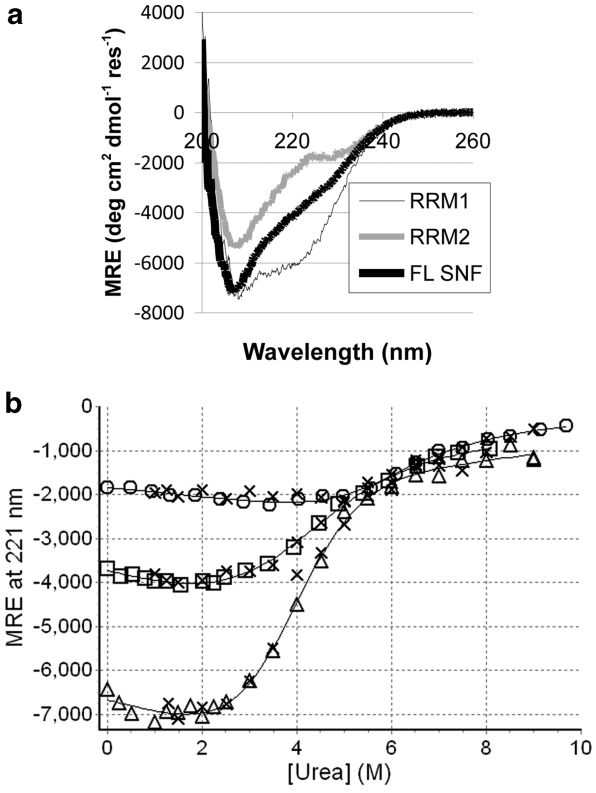

FIGURE 4: (a) Far-UV CD spectra of RRM1, RRM2, and full-length SNF plotted as mean residue ellipticity (MRE) as a function of wavelength (nm). (b) Representative unfolding curves for the chemical denaturation of FL SNF (□), RRM1 (△), and RRM2 (○). The mean residue ellipticity (deg cm^2 dmol^{-1} residue^{-1}) at 221 nm is plotted as a function of urea concentration (M). Fits to a two-state unfolding model are shown, and data from refolding experiments are indicated (×). All spectra were recorded in 50 mM KCl, 10 mM sodium cacodylate, pH 7.

the RRM1 and RRM2 constructs add to yield approximately the molar ellipticity of the full protein (not shown). We attribute the difference to the disordered linker, which is present only in the full-length construct.

(*ii*) *Stability*. Chemical denaturation of RRM1, RRM2, and FL SNF with urea is fully reversible and can be fit to a two-state model of unfolding (Figure 4b). RRM1 has an unfolding free energy ($\Delta G°_{D,H_2O}$) of 3.5 ± 0.3 kcal/mol. RRM2 has a $\Delta G°_{D,H_2O}$ of 4.8 ± 0.6 kcal/mol (Table 1). Both of these values are significantly reduced from those of U1A. In U1A, RRM1 has a $\Delta G°_{D,H_2O}$ of 9.4 ± 0.5 kcal/mol (*18*) and RRM2 has a $\Delta G°_{D,H_2O}$ of 8.3 ± 0.8 kcal/mol (*17*). Clearly, both the RNA recognition motifs of SNF are significantly destabilized as compared with their U1A counterparts. While it was possible to fit the unfolding of FL SNF according to a two-state model, the unfolding free energies of RRM1 and RRM2 are similar enough that it is not possible to distinguish whether the domains unfold independently in the full-length protein.

RNA Binding. Human U1A binds only to human SLII, whereas the RNA binding properties of human U2B″ are unclear. Accounts about whether U2B″ can recognize both SLII and SLIV and whether an auxiliary protein, U2A′, is required for the protein to bind RNA differ. A rigorous thermodynamic analysis of RNA binding has not been undertaken for U2B″. While studies from genetics experiments suggested that SNF could bind both U1 SLII and U2 SLIV, it was not clear from the experiments whether SNF required U2A′ to bind SLIV. Certainly, these experiments provided no sense of the affinity SNF had for its RNA targets. Additionally, it was possible that changes to the RNA, as well as to the protein, were contributing to SNF's ability to recognize two RNA targets. To address these questions, we performed nitrocellulose filter binding experiments with full-length SNF and ^{32}P-labeled RNA hairpins. The hairpins contained human and *Drosophila* loop sequences and loop closing base pairs and were synthesized with 8 or 9 bp stems for binding assays. Data from protein titrations were fit assuming a 1:1 stoichiometry, which we determined for SNF and dSLIV by fluorescence experiments (data not shown). Apparent binding constants (K_d) for SNF binding to human and *Drosophila* RNAs are given in Table 2 at two salt concentrations. Under both conditions, SNF binds to human and *Drosophila* SLII with similar affinity, but with a small loss of affinity compared to U1A. In contrast, at 100 mM KCl, SNF binds to human SLIV with an affinity that is approximately 2 orders of magnitude weaker than its affinity for *Drosophila* SLIV. In experiments conducted in 250 mM KCl, SNF bound dSLIV with an affinity of (7.7 ± 3.0) × 10^{-8} M, but binding to hSLIV could not be distinguished from binding to a random 25mer (Figure 5). Binding to either hSLIV or the random 25mer was very weak (micromolar at best) and could not be fit to yield a binding constant. It is evident that SNF can bind dSLIV in the absence of U2A′ and that the ability to bind dSLIV has resulted from evolutionary adaptations in the RNA, as well as in the protein.

Table 1: SNF Unfolding Free Energies[a]

	$\Delta G°_{D,H_2O}$ (kcal/mol)	m_D
RRM1	3.5 ± 0.3	0.85 ± 0.11
RRM2	4.8 ± 0.6	0.92 ± 0.07

[a] $\Delta G°_{D,H_2O}$ is the standard unfolding free energy in the absence of denaturant as determined by linear extrapolation. m_D is the slope of ΔG_D vs denaturant concentration and is one of the terms of the linear extrapolation model. All experiments were conducted in duplicate, and the estimates from nonlinear least-squares fits were averaged to give the values shown. Errors for $\Delta G°_{D,H_2O}$ and m_D are estimated from the nonlinear least-squares fit.

Salt Dependence of RNA Binding. It is impossible to consider mechanisms for RNA−protein interactions without considering electrostatic contributions to binding. The association of cations with highly charged nucleic acids reduces the unfavorable charge density of the polyphosphate backbone through counterion condensation and screening effects. When the ligand binds, some of the associated cations are displaced to allow association of the ligand. This results in a salt dependence of the apparent binding constant, K_{obs}. For an oligopeptide binding to a nucleic acid, analysis of the salt dependence of binding can yield the number of cations released from the nucleic acid and thus provide mechanistic information about binding. For instance, the release of the cation from the RNA is entropically favored, and in some cases, it has been shown that a binding reaction can be driven by cation release (*19*). As important, the salt dependence of these interactions provides a framework for understanding the extent to which electrostatics contribute to interaction specificity between the protein and RNA target (*20*). To this end, we wanted to understand whether electrostatic contributions to binding were different for the U1A−SLII interaction versus the SNF−SLII interaction. Additionally, we wanted to know whether electrostatics contributed to the difference in binding affinity for SNF binding to SLII as compared with SLIV.

Table 2: Apparent Binding Constants (K_d) for SNF−RNA Interactions

	100 mM KCl		250 mM KCl	
stem−loop	hSL (M)	dSL (M)	hSL (M)	dSL (M)
II	$(1.6 \pm 0.8) \times 10^{-10}$	$< 1 \times 10^{-10}$	$(3.3 \pm 0.2) \times 10^{-9}$	$(7 \pm 6) \times 10^{-10}$
IV	$(2.9 \pm 0.5) \times 10^{-7}$	$(5.6 \pm 1.6) \times 10^{-9}$	micromolar	$(7.7 \pm 3.0) \times 10^{-8}$

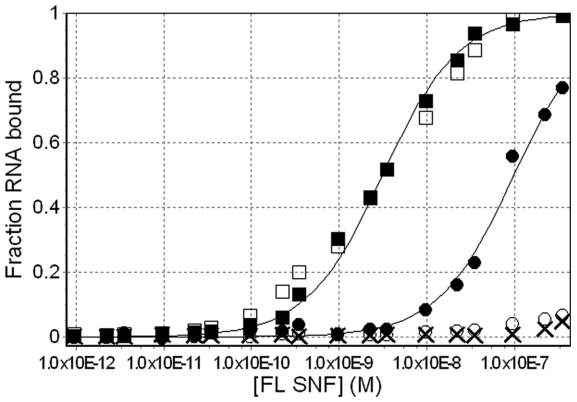

FIGURE 5: Titrations for FL SNF binding different RNAs. Nitrocellulose filter binding experiments were conducted to determine binding constants for FL SNF binding to dSLII (■), hSLII (□), dSLIV (●), hSLIV (○), or a random 25mer sequence (×). Fits to a model of single-site binding are shown for SNF binding to dSLII, hSLII, and dSLIV. Experiments were conducted at 22 °C in 250 mM KCl, 10 mM cacodylate, and 1 mM $MgCl_2$ (pH 7).

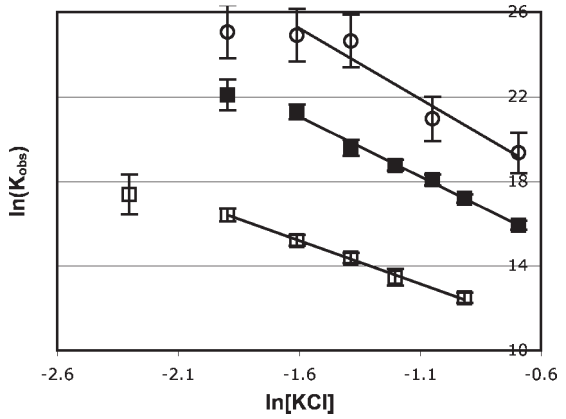

FIGURE 6: Comparison of the KCl dependence of binding for U1A binding to hSLII (○), SNF binding to dSLII (□), and SNF binding to dSLIV (■). Individual binding constants were obtained from nitrocellulose filter binding experiments. Experiments were conducted in 10 mM cacodylate and 2 mM $MgCl_2$ (pH 7) at 22 °C. Data for U1A binding to human SLII are from ref *16*. For the U1A data set, sodium was the monovalent cation used.

Analysis of protein−nucleic acid interactions is complicated by a number of factors, including preferential interactions of cations, anions, protons, and water with the protein. The resulting salt dependence of the interaction is therefore described by

$$\partial(\log K_{obs})/\partial(\log[MX]) = -(\Delta c + \Delta a) + 2[MX] \times \Delta w/[H_2O] \quad (1)$$

where Δc, Δa, and Δw represent the net gain or loss of cations (Δc), anions (Δa), and water (Δw), respectively (*21*). Unless the salt concentration is quite high (> 0.5 M), preferential hydration effects are expected to be small compared with the ion terms, simplifying the analysis, and the salt dependence of binding can be understood in terms of the net release of cations and anions from the interacting species. In a plot of $\log(K_{obs})$ as a function of the log of the salt concentration, the slope yields the net release (slope < 0) or uptake (slope > 0) of ion pairs upon protein−RNA association.

We measured the association constants (K_{obs}) of SNF−RNA interactions at multiple concentrations of KCl (Figure 6). When U1A binds hSLII, there is a net release of 6.7 ± 1.1 ions. These data were obtained from titrations conducted in NaCl, but it has been shown for U1A that binding is equivalent in NaCl or KCl; the binding affinity is independent of the nature of the monovalent cation used. SNF binding to dSLII results in a net release of 5.7 ± 0.2 ions, but these results are the same within error as those obtained for U1A. In contrast, dSLIV binds to SNF with a net release of 4.0 ± 0.2 ions, 1.7 fewer than the number released by dSLII binding. This suggests that the mechanism of binding of dSLII and dSLIV by SNF is different.

If the assumption is made that SLIV and SLII bind to the same site on the protein, it follows that the anion effects can be considered equivalent for the binding of SNF to both dSLII and dSLIV. If the added specificity of the protein for SLII over SLIV is independent of salt, this would indicate that electrostatics do not contribute to binding specificity (*20*). However, our analysis shows that for SNF, the salt dependence of binding its two partners is indeed different; electrostatic interactions affect the specificity of binding.

Temperature Dependence of RNA Binding. The temperature dependence of the observed binding constants provides information about the standard enthalpy and entropy change associated with the binding reaction. In addition, a nonlinear van't Hoff plot indicates an apparent heat capacity change that can be described by the following equation (*19*):

$$\ln(K_{obs}) = (\Delta C_{p,obs}/R)[T_H/T - \ln(T_S/T) - 1] \quad (2)$$

where the enthalpy and entropy are considered to be temperature-dependent. T_H and T_S are the temperatures at which the enthalpy and entropy of complex formation, $\Delta H°$ and $\Delta S°$, respectively, equal 0. $\Delta C_{p,obs}$ is the apparent heat capacity change that characterizes the reaction. There are a number of reasons why an association may indicate an apparent heat capacity change, but in general, an apparent ΔC_p accompanying a protein−ligand interaction is interpreted as resulting from hydrophobic surface burial or a coupled conformational change, of either the ligand, protein, or both.

The van't Hoff plot of U1A binding to SLII shows significant deviation from linearity, indicative of a large apparent ΔC_p. The binding event occurs with significant conformational changes in both the RNA and the protein: base stacking is disrupted in the RNA, and loop 3 of the protein is rigidified as it protrudes through the RNA loop. These changes likely contribute to the apparent ΔC_p. Given the similarities of protein and RNA structures, we thought that binding would occur with similar changes to the RNA and protein, resulting in an apparent ΔC_p. However, we investigated whether the apparent heat capacity

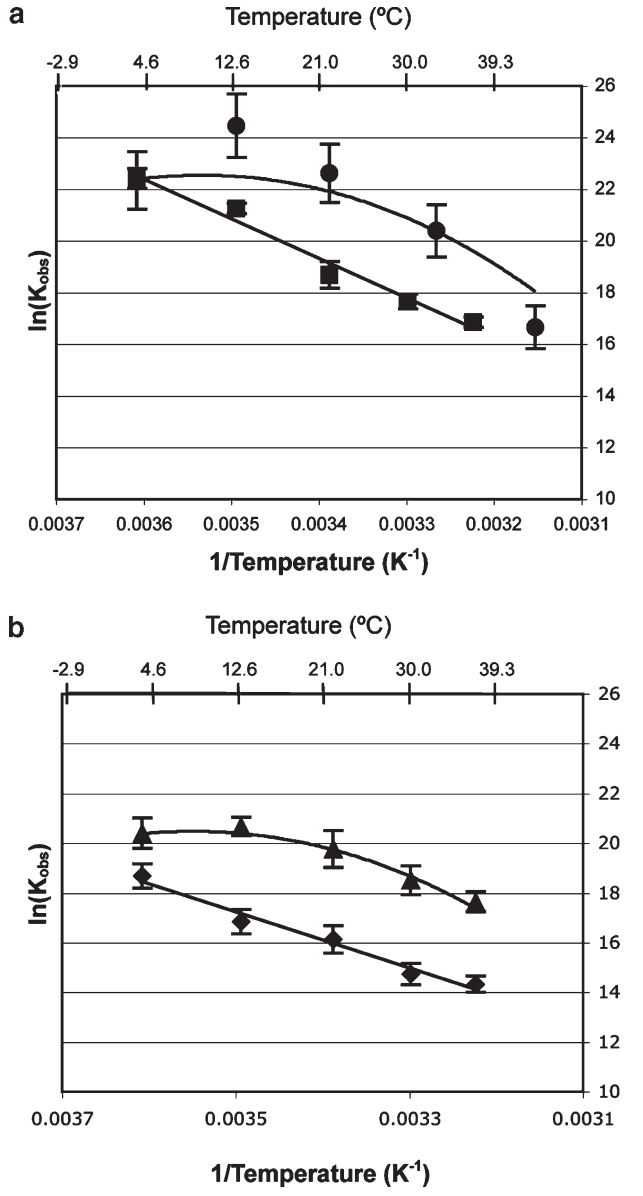

FIGURE 7: van't Hoff plots for SNF and U1A binding to target RNAs. (a) Human U1A binding to hSLII (●) and SNF binding to dSLII (■). (b) SNF binding to dSLIV in 100 mM KCl (▲) and 250 mM KCl (◆). Binding experiments for U1A were conducted in 200 mM NaCl, 10 mM cacodylate, and 1 mM $MgCl_2$. Binding experiments for SNF were conducted in 10 mM cacodylate and 1 mM $MgCl_2$ (pH 7) at variable concentrations of potassium chloride. Errors in K_{obs} in SNF experiments were estimated from the standard deviation of at least two separate results.

change was similar in magnitude for both RNA targets and whether this was similar to the ΔC_p observed for the U1A−SLII interaction. Along with comparison of the ΔC_p values for these different interactions, comparing the characteristic temperatures T_H and T_S could provide information about mechanistic differences in binding.

We measured the temperature dependence of K_{obs} for SNF−RNA interactions at a variety of salt concentrations (Figure 7 and Tables 3 and 4). We varied the salt concentration between experiments because of the difficulties in accurately measuring K_{obs} for both RNAs over the assessed temperature range (4−37 °C). In particular, binding of dSLII to SNF was too tight to measure the K_{obs} accurately at low temperatures, especially when the salt concentration was below 300 mM KCl.

Table 3: Enthalpies and Entropies of Binding for Data Fit to the van't Hoff Equation

protein−RNA	[KCl] (mM)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal mol^{-1} K^{-1})
SNF−dSLII	300	−31.1 ± 2.0	−67.3 ± 6.9
SNF−dSLIV	250	−23.4 ± 2.2	−47.6 ± 7.4
SNF−dSLIV[†]	250	−23.4 ± 3.5	−48.9 ± 11.7
SNF−dSLIV	300	−24.4 ± 1.9	−53.2 ± 6.5

Table 4: Fits of Nonlinear van't Hoff Plots[a]

protein−RNA	[KCl] (mM)	$\Delta C_{p,obs}$ (kcal/mol)	T_H (K)	T_S (K)
U1A−hII	200	−3.1 ± 0.4	284 ± 5	288 ± 2
SNF−dSLIV	100	−1.37 ± 0.42	281.7 ± 3.6	290.2 ± 1.6

[a]Apparent heat capacity change ($\Delta C_{p,obs}$), T_H, and T_S values for interactions that showed curvature in the van't Hoff plots. All results for U1A are from ref (*25*).

The results of these experiments were surprising. Unlike U1A binding to hSLII, the van't Hoff plot for SNF binding to dSLII was linear and could thus be analyzed without invoking an apparent ΔC_p. More surprising was that for dSLIV, the presence of a heat capacity change was found to be salt-dependent. Whereas in 100 mM KCl, the van't Hoff plot was nonlinear and could be fit to eq 2 to yield a $\Delta C_{p,obs}$ of −1.4 kcal/mol, at 250 and 300 mM KCl, the apparent heat capacity change vanished. While the $\Delta H°$ of complex formation was the same at these higher salt concentrations (~24 kcal/mol), $\Delta S°$ was more variable. The data also show that for SNF binding to both RNAs at higher salt concentrations, the reactions are enthalpy-driven. For SNF binding to dSLIV at 100 mM KCl, the reactions are enthalpy-driven above 17 °C but entropy-driven below 9 °C. We conclude that SNF uses different binding mechanisms to bind these two RNAs.

Comparison of FL SNF and RRM1 Binding. In U1A, the first RRM is sufficient for full RNA binding. The second RRM does not bind RNA and does not contribute to the binding affinity of the protein−RNA interaction. We conducted binding experiments with SNF RRM1 to see if this domain was sufficient for full RNA binding. Like U1A, SNF RRM2 does not bind either dSLII or dSLIV on its own (data not shown). Unlike U1A, loss of the linker and RRM2 of SNF is significantly detrimental to SNF's ability to bind both its targets. Table 5 gives binding affinities (K_{obs}) and free energies ($\Delta G°$) for full-length SNF and RRM1 of SNF and U1A binding to their respective RNAs. In 250 mM KCl, SNF truncation results in a 30-fold decrease in binding affinity for dSLII and a 3-fold decrease in binding affinity for dSLIV. Analysis of binding isotherms monitored by fluorescence enhancement suggests a possible explanation of the effect of truncation. Figure 8a shows representative binding curves for FL SNF and RRM1 binding to dSLIV. In these experiments, 6-carboxyfluorescein (6-FAM) was attached to the 5′ end of the RNA, and protein binding was observed by fluorescence enhancement. Affinities are equivalent to those obtained from nitrocellulose filter binding experiments under similar conditions.

As previously noted, the binding isotherms show that protein truncation results in a loss of affinity for SLIV. However, just as striking is the fact that fluorescence enhancement upon binding is greatly decreased when SNF is truncated to RRM1 alone (Figure 8a). For titrations conducted in 100 mM KCl, when the

Table 5: Comparison of FL SNF versus RRM1 Binding Free Energies[a]

RNA	$K_{obs,FL}$ (M)	$\Delta G°_{FL}$ (kcal/mol)	$K_{obs,RRM1}$ (M)	$\Delta G°_{RRM1}$ (kcal/mol)	$K_{obs,RRM1}/K_{obs,FL}$
U1A−hSLII	$(4 \pm 3) \times 10^{-10}$	-13 ± 0.4	$(4 \pm 3) \times 10^{-10}$	-13 ± 0.4	1
SNF−dSLII	$(7 \pm 6) \times 10^{-10}$	-12 ± 0.5	$(2.18 \pm 0.16) \times 10^{-8}$	-10.3 ± 0.1	30
SNF−dSLIV	$(7.7 \pm 3.0) \times 10^{-8}$	-9.6 ± 0.2	$(2.33 \pm 0.66) \times 10^{-7}$[b]	-8.95 ± 0.17	3
SNF−dSLIV with 100 mM KCl	$(5.6 \pm 1.6) \times 10^{-9}$	-11 ± 0.17	$(3.33 \pm 0.35) \times 10^{-8}$[c]	-10.1 ± 0.1	5.9

[a]Experiments were conducted in 250 mM KCl (unless noted), 10 mM cacodylate, and 1 mM $MgCl_2$ (pH 7). [b]Determined from fluorescence experiments rather than filter binding experiments. Unless noted, errors were determined as the standard deviation of at least two data sets. This generally gave a higher estimation of uncertainty than what was achieved with error propagation analysis. [c]The error was determined via the propagation of uncertainty.

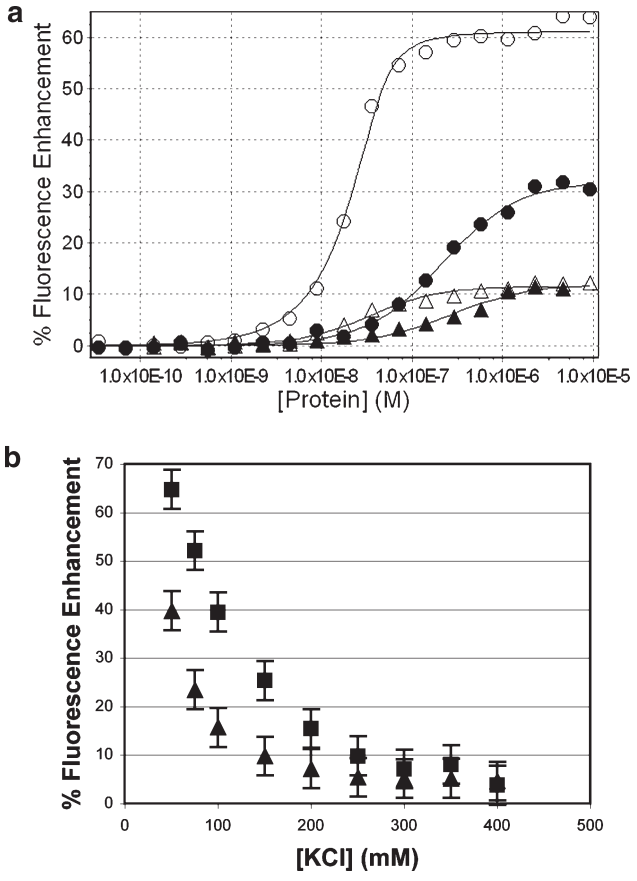

FIGURE 8: (a) Representative binding isotherms for full-length SNF (● and ○) and RRM1 (▲ and △) binding to dSLIV at 22 °C in 10 mM potassium phosphate and 1 mM $MgCl_2$ (pH 8). Empty symbols are data points for experiments conducted in 100 mM KCl. Filled symbols are data points for experiments conducted in 250 mM KCl. Lines represent best fits for the individual data sets. (b) Salt titrations of maximal fluorescence enhancement of FAM-dSLIV upon SNF binding. Fluorescence enhancement was measured by titrating the RNA alone, as well as the RNA and either 10 μM FL SNF (●) or 10 μM RRM1 (▲) with KCl. Titrations were performed in 10 mM potassium phosphate and 1 mM $MgCl_2$ (pH 8).

RNA is fully protein-bound, there is an approximately 85% loss of fluorescence enhancement using RRM1. For FL SNF, fluorescence enhancement upon binding is clearly salt-dependent; when the salt concentration is increased to 250 mM, the fluorescence enhancement is greatly reduced. A salt titration was performed to assess the effect of salt upon fluorescence enhancement (Figure 8b). Experiments were performed at constant protein and RNA concentrations of 10 μM and 10 nM, respectively. At this protein concentration, SLIV is fully bound to FL SNF at all salt concentrations tested and SLIV is fully bound to RRM1 up to at least 300 mM KCl (data not shown). The fluorescence enhancement of the RNA upon protein binding is shifted to lower salt concentrations for RRM1 as compared to FL SNF, but the shape of the salt dependence is similar. Given that the dye is attached to the 5′ end of the RNA stem, these data suggest that the linker or RRM2 interacts with the RNA stem at low salt concentrations, enhancing the fluorescence of the dye. Alternatively, the linker or RRM2 may be interacting with RRM1 to facilitate specific interactions between RRM1 and the RNA stem. The difference in RNA binding affinities for FL SNF compared with RRM1 underscores a major mechanistic difference in binding when SNF is compared with U1A. Although the binding affinities of FL SNF and U1A for SLII are quite similar, SNF RRM1 is not able to achieve a high binding affinity for both of its targets without compensating interactions from other parts of the protein.

DISCUSSION

Structural and Thermodynamic Properties of SNF. The far-UV CD spectra of the two RRM domains of SNF are almost identical to those of U1A, indicating that both proteins have similar secondary structure. Furthermore, the unfolding curves of both domains fit well to a two-state model, characteristic of many globular proteins. In spite of these similarities, the thermodynamic stability of the first and second domains is dramatically reduced in SNF; RRM2 is destabilized by 3.5 kcal/mol and RRM1 by 6 kcal/mol when compared with U1A. This suggests there was an evolutionary capacity for SNF to be more stable than it is. The marginal stability of SNF RRM1 is therefore surprising, but it is not unique. The folded state of globular proteins is generally only slightly stabilized compared to the unfolded state (folding free energies typically range from 5 to 10 kcal/mol) (*22*). Much more stable proteins have been engineered, suggesting that there is little evolutionary pressure for highly stable proteins and that perhaps the converse is true, that marginal thermodynamic stability is in fact a biologically important characteristic of proteins. However, the relationship between protein instability and protein function has not been well-defined. The case of SNF's instability is particularly dramatic, and it is tempting to speculate that SNF's marginal stability is important for its ability to bind two different targets.

An unresolved structural question is the presence or absence of a third α-helix at the C-terminus of SNF RRM1. In U1A, the C-terminal tail and its helix contribute directly and indirectly to substrate discrimination and specificity (*18*, *23*). This helix is not a feature of canonical RRMs, and although the U1A sequence is similar to that of SNF, it is not identical. The sensitivity of far-UV CD to helical content renders CD suitable for addressing whether the third helix is present. The similarity between the CD spectra of RRM1 of U1A and SNF suggests that this helix does exist in SNF.

SNF−RNA Interactions. A comparison of the binding partners of SNF and U1A reveals that the differences between

Drosophila and human SLII are minimal. Both loops close with a C·G base pair, and the single-stranded AUUGCAC loop RNA sequence recognized in the U1A−hSLII complex (*13*, *23*, *24*) is identical between the human and *Drosophila* RNAs. In the U1A−SLII complex, the three remaining residues at the 3′ end of the RNA loop extend away from the protein surface and do not make any contacts with the protein (*23*). They are important in that they act as a spacer and therefore must not form base pairs; otherwise, the sequence of these residues is unimportant (*25*). Individual mutations of the two discrepant nucleotides to their *Drosophila* counterparts result in no change in binding affinity for U1A (*24*).

To improve our understanding of the nature of RNA−protein interactions, it would be useful to compare the thermodynamics of U1A and SNF binding to their respective RNA targets under similar conditions. Unfortunately, we were not able to measure the binding of SNF to dSLII at 250 mM KCl, as it bound too tightly for accurate determination of the K_{obs}, especially at low temperatures. The comparison between SNF binding dSLII and U1A binding hSLII is therefore made with some reservations. At 250 mM NaCl, the van't Hoff plot of U1A binding hSLII clearly deviates from linearity and shows an apparent ΔC_p. Despite the similarity of both proteins and SLII RNAs, SNF binding to dSLII is linear at 300 mM KCl; there is no $\Delta C_{p,obs}$. This indicates that binding occurs through different mechanisms in the two proteins. At high salt concentrations, there is also no observed heat capacity change for the interaction of SNF with dSLIV. It is surprising that binding to both RNAs occurs without an observed heat capacity change, as we expect that the protein−RNA interaction should take place in the context of coupled conformational changes. In the U1A−hSLII interaction, conformational transitions occur in both the RNA and the protein. In the RNA, the loop must be splayed open by part of the protein. Along with the disruption of stacking interactions throughout the loop, this is likely the basis of changes in the RNA loop structure that occur upon protein binding and contribute to the $\Delta C_{p,obs}$ (*13*, *26*). Given that dSLII and hSLII are essentially identical, the disparity in $\Delta C_{p,obs}$ is unlikely to result from differences intrinsic to the RNA. Additionally, the U1A−hSLII interaction results in changes to the protein structure. Notably, loop 3 becomes confined through its interactions with the RNA. While the exact nature of the SNF−RNA interface and hydrogen bonding network is unknown, it is likely that binding to either RNA target results in conformational changes in both the RNA and the protein. It seems unlikely that the RNA−protein interaction would result in significant hydrophobic surface burial of the human protein but not of SNF. It also seems unlikely that the SNF−dSLII interaction occurs through a ligand docking mechanism, but such a mechanism is consistent with the results. It is possible that changes in the protein and RNA compensate for each other so that there is no $\Delta C_{p,obs}$.

A large $\Delta C_{p,obs}$ characterized the SNF−dSLIV interaction at lower salt concentrations. The strong salt dependence of $\Delta C_{p,obs}$ for dSLIV was surprising. This apparent salt dependence could result from a shift in the van't Hoff plot such that over the temperature range measured, the curve appears linear when in fact it is not. This could also account for the apparent linearity of the van't Hoff plot observed for SNF binding to dSLII. Alternatively, the salt dependence of the $\Delta C_{p,obs}$ for the SNF−dSLIV interaction may reflect a large electrostatic contribution to the observed ΔC_p. It is possible that in this system, the linker between RRM1 and RRM2 interacts with the RNA. The linker contains a high density of positive charges that could contribute to a strong electrostatic interaction, and we find that removing the linker and RRM2 results in a significant loss of binding affinity. For SLIV, we can compare the loss of binding affinity at 100 and 250 mM KCl, and we find that at the lower salt concentration there is a 6-fold loss of binding affinity upon truncation whereas at 250 mM KCl there is an only 3-fold loss of affinity. Truncation also leads to a loss of discrimination between RNAs: $\Delta\Delta G^\circ = \Delta G^\circ_{(dSLII)} - \Delta G^\circ_{(dSLIV)} = -2$ kcal/mol for SNF but only −0.6 kcal/mol for RRM1. Fluorescence enhancement upon binding is also diminished with protein truncation. Fluorescence enhancement upon binding is highly salt-dependent for both FL SNF and RRM1, but fluorescence enhancement decreases at much lower salt concentrations for RRM1. We suggest that the highly charged linker binds the phosphate backbone of the RNA stem, contributing to an overall increase in binding affinity. Furthermore, we propose that these additional interactions are necessary to compensate for the compromised binding that SNF RRM1 exhibits for dSLII as compared with the affinity that U1A RRM1 has for hSLII.

One of the most striking differences between U1A and SNF is that while U1A does not detectably bind hSLIV, SNF binds dSLIV with nanomolar affinity. At 300 mM KCl, van't Hoff plots for SNF binding both dSLII and dSLIV are linear, which makes it possible to compare the entropic and enthalpic contributions to binding. Binding of the two RNAs is distinguished by $\Delta\Delta S^\circ = \Delta S^\circ_{SNF-dSLII} - \Delta S^\circ_{SNF-dSLIV} = -14.1 \pm 9.5$ cal mol^{-1} K^{-1} ($T\Delta\Delta S^\circ$ is −4.2 kcal/mol at 298 K), whereas $\Delta\Delta H^\circ = \Delta H^\circ_{SNF-dSLII} - \Delta H^\circ_{SNF-dSLIV} = -6.7 \pm 2.8$ kcal/mol. While binding of both RNA targets is enthalpically driven, the additional binding affinity of SNF for SLII versus SLIV appears to be completely enthalpic in origin. While there is a considerable entropic penalty for binding both RNAs, at 22 °C the penalty associated with binding SLII is larger than that of binding SLIV by 4 kcal/mol. More ions are released upon SLII binding than upon SLIV binding, so the difference in the entropic penalty cannot be attributed to differential ion release.

Enthalpic contributions to binding, including hydrogen bonding, salt bridges, and stacking interactions between bases and amino acid side chains, appear to be more favorable for the SNF−SLII interaction, but more interesting is the fact that enthalpic and entropic contributions to binding appear to favor different RNAs. This presents intriguing possibilities for tweaking the binding system. In biological contexts, SNF binds at least one other protein partner, U2A′. Although it is unclear how U2A′ alters SNF's interactions with RNA, studies of the mammalian U2A′−U2B″ interaction suggest that U2A′ alters the specificity of U2B″ for its RNA target. Shifting the balance between entropic and enthalpic effects may well provide a mechanism for driving specificity, and this may occur in the context of protein modulators.

In both *Drosophila* and humans, U1 SLII and U2 SLIV snRNAs are extremely similar. Nevertheless, U1A is able to bind SLII but not SLIV, raising the question of how the protein achieves specificity for a single target. In humans, there are three differences between SLII and SLIV. The first is that in SLIV, there is a cytosine to guanine substitution at the seventh loop nucleotide. This single mutation in SLII decreases the binding affinity for U1A 10-fold (*24*). The second difference is that SLIV has an 11-nucleotide loop rather than a 10-nucleotide loop, with an extra adenine inserted between nucleotides 8 and 9. This region of SLIV appears to interact with U2B″, but the corresponding region of SLII does not interact with U1A. Although it

has been shown that a nucleotide insertion at this site in SLII has no effect on protein binding (*25*), it is not clear whether this is true for a specific adenosine insertion. The adenine at this site has the potential to form a base pair across the SLIV loop with the third uridine, an effect that may compromise binding to the protein if the base pair alters the conformational equilibrium of the loop such that protein binding becomes unfavorable. The last difference between the two stem−loop is the loop-closing base pair. In SLII, this base pair is a Watson−Crick C·G pair, whereas hSLIV contains a noncanonical U·U pair at this position. We propose that this last feature is the critical determinant for SNF−dSLIV binding: although other features are shared between hSLIV and dSLIV, the loop-closing base pair in dSLIV is a U·G wobble pair. SNF is capable of binding dSLIV with quite high affinity, but it is compromised in binding hSLIV. The binding data, therefore, suggest that in *Drosophila*, both the RNAs and protein have adapted in the course of evolutionary history so that a single protein can bind RNAs that are more commonly bound by two distinct proteins.

Evolution of snrpA/snrpB2 Proteins, Their RNAs, and Implications for Multiple Binding Mechanisms. Phylogenetic analysis of *snrpA/snrpB2* proteins shows that in eukaryotes, most organisms contain two distinct proteins that recognize two distinct, conserved RNAs (U1 SLII and U2 SLIV). In different phyla, the two proteins share more in common with each other than they do with respective proteins from other phyla. Although it is possible that this situation arose through numerous gene duplications throughout evolutionary history, it is more likely that there was a single gene duplication event that occurred in a common ancestor of higher eukaryotes and that U1A and U2B″ have been coevolving.

At the time this alternative was proposed, it was favored solely on the basis of its parsimony, and the authors speculated that as yet unidentified protein−protein interactions were providing the evolutionary pressure for U1A and U2B″ to coevolve (*15*); that is, an unidentified protein was interacting with both U1A and U2B″, thus forcing the two proteins to coevolve with this third protein. More recently, experiments with *C. elegans* U1A and U2B″ showed that in the worm, U1A and U2B″ are functionally redundant (*4*). In worms, when both proteins are present, each is segregated to a specific snRNP. However, the absence of either protein results in minimal phenotypic disruption, and the remaining protein can be found in both snRNPs. Deletion of either U1A or U2B″ results in no apparent phenotype, but the deletion of both proteins is embryonic-lethal. The authors conclude that in the worm, there has been selective pressure for the proteins to remain redundant. It is therefore not necessary to invoke unknown protein−protein interactions to account for the coevolution of these two proteins. However, it is still unclear why redundancy of these proteins is important and has been selected in worms.

The extent to which evolution has selected for functional redundancy of U1A and U2B″ proteins is also unclear. Is this a feature of many *snrpA/snrpB2* proteins, or is it specific to the worm? Presumably, if the need for functional redundancy is the only selective pressure guiding coevolution of the two proteins, functional redundancy should be pervasive in this protein family. In apparent contradiction to this, human U1A is not capable of binding hSLIV. However, it is not known whether U2B″ can recognize both SLII and SLIV. It may be that in vertebrates, SLII binding is redundant, whereas SLIV binding is in the unique purview of U2B″. The SNF proteins, in contrast, seem to represent an alternative solution to the problem of RNA target recognition, where selective pressure for adaptations occurs in target RNAs. In organisms containing SNF proteins, there is clearly no requirement for two redundant proteins.

Upon consideration of the variations in proteins from the *snrpA/snrpB2* family, different regions of the proteins, all of which are known to be important for interacting with RNA, seem to have adapted under different evolutionary pressures. A number of protein residues are universally conserved, suggesting that the proteins utilize these residues as part of a shared binding framework. However, this framework seems to allow some flexibility in the RNAs recognized. Most characteristically, this framework permits the recognition of both SLII and SLIV RNAs, although not necessarily by the same protein. In U2B″, the VALKT sequence of loop 3 is highly conserved, and this sequence presumably contacts the 3′ UACC nucleotides of the SLIV loop [as inferred from a cocrystal of U2B″ and SLIV (*8*)]. The UAC nucleotides are conserved in SLIV RNAs from vertebrates, plants, *Drosophila*, *Ciona*, and *Schizosaccharomyces pombe*. The implication is that loop 3−UACC interactions are conserved in the U2B″−SLIV complex and have been selected through the interaction of the protein with the RNA. Conversely, corresponding nucleotides in SLII RNAs are variable, and biochemical experiments show that no interactions occur between human U1A and these 3′ nucleotides. It is curious, however, that *S. pombe* and plant U1A proteins retain the A, K, and T amino acids in this loop 3 sequence, as these residues are unlikely to aid in SLII recognition. It seems more likely that in plants and *S. pombe*, U1A may be able to recognize SLIV as well as SLII and that U1A loop 3 contributes to the functional redundancy of the protein. In vertebrates, none of the corresponding loop 3 residues are shared between U1A and U2B″, which may account, at least in part, for the inability of human U1A to bind U2 SLIV.

The TDS sequence (residues 89−91), RNP1 (especially at the variable position 55), and the second half of the third α-helix have diverged considerably between species but exhibit remarkable similarity between U1A and U2B″ within the same species. These are sites in which the two proteins appear to have coevolved. One explanation for this is that while many of the specific RNA−RRM interactions are conserved among all of these proteins, organisms have adopted a unique set of interactions that is defined by these nonconserved but coevolving sites. This suggests that different organisms have found alternative solutions to snRNA recognition.

While we know much about the *snrpA/snrpB2* protein family and its interactions with RNA, many questions remain unresolved. The most basic of these questions is what role these proteins play in splicing or other cellular processes. Included in this are questions about the function of the proteins' second RRM, how protein−protein interactions regulate RNA binding and protein function, and whether functional redundancy is a common feature of these proteins. The *Drosophila snf* protein provides an opportunity for a more complete understanding of many aspects of these questions, because *Drosophila* is a model organism for genetic studies and the protein itself has many unique features.

SUPPORTING INFORMATION AVAILABLE

U1A and U2B″ phylogenetic trees, as well as protein and RNA sequence alignments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J. 272*, 2118–2131.
2. Will, C. L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W. J., and Luhrmann, R. (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: The U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res. 24*, 4614–4623.
3. Wahl, M. C., Will, C. L., and Luhrmann, R. (2009) The spliceosome: Design principles of a dynamic RNP machine. *Cell 136*, 701–718.
4. Saldi, T., Wilusz, C., MacMorris, M., and Blumenthal, T. (2007) Functional redundancy of worm spliceosomal proteins U1A and U2B″. *Proc. Natl. Acad. Sci. U.S.A. 104*, 9753–9757.
5. Salz, H. K., and Flickinger, T. W. (1996) Both loss-of-function and gain-of-function mutations in snf define a role for snRNP proteins in regulating Sex-lethal pre-mRNA splicing in *Drosophila* development. *Genetics 144*, 95–108.
6. Stitzinger, S. M., Conrad, T. R., Zachlin, A. M., and Salz, H. K. (1999) Functional analysis of SNF, the *Drosophila* U1A/U2B″ homolog: Identification of dispensable and indispensable motifs for both snRNP assembly and function in vivo. *RNA 5*, 1440–1450.
7. Clery, A., Blatter, M., and Allain, F. H. (2008) RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol. 18*, 290–298.
8. 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.
9. Nagengast, A. A., and Salz, H. K. (2001) The *Drosophila* U2 snRNP protein U2A′ has an essential function that is SNF/U2B″ independent. *Nucleic Acids Res. 29*, 3841–3847.
10. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res. 15*, 8783–8798.
11. Stump, W. T., and Hall, K. B. (1993) SP6 RNA polymerase efficiently synthesizes RNA from short double-stranded DNA templates. *Nucleic Acids Res. 21*, 5480–5484.
12. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol. 131*, 266–280.
13. Hall, K. B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry 33*, 10076–10088.
14. Hall, K. B., and Kranz, J. K. (1999) Nitrocellulose filter binding for determination of dissociation constants. *Methods Mol. Biol. 118*, 105–114.
15. Polycarpou-Schwarz, M., Gunderson, S. I., Kandels-Lewis, S., Seraphin, B., and Mattaj, I. W. (1996) *Drosophila* SNF/D25 combines the functions of the two snRNP proteins U1A and U2B′ that are encoded separately in human, potato, and yeast. *RNA 2*, 11–23.
16. Hall, K. B., and Stump, W. T. (1992) Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res. 20*, 4283–4290.
17. Lu, J., and Hall, K. B. (1995) An RBD that does not bind RNA: NMR secondary structure determination and biochemical properties of the C-terminal RNA binding domain from the human U1A protein. *J. Mol. Biol. 247*, 739–752.
18. Zeng, Q., and Hall, K. B. (1997) Contribution of the C-terminal tail of U1A RBD1 to RNA recognition and protein stability. *RNA 3*, 303–314.
19. deHaseth, P. L., Lohman, T. M., and Record, M. T., Jr. (1977) Nonspecific interaction of lac repressor with DNA: An association reaction driven by counterion release. *Biochemistry 16*, 4783–4790.
20. Record, M. T., Jr., Lohman, M. L., and De Haseth, P. (1976) Ion effects on ligand-nucleic acid interactions. *J. Mol. Biol. 107*, 145–158.
21. Lohman, T. M., and Mascotti, D. P. (1992) Thermodynamics of ligand-nucleic acid interactions. *Methods Enzymol. 212*, 400–424.
22. Pace, C. N. (1990) Conformational stability of globular proteins. *Trends Biochem. Sci. 15*, 14–17.
23. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*, 432–438.
24. Stump, W. T., and Hall, K. B. (1995) Crosslinking of an iodo-uridine-RNA hairpin to a single site on the human U1A N-terminal RNA binding domain. *RNA 1*, 55–63.
25. Williams, D. J., and Hall, K. B. (1996) Thermodynamic comparison of the salt dependence of natural RNA hairpins and RNA hairpins with non-nucleotide spacers. *Biochemistry 35*, 14665–14670.
26. Hall, K. B., and Kranz, J. K. (1995) Thermodynamics and mutations in RNA-protein interactions. *Methods Enzymol. 259*, 261–281.