

Washington University School of Medicine Digital Commons@Becker

Open Access Publications

2014

Climbing the vertebrate branch of U1A/U2B^{''} protein evolution

Kimberly J. Delaney

Washington University School of Medicine in St. Louis

Sandra G. Williams

Washington University School of Medicine in St. Louis

Mariah N. Lawler

Washington University School of Medicine in St. Louis

Kathleen B. Hall

Washington University School of Medicine in St. Louis

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation

Delaney, Kimberly J.; Williams, Sandra G.; Lawler, Mariah N.; and Hall, Kathleen B., "Climbing the vertebrate branch of U1A/U2B^{''} protein evolution." *RNA*.20,7. 1035-1045. (2014).
http://digitalcommons.wustl.edu/open_access_pubs/4161

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.

Climbing the vertebrate branch of U1A/U2B'' protein evolution

KIMBERLY J. DELANEY,^{1,2} SANDRA G. WILLIAMS,^{1,2} MARIAH LAWLER,¹ and KATHLEEN B. HALL^{1,3}

¹Department of Biochemistry and Molecular Biophysics, Washington University Medical School, St. Louis, Missouri 63110, USA

ABSTRACT

In the vertebrate lineage of the U1A/U2B''/SNF protein family, the U1A and U2B'' proteins bind to RNA stem-loops in the U1 or U2 snRNPs, respectively. However, their specialization is fairly recent, as they evolved from a single ancestral protein. The progress of their specialization (subfunctionalization) can be monitored by the amino acid sequence changes that give rise to their modern RNA-binding specificity. Using ancestral sequence reconstruction to predict the intermediates on the evolutionary branch, a probable path of sequential changes is defined for U1A and U2B''. The RNA-binding affinity for U1A/U2B'' protein ancestors was measured using modern U1 and U2 snRNA stem-loops and RNA stem-loop variants to understand how the proteins' RNA specificities evolved.

Keywords: RRM; RRM:RNA binding; U1A; ancestral reconstruction; snRNP

INTRODUCTION

The U1A/U2B''/SNF family of spliceosomal proteins is an elegant system for studying the evolution of RNA-binding modes and RNA-binding specificity. In extant organisms, each family member contains two RNA recognition motifs (RRMs) connected by a flexible linker, the first of which (RRM1) interacts with an RNA stem-loop within the U1 and/or U2 snRNP with high affinity and specificity. Human U1A and U2B'' are the proteins belonging to this family in *Homo sapiens*, and they share >74% sequence identity in the RRM1. In spite of this, they exhibit remarkably different RNA-binding activity. In vivo, human U1A segregates to the U1 snRNP, where it binds stem-loop II (SLII), while U2B'' segregates to the U2 snRNP, where it binds stem-loop IV (SLIV) as well as the U2A' protein. In vitro binding assays have shown that U1A binds SLII with extremely high affinity but binds SLIV with weak affinity (Hall and Stump 1992; Stump and Hall 1995). U2B'' binds both stem-loops with modest affinity (Williams and Hall 2011). Clearly, these modern proteins have developed specialized behaviors which impact their ability to segregate to the U1 or U2 snRNP, though the molecular basis for this difference has, as yet, not been fully ascertained.

To examine these proteins within their evolutionary context, we recently undertook ancestral reconstruction of

this family as far back as the last common ancestor of all bilaterians (Williams et al. 2013). This approach analyzes modern protein sequences to infer both the evolutionary relationship between modern-day proteins (phylogeny) and the most likely amino acid sequence of the protein as it evolved, following both speciation and gene duplication events (Harms and Thornton 2013). Ancestral sequence reconstruction, therefore, reveals the historical mutations separating different evolutionary nodes. To understand the molecular evolution of U1A/U2B''/SNF proteins, modern metazoan protein sequences were used to generate the most statistically likely phylogeny of this family. Ancestral protein sequences at each node of the phylogenetic tree were inferred using information in the reconstructed tree and modern protein sequences (Williams et al. 2013). Surprisingly, this analysis led to the realization that the gene duplication that gave rise to modern U1A and U2B'' occurred in an ancestor of jawed vertebrates (Fig. 1), which is much more recent than the former hypothesis that the gene duplication occurred prior to the divergence of the eukaryotic kingdoms (Polycarpou-Schwarz et al. 1996; Saldi et al. 2007). Consistent with this finding, most metazoans have a single SNF protein.

Through thorough mutational analysis, ancestral sequence reconstruction enables the determination of which

²These authors contributed equally to this work.

³Corresponding author

E-mail kathleenhal@gmail.com

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.044255.114>.

© 2014 Delaney et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

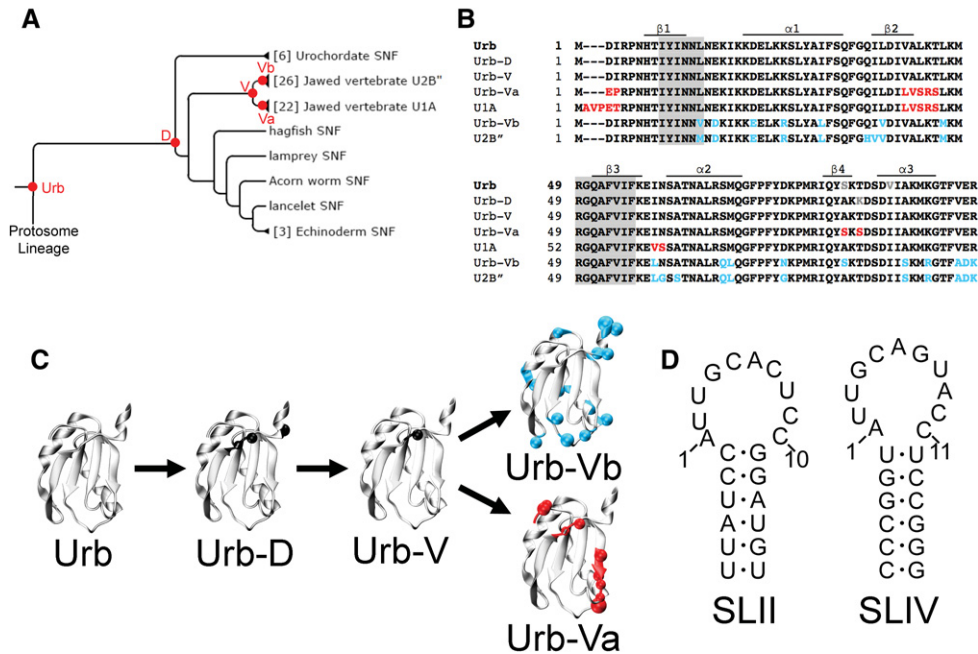


FIGURE 1. The deuterostome branch of the U1A/U2B''/SNF phylogenetic tree. (A) The deuterostome subset from the original phylogenetic tree (Williams et al. 2013). Red circles indicate nodes that were resurrected for biochemical experiments. (B) Alignments of the maximum-likelihood sequences of each resurrected protein as well as modern U1A and U2B''. Amino acids in gray indicate variation from Urb-V prior to the gene duplication. Amino acids in red indicate variation from Urb-V in the Urb-Va lineage. Amino acids in cyan indicate variation from Urb-V in the Urb-Vb lineage. (C) Amino acid divergence is plotted on the RRM structure. Colored amino acids indicate change from immediate predecessor. (D) Sequences of modern human stem-loop II and stem-loop IV.

mutations were ancillary and which were functionally significant between two historically important nodes. In order to understand the molecular mechanisms that gave rise to modern protein RNA-binding behaviors, we have examined the proteins immediately preceding and following this vertebrate gene duplication and compared their RNA-binding activities to their Urbilateralian ancestor (Urb) as well as to modern U1A and U2B''. Urb-V is the protein at the node immediately preceding the duplication; Urb-Va and Urb-Vb are the resultant nodes of the duplication that gave rise to U1A and U2B'', respectively. We found that Urb-V RNA-binding activity is indistinguishable from its predecessors as far back as the Urbilateralian ancestor. However, following the gene duplication, the RNA-binding activity of Urb-Va and Urb-Vb changed very rapidly to acquire modern U1A and U2B'' characteristics.

We have used modern SLII and SLIV RNAs to compare binding affinities of the RRM1 and stem-loop variants to probe their RNA-binding specificity. Our results show that five mutations within β2/Loop 3 of RRM1 are sufficient to increase the specificity of Urb-Va for SLII, which is accomplished via a dramatic decrease in affinity for SLIV and a small increase in affinity for SLII. In contrast, many more mutations occurred between Urb-V and Urb-Vb and were distributed throughout the body of the RRM. We find that Urb-Va and Urb-Vb have distinct binding mechanisms with the SLII or SLIV RNA hairpins; in contrast, Urb-V ap-

pears to be capable of both interchanging between Urb-Va-like behaviors and Urb-Vb-like behaviors. We conclude that the gene duplication allowed each protein (Urb-Va and Urb-Vb) to specialize to a single binding mode that is optimal for its in vivo functions.

RESULTS

The reconstruction of the deuterostome branch of U1A/U2B''/SNF phylogeny is shown in Figure 1A. This is a subtree of the larger U1A/U2B''/SNF phylogeny that was previously reported (Williams et al. 2013). Proteins at each significant node are designated as follows: Urb-D is the last common ancestral protein for modern deuterostomes; Urb-V for jawed vertebrate proteins (preceding the gene duplication); Urb-Va for the last common ancestral protein of modern U1A proteins; and Urb-Vb for U2B'' proteins. Figure 1B shows a sequence alignment of the first RRM of these proteins, as well as alignments of the corresponding sequences of human U1A and U2B''. Figure 1C shows the location of the evolving residues plotted on the RRM. Urb-Va has an ambiguously reconstructed amino acid in Loop 3 and β2, but many of the ambiguities are at the N- and C-termini of the RRM (see Supplemental Table 1).

RRM1 evolved very little between Urb, Urb-D, and Urb-V; the amino acid sequences of the three proteins are very similar. Following the gene duplication, the Urb-Va/U1A branch

accumulated several changes within the N-terminal tail, one mutation in the $\beta 4/\alpha 3$ junction at a relatively nonconserved position, and five simultaneous changes within $\beta 2$ /Loop 3. These five changes in $\beta 2$ /Loop 3 are preserved from reconstructed Urb-Va to modern U1A. Only two additional mutations (one in Loop 4 and one in $\beta 4$) distinguish Urb-Va from U1A RRM1. Many more mutations were introduced in the Urb-Vb/U2B'' branch between the Urb-V and Urb-Vb nodes than in the U1A branch. These include mutations throughout the RRM (Fig. 1C). On the U2B'' branch, seven additional mutations throughout the domain were accumulated between Urb-Vb and human U2B'' RRM1.

From our deuterostome phylogeny, we selected RRMs at critical nodes (Urb, Urb-D, Urb-V, Urb-Va, and Urb-Vb) for further functional characterization. We used the modern human U1 SLII and U2 SLIV RNA hairpins to assess the RNA-binding properties of these proteins, as these RNAs have evolved very little and are considered reasonable surrogates for their ancestral counterparts (Fig. 2). Results of the binding experiments are shown in Table 1. Results for human U1A RRM1 and full-length human U2B'' are shown for comparison (these values were previously published [Hall and Stump 1992; Williams and Hall 2011]). Urb, Urb-D, and Urb-V show almost identical binding to both SLII and SLIV. In contrast, Urb-Va has gained affinity for SLII and lost affinity for SLIV. Urb-Vb has lost affinity for both SLII and SLIV and does not appear to discriminate between the two RNAs. Comparing Urb-Va with U1A RRM1 and Urb-Vb with (FL) U2B'', we find that the ancestral RRMs and their modern orthologs have very similar RNA-binding specificities for their *in vivo* RNA targets.

Source of changes to binding affinity in the U1A branch

Prior *in vitro* experiments with U1A/U2B'' chimeric RRMs showed that $\beta 2$ /Loop 3 (LVSRS LKMRG₅₃ in U1A; VALKTMKMRG₅₀ in U2B'') was important for determining RNA-binding specificity (Scherly et al. 1990a; Laird-Offringa and Belasco 1995; Katsamba et al. 2002). In our reconstruction, $\beta 2$ /Loop 3 (VALKTMKMRG₅₀) is unchanged in ancestors of U1A and U2B'' prior to the vertebrate (gnathostome) gene duplication. Between the Urb-V and Urb-Va nodes, $\beta 2$ /Loop 3 accumulates five mutations, changing VALKT to LVSRS. In modern proteins, U1A con-

tains the LVSRS sequence, and U2B'' has VALKT; the positions of these amino acids in the structures of the respective RNA:protein complexes are shown in Figure 2. It is important to appreciate that these amino acids do not directly participate in RNA recognition, so any effect they have on RNA binding must be a consequence of changes to the RRM.

While the individual mutations VALKT to LVSRS are relatively conservative, the appearance of the mutations in this branch suggests that they confer new functionality. We have systematically altered VALKT/LVSRS sequences to evaluate the amino acid contributions to RNA binding.

Our most substantial alteration was the reversion of $\beta 2$ /Loop 3 in Urb-Va to the sequence of Urb-V (prior to the gene duplication) to create the Urb-Va-VALKT RRM. This protein differs from Urb-V at sites in the N terminus and in the $\beta 4/\alpha 3$ junction. Urb-Va-VALKT does not discriminate between SLII and SLIV and binds both with an affinity comparable to that of Urb-V (Table 2). The $\beta 2$ /Loop 3 sequence in the Urb-Va protein is, therefore, sufficient to revert the binding specificity of the RRM to that of its immediate ancestor.

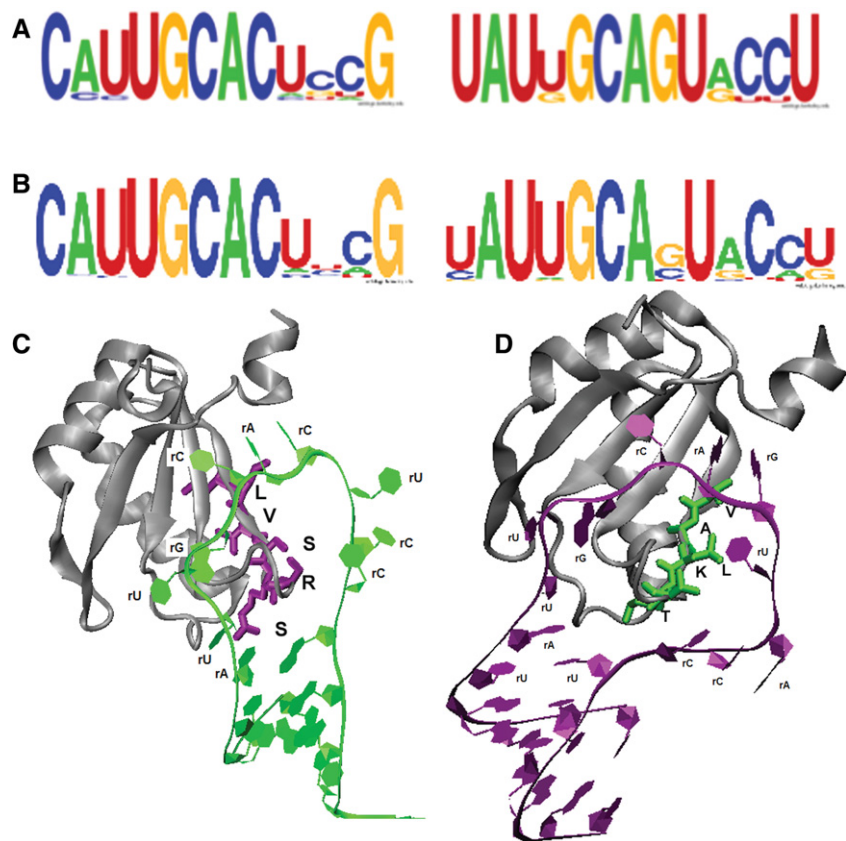


FIGURE 2. RNA sequence conservation. (A) Vertebrate SLII (left) and SLIV (right) sequence logos-loop and loop-closing base pair. (B) Metazoan SLII (left) and SLIV (right) sequence logos-loop and loop-closing base pair, as reported in Williams et al. (2013). (C) Cocystal of U1A RRM1:SLII (1URN) and (D) U2B'' RRM1 and SLIV from the cocystal of the ternary complex (1A9N). LVSRS amino acids are shown in U1A, and VALKT in U2B''. Graphics using VMD (Humphrey et al. 1996).

TABLE 1. RNA-binding affinities for wild-type RNA stem-loops

Protein	SLII (M)	SLIV (M)	$\Delta\Delta G^\circ$ (kcal/mol)
Urb	$1.2 \pm 0.2 \times 10^{-9}$	$1.5 \pm 0.2 \times 10^{-8}$	-1.5 ± 0.1
Urb-D	$1.5 \pm 0.2 \times 10^{-9}$	$2.6 \pm 0.4 \times 10^{-8}$	-1.7 ± 0.1
Urb-V	$1.2 \pm 0.1 \times 10^{-9}$	$2.5 \pm 0.4 \times 10^{-8}$	-1.8 ± 0.1
Urb-Va	$4.4 \pm 0.2 \times 10^{-10}$	$2.8 \pm 0.04 \times 10^{-7}$	-3.8 ± 0.03
hU1A RRM1	$4 \pm 3 \times 10^{-10}$	$>1 \times 10^{-6}$	<-4.6
Urb-Vb	$5.0 \pm 0.2 \times 10^{-7}$	$3.8 \pm 1.8 \times 10^{-7}$	0.15 ± 0.28
hU2B'' FL	$3.3 \pm 1 \times 10^{-8}$	$2.6 \pm 0.6 \times 10^{-8}$	0.14 ± 0.22

Dissociation constants for binding to WT SLII or WT SLIV are shown. Data for hU1A RRM1 and hU2B'' FL were previously reported (Hall and Stump 1992; Williams and Hall 2011). $\Delta\Delta G^\circ = \Delta G^\circ$ (SLII) $- \Delta G^\circ$ (SLIV). All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

Changing five amino acids simultaneously masks the relative contribution of each residue to RNA specificity. Point mutations that individually change the residues of Urb-Va-VALKT to the Urb-Va LVRSRS residues have interesting and distinct effects on RNA binding. Results for these experiments are shown in Table 2. Some individual substitutions result in modest increases in affinity for SLII and modest decreases in affinity for SLIV. However, the two single substitutions at the N terminus of the VALKT sequence (either V to L or A to V) result in changes to RNA binding that almost completely mimic those seen in the protein with the full loop substitution. These individual mutations particularly decrease the RRM's affinity for SLIV. In contrast, the largest increases in affinity for SLII result from substitutions in the middle of the sequence (L to S or K to R). Thus, different sites in the loop are implicated in the recognition of SLII vs. SLIV. As quantified by calculations of the difference in free energies of binding, $\Delta\Delta G^\circ$, the contribution from each site to RNA binding is not additive, indicative of a complicated network of interactions throughout the loop that is coupled to the binding of both RNAs.

The corresponding evolution in the U2B'' branch is far more complex. Preliminary results (data not shown) suggest that mutations throughout the body of the RRM (and not just those restricted to Loop 3) have consequences for RNA binding. These evolutionary substitutions are supported by previous data (Scherly et al. 1990b, 1991) that also show RNA binding can be altered by mutations that are presumably outside the U2B'' RNA-binding surface. The underlying mechanisms by which such mutations perturb RNA binding can be relatively simple (e.g., disrupting the RRM fold) or complicated (e.g., selec-

tively enhancing protein flexibility that becomes detrimental to RNA binding). Evaluating evolutionary changes in the U2B'' branch, given the positions of the mutations, requires a comprehensive study that is beyond the scope of this work, so we restrict our analysis to the evolution of the U1A branch.

RNA binding is salt-dependent

Protein binding to nucleic acids typically has an electrostatic component to the association. The net contribution of the electrostatics to the binding free energy is unique to each complex, but the salt dependence of the interaction can be used to determine if ions are taken up or released and so provides a means to compare binding modes of the interactions. Electrostatic interactions are known to play a key role in the interaction of U1A and SLII (Hall and Stump 1992; Law et al. 2006a), and so we examined the dissociation constant of Urb proteins binding to SLII and SLIV as a function of $[K^+]$. These data are plotted as $\log(K_{D, app})$ vs. $\log [KCl]$, using the formalism of Record et al. (1976). A positive slope indicates ion uptake; a negative slope indicates net ions released. We measured binding of each protein to SLII and SLIV over a range of KCl concentrations (Fig. 3), and as expected, the negative slope of the salt dependence curve indicates that a net number of ions are released upon binding. However, we find that each protein:RNA interaction is unique: Urb-V binding to SLII/SLIV releases 3.3/4.2 ions; Urb-Va releases 5.4/2.1; and Urb-Vb releases 2.6/5.9 ions. This diversity of response indicates a significant difference in how each protein interacts with each target RNA.

RNA mutagenesis

Stem-loops II and IV are both highly conserved throughout metazoans (Fig. 2) and are remarkably similar in size and structure (Fig. 1D). In the vertebrate lineage, and specifically

TABLE 2. RNA-binding affinities of Urb-Va loop 3 mutants for wild-type RNA stem-loops

Protein	SLII		SLIV	
	K_d (M)	$\Delta\Delta G^\circ$ (kcal/mol)	K_d (M)	$\Delta\Delta G^\circ$ (kcal/mol)
Urb-V	$1.2 \pm 0.1 \times 10^{-9}$		$2.5 \pm 0.4 \times 10^{-8}$	
Urb-Va-VALKT	$1.0 \pm 0.1 \times 10^{-9}$		$1.4 \pm 0.1 \times 10^{-8}$	
Urb-Va-LALKT	$5.3 \pm 0.5 \times 10^{-10}$	-0.4 ± 0.1	$1.6 \pm 0.1 \times 10^{-7}$	1.4 ± 0.1
Urb-Va-VVLKT	$1.7 \pm 0.1 \times 10^{-9}$	0.3 ± 0.1	$1.5 \pm 0.2 \times 10^{-7}$	1.4 ± 0.1
Urb-Va-VASKT	$2.6 \pm 0.1 \times 10^{-10}$	-0.8 ± 0.1	$2.9 \pm 0.2 \times 10^{-8}$	0.4 ± 0.1
Urb-Va-VALRT	$2.1 \pm 0.2 \times 10^{-10}$	-0.9 ± 0.1	$1.1 \pm 0.1 \times 10^{-8}$	-0.2 ± 0.1
Urb-Va-VALKS	$5.8 \pm 0.8 \times 10^{-10}$	-0.3 ± 0.1	$5.2 \pm 0.4 \times 10^{-8}$	0.8 ± 0.1
Sum-point mutations		-2.1 ± 0.2		3.8 ± 0.1
Urb-Va(LVRSRS)	$4.4 \pm 0.2 \times 10^{-10}$	-0.5 ± 0.1	$2.75 \pm 0.04 \times 10^{-7}$	1.74 ± 0.03

Dissociation constants for binding to WT SLII or WT SLIV. $\Delta\Delta G^\circ = \Delta G^\circ$ (Mut Prot) $- \Delta G^\circ$ (Urb-V). All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

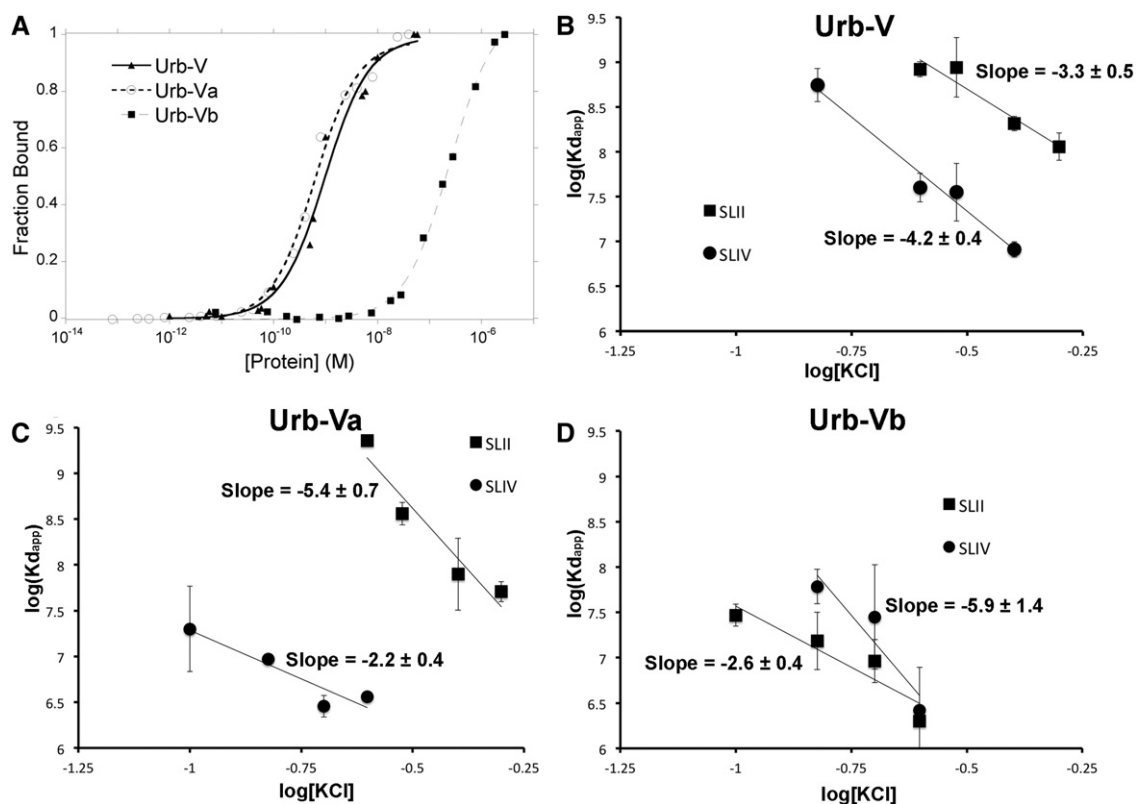


FIGURE 3. Salt dependence of binding to wild-type RNA stem-loops. (A) Binding isotherm data and fits to a 1:1 complex; 250 mM KCl, 1 mM MgCl₂, 10 mM sodium cacodylate, pH 7. (B–D) Protein-RNA pairs are indicated in the panels. (■) SLII, (●) SLIV. All experiments were performed in 10 mM cacodylate and 1 mM MgCl₂ at room temperature with indicated salt. Slopes of the lines are interpreted in terms of net ions released and are reported adjacent to each line.

in humans, there are three main differences between SLII and SLIV: (1) Loop position 7 is cytosine in SLII but guanosine in SLIV; (2) the loop-closing base pair is a C:G in SLII and a noncanonical U:U in SLIV; and (3) the 3' loop nucleotides (UCC) in SLII RNA are poorly conserved, but in SLIV the analogous UACC is conserved. SLII has 10 loop nucleotides, while SLIV has 11. The high degree of sequence conservation in these large RNA loops is likely to be predominantly driven by protein recognition. However, the ancestral RRMs clearly show differences in RNA-binding affinity and specificity, which can account for much of the specific protein localization seen in modern snRNPs. A series of RNA mutations were made to probe the source of differences in the RNA-binding specificities of Urb-V, Urb-Va, and Urb-Vb.

Loop size and structure

Extensive work with U1A has previously shown that the secondary structure of its RNA target is important for high-affinity binding: the recognition sequence must occur within the context of a stem-loop structure, and the optimal loop size is at least 10 nt (Williams and Hall 1996; Law et al. 2006b). We used RNA variants to address secondary structure requirements and optimal loop size for Urb-V, Urb-

Va, and Urb-Vb. The ssLoop RNA construct puts the loop sequence of SLII in a completely single-stranded RNA context (secondary structure predictions via *mfold* [Zuker 2003]). As shown in Figure 4, all three proteins experience a significant loss of binding affinity in the absence of a stem-loop: binding of both Urb-Va and Urb-Vb is weaker than the sensitivity of the filter binding assay. This is unsurprising as crystal structures of U1A and U2B' RRM1 in complex with their RNA targets indicate interactions between Loop 3 of the protein and the loop-closing base pair of the stem (Oubridge et al. 1994; Price et al. 1998). It appears that the ancestral proteins share a similar requirement for the presence of a stem.

We probed the loop size requirement by deleting one (9-nt loop) or two (8-nt loop) nucleotides from the 3' side of the loop. Previous work with U1A showed that the protein does not specifically recognize these nucleotides in the U1A:SLII complex (Stump and Hall 1995; Williams and Hall 1996), and in the U1A:SLII cocrystal, the UCC bases have no protein contacts (Oubridge et al. 1994). Consistent with the U1A results, all three proteins show a loss of binding affinity to the 9-nt loop RNA and an even greater loss for the 8-nt loop RNA (Fig. 4). This result is supported by cocrystal structures of modern protein-RNA complexes (Oubridge et al. 1994; Price et al. 1998), which show that Loop 3 of

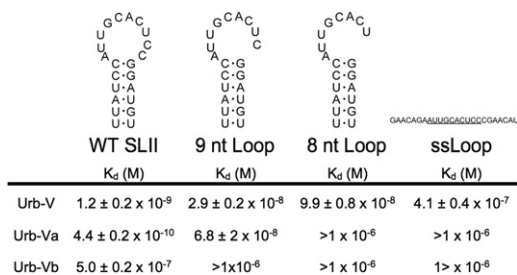


FIGURE 4. Binding to loop size and structure mutants. Dissociation constants for binding to each RNA are shown. All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

the protein protrudes through the RNA loop; the loop size must be large enough to accommodate the insertion of the protein and position the central nucleotides near the β4/α3 junction.

Loop position 1

Both biochemical and structural data have shown that the adenosine in loop position 1 (A1) is specifically recognized by U1A RRM (Hall 1994; Oubridge et al. 1994; McConnell et al. 2003; Benitex and Baranger 2007). A1 is conserved in both SLII and SLIV. To investigate the importance of A1 in the ancestral complexes, we replaced it with cytosine (A1C) or deleted it (ΔA1, with a 3' C insertion to maintain a 10-nt loop) in SLII. All three proteins exhibit decreased affinity for these RNAs (Fig. 5), consistent with A1 conservation in SLII. However, it is clear that perturbation of this nucleobase impedes Urb-Va binding significantly more than it disturbs Urb-Vb binding.

Urb-V binding affinities for A1C and ΔA1 SLII indicated some dependence of this loop position (Fig. 5). Binding free energies of both SLII mutant RNAs to Urb-Va were significantly perturbed ($\Delta\Delta G^\circ = (\Delta G^\circ_{\text{MUT}} - \Delta G^\circ_{\text{WT}}) = +2.3$ – 2.4 kcal/mol), but Urb-Vb binding was nearly unchanged ($\Delta\Delta G^\circ \sim +0.3$ – 0.6 kcal/mol), indicating that this protein is not very sensitive to this position. In fact, deletion of A1 resulted in less disruption of Urb-Vb:SLII binding than a mutation, suggesting it has no need for this nucleobase. This constitutes a major difference in the RNA recognition mechanisms of Urb-Va and Urb-Vb. While these results provide a rationale for the phylogenetic conservation of A1 in SLII, they do not explain the conservation of A1 in vertebrate U2 SLIV snRNAs.

SLIV 3' side

Previous mutational analysis and structural work with U1A showed that there is no sequence recognition of the 3' side of SLII (Oubridge et al. 1994; Stump and Hall 1995; Williams and Hall 1996; Law et al. 2006b). However, the importance of this region of the SLIV hairpin has not been studied. Interest-

ingly, phylogenetic analysis of the RNA sequences indicates that the 3' side of SLIV shows more sequence conservation than that of SLII and that, in particular, position U8 [AUUGCAGU8ACC] is universally conserved among metazoans (Fig. 2). This level of conservation suggests a strong evolutionary pressure against mutation, which could be consistent with protein interaction with the 3' side of the loop and the U8 nucleotide in particular. However, Urb-V and Urb-Vb exhibit very little change in binding affinity to the SLIV U8C mutant (Fig. 6). Surprisingly, Urb-Va exhibited the most dramatic response to this mutation with a significant loss of binding affinity (at least one order of magnitude). Previous work has shown that the 3' UCC does not contact the protein in the U1A:SLII complex; the nucleotides were replaced with a polyethylene glycol linker with no loss of binding affinity (Williams and Hall 1996), so the response of Urb-Va to the U8C substitution is rather mysterious. The SLIV A9G mutation also resulted in drastically reduced binding affinity by Urb-Va but insignificant changes in Urb-V and Urb-Vb affinity. The unexpectedly strong dependence of Urb-Va for this side of the loop reveals a new aspect of the different binding mechanisms of Urb-Va:SLIV and Urb-Va:SLII.

SLII and SLIV differences

To further probe the different RNA-binding mechanisms of the proteins and their specific nucleobase recognition, we constructed three RNA mutants: SLII C7G, SLIV G7C, and SLIV LCB in which the U-U loop-closing base pair (LCB) was replaced with the C-G that is found in SLII. In metazoans, C7 appears to be universally conserved in SLII sequences (Fig. 2). In metazoan SLIV sequences, the 7 loop position is considerably more variable; it is most commonly G or C, but U can also be found. However, in vertebrate RNA sequences, it appears to be universally conserved as a G (Fig. 2; Williams et al. 2013).

As expected, the SLII C7G mutation resulted in decreased affinity of Urb-Va, consistent with previous U1A binding data (Hall and Stump 1992). Urb-V and Urb-Vb binding to SLII C7G was not perturbed, consistent with a lack of specific

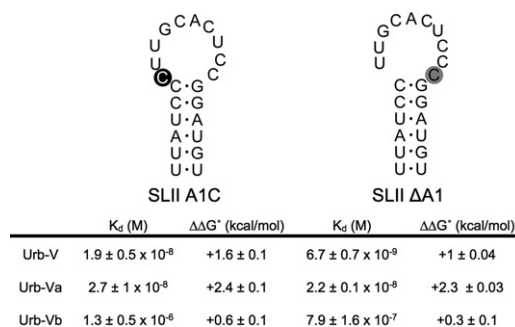


FIGURE 5. Binding to nucleotide A1 mutants. Dissociation constants for binding to each RNA are shown. $\Delta\Delta G^\circ = \Delta G^\circ$ (Mutant RNA) – ΔG° (WT SLII). All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

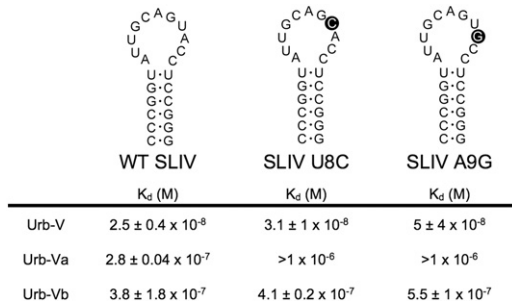


FIGURE 6. Binding to 3' loop mutants. Dissociation constants for binding to each RNA are shown. All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

recognition of this nucleotide (Fig. 7). Conversely, the SLIV G7C mutation resulted in a significant increase in affinity of Urb-Va for SLIV; again, there was little or no change in Urb-Vb or Urb-V affinity for these RNAs. These data indicate that prior to the gene duplication, Urb-V did not specifically recognize the nucleobase at position 7 in either SLII or SLIV. Urb-Vb retained this lack of discrimination, but this site is recognized specifically by Urb-Va, and the C at this position is important for high-affinity binding. The free energy associated with this recognition largely accounts for the increased affinity of Urb-Va for SLII, compared with Urb-V. Nucleotide C7 interacts with the peptide backbone in the $\beta 4/\alpha 3$ junction (Oubridge et al. 1994), suggesting that the mode of recognition of this base in Urb-Va/U1A will depend on orientation and dynamics of this region of the protein.

The SLIV LCB UU to CG mutation resulted in an increase in affinity for Urb-V and Urb-Va but resulted in no change to the binding affinity of Urb-Vb (Fig. 7). This mutation results in (almost) identical increases to Urb-V and Urb-Va binding affinity: $\Delta\Delta G^{\circ}_{\text{binding}}$ (Mutant – WT) = $-1.5/-1.2$ kcal/mol, respectively (Table 3). C7 and the loop-closing base pair account for much of the difference in RNA discrimination between Urb-Va and Urb-Vb.

Because SLIV G7C and SLIV LCB mutations account for two of the three differences between SLII and SLIV, we summed the binding free energy differences of these mutations to compare their contributions to protein specificity for SLII over SLIV (Table 3). A simple sum of the $\Delta\Delta G^{\circ}$ for Urb-V binding to both SLIV mutations results in a -1.8 kcal/mol preference for the mutants over SLIV, indicating that position 7 and the LCB could entirely account for Urb-V's preference for SLII over SLIV (assuming no cooperativity or contribution from other factors). In 250 mM KCl, Urb-Vb shows no significant difference in binding affinity for WT SLII vs. SLIV, SLIV G7C, or SLIV LCB, indicating that loop position 7 and the LCB do not contribute to Urb-Vb recognition of the stem-loops. However, Urb-Va presents a different pattern. While Urb-Va has $\Delta\Delta G^{\circ}_{\text{binding}}$ (SLII – SLIV) = -3.8 kcal/mol preference for SLII over SLIV, the additive contributions of the LCB mutation and the mutation at

position 7 result in a more modest -2.4 kcal/mol preference for the mutants over SLIV. Clearly, position 7 and the LCB are not sufficient to account for Urb-Va's specificity for SLII over SLIV, assuming a model of site independence. Like the modern U1A RRM, the binding mechanism of Urb-Va appears to be quite complex.

DISCUSSION

The conservation of the RRM protein sequence and RNA-binding activity among Urb, Urb-D and Urb-V represents an estimated 150 million years of evolution in which this family remained remarkably stable. However, a gene duplication in an ancestor of jawed vertebrates resulted in a short period of RRM sequence and functional divergence in both protein paralogs. Urb-Va and Urb-Vb are reconstructions of the last common ancestral RRM1 of modern vertebrate U1A and U2B' proteins, respectively. Given the similarities in RNA-binding specificity between Urb-Va and U1A and between Urb-Vb and U2B', it is likely that the mutations to the proteins (following the gene duplication) rapidly resulted in subfunctionalization of protein binding and localization.

In the Urb-Va/U1A lineage, we are able to identify five amino acids in $\beta 2/\text{Loop 3}$ that are responsible for the protein's specificity. The transition from VALKT (found in Urb, Urb-D, Urb-V, and Urb-Vb) to LVSRs (in Urb-Va), though conservative at each amino acid position, is sufficient to both increase affinity for SLII and decrease affinity for SLIV compared to its predecessor (Urb-V). This is consistent with previous structural and mutagenesis data that show U1A Loop 3 interacting with the 5' side of the RNA loop and the loop-closing base pair of SLII (Oubridge et al. 1994; Scherly et al. 1990a, 1991). Some of the single amino acid mutations in the loop sequence result in binding effects that are comparable to that of the full loop substitution. The single mutation of VALKT to LALKT is one of the mutations that has the most dramatic effect on RNA binding. In U1A, the comparable reciprocal substitution, L44V (U1A numbering), confers opposite changes to binding affinity and specificity (Rimmele

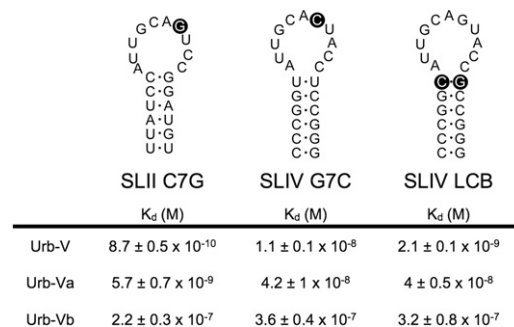


FIGURE 7. Binding to SLII and SLIV conversion mutants. Dissociation constants for binding to each RNA are shown. All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature.

TABLE 3. Protein family RNA affinity of RNA mutants vs. WT RNA

	Urb-V	Urb-Va	Urb-Vb
$\Delta\Delta G^\circ(\text{SLII-SLIV})$	-1.8 ± 0.1	-3.8 ± 0.1	$+0.15 \pm 0.3$
$\Delta\Delta G^\circ(\text{G7C-SLIV})$	-0.3 ± 0.1	-1.2 ± 0.1	-0.03 ± 0.2
$\Delta\Delta G^\circ(\text{LCB-SLIV})$	-1.5 ± 0.1	-1.2 ± 0.1	-0.1 ± 0.2
Sum of G7C + LCB	-1.8 ± 0.1	-2.4 ± 0.1	-0.13 ± 0.1

Difference in free energies of binding of each protein to the indicated RNAs is shown. All binding experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl_2 (pH 7) at room temperature.

and Belasco 1998). Our phylogenetic analysis and protein sequence reconstruction cannot determine which of these mutations came first.

In contrast to the Urb-Va branch, where functional changes are localized to a short stretch of residues in $\beta 2/\text{Loop } 3$, the evolution of Urb-Vb RRM1 after the gene duplication was much more complicated. Fourteen amino acid changes accumulated between Urb-V and Urb-Vb; these are not localized to any single region of the RRM. Preliminary results (data not shown) suggest that mutations throughout the RRM (and not localized to $\beta 2/\text{Loop } 3$) are important for changing the RNA-binding properties of the molecule.

Our structural and biochemical mutational studies of U1A and *Drosophila* SNF RRMs have identified a network of hydrogen bonding interactions between amino acid side chains that are important for RNA binding (data not shown; Kranz et al. 1996; Kranz and Hall 1998). This network is quite extensive, encompassing nearly the entire RRM surface, but it is different in *Drosophila* SNF and human U1A. We propose that the network of interactions in each RRM contributes substantially to their different RNA-binding properties. The similarity in slow and fast timescale conformational dynamics between SNF and Urb RRM1 (Williams et al. 2013) makes it likely that preformed networks similar to SNF were present in Urb-family proteins and were conserved prior to the vertebrate (gnathostome) gene duplication to allow these proteins to bind both SLII and SLIV with high affinities, while still preserving some discrimination between RNAs. However, the network was likely altered in both branches following the gene duplication. In U1A and SNF, engineered mutations of residues on the β -sheet have substantial effects on both the surface hydrogen bonding network and RNA binding (data not shown; Kranz et al. 1996; Kranz and Hall 1998). Given the many changes throughout the body of the RRM in the Urb-Vb branch, it is likely that the hydrogen bonding network has been very substantially altered.

U2A' interactions

In addition to binding RNA, U2B'' and SNF also bind the U2A' protein in the U2 snRNP. U2A' is conserved in eukary-

otes, indicating that it is an ancient protein (Collins and Penny 2005) that is found only in the U2 snRNP. When analyzing the effects of evolutionary changes between Urb, Urb-Va, and Urb-Vb, interactions with the protein binding partner U2A' may also have altered.

In the absence of RNA, U2A' binds tightly to human U2B'' but binds with much weaker affinity to both U1A and *Drosophila* SNF (data not shown). It is tempting to speculate that prior to the vertebrate gene duplication, Urb proteins bound ancestral U2A' proteins with weak affinity and that, at some point in the evolution of the Urb-Vb/U2B'' branch, the RRM adapted into a high-affinity binder for U2A'. Whether this is true and whether the subsequent high-affinity U2B''-U2A' interaction was the result of protein-protein coevolution or adaptations of a single protein remain to be determined. Recent work on the evolution of yeast transcription factors has suggested that in macromolecular assemblies, one of the consequences of gene duplication followed by protein subfunctionalization is a dominant-negative effect between the paralogous proteins with respect to other components of the assembly. Because this can be functionally deleterious, in order for the duplicated proteins to persist, there is a strong evolutionary pressure to minimize paralog interference, thus accounting for differences in binding to other members of the macromolecular complex (Baker et al. 2013). The difference in the binding affinities of U1A and U2B'' for U2A' is suggestive of adaptations to minimize paralog interference, but conclusive evidence for such an evolutionary pressure would improve our understanding of the functional role of these proteins in the snRNPs.

RNA recognition

The high level of conservation of SLII and SLIV in all family lineages implies significant pressure against change. RNA mutagenesis allowed us to examine the binding requirements of each of the proteins surrounding the gene duplication to see how these pressures may have changed with protein mutations.

The adenosine residue in loop position 1 (A1) is found in both SLII and SLIV. Mutations to A1 resulted in weakened binding affinity by Urb-V and Urb-Va but, surprisingly, not Urb-Vb. Urb-Va's requirement for A1 provides the evolutionary pressure to maintain this nucleotide in SLII. Prior to the gene duplication, SLIV was bound by Urb-V, which also specifically recognized A1. However, following the duplication, all vertebrate SLIV loops retain A1 despite a seeming lack of recognition by Urb-Vb. It is possible that in the ternary SLIV:RRM:U2A' complexes of vertebrates, the RRM interacts directly with A1, particularly if Loop 3 undergoes a conformational change upon U2A' binding. However, if this is not the case, it is possible that this nucleotide may eventually be mutated or eliminated in vertebrate SLIV RNAs.

Caenorhabditis elegans evolution has resulted in a different solution to U1A/U2B''/SLII/SLIV recognition that is

pertinent to vertebrate SLIV A1. In worms, the U1A and U2B' proteins are redundant, and the worm is viable upon loss of either one (but not both) (Saldi et al. 2007). The snRNAs in *C. elegans* differ from those in vertebrates: in particular, SLIV has lost nucleotide A1 (loop sequence: UU GCACUGC), although SLII retains it. In the absence of biochemical data describing specific RRM:RNA interactions, *in vivo* data show that in *C. elegans*, U1A and U2B' proteins are able to bind both RNAs with sufficient affinity to be retained in both snRNPs.

The 3' side of SLIV has four nucleotides (AUUGCAGU8A 9C10C11) that are fairly well-conserved in vertebrates (U8 is universally conserved, not only in vertebrates but in all metazoans) (Fig. 2). Urb-V and Urb-Vb, two proteins that likely bound SLIV *in vivo*, do not discriminate among U8C and A9G mutations, which indicates that the nucleobases are not recognized specifically by the proteins in the bimolecular complex. Unexpectedly, we find that Urb-Va exhibits significantly weaker binding affinity for U8C and A9G 3' loop mutants of SLIV RNAs. While it is important to remember that nucleotide substitutions within the loop could potentially alter its secondary structure, the unique response of Urb-Va binding shows that it is a protein-specific phenomenon. The result is unexpected since, in SLII, this side of the RNA loop does not make contact with the modern U1A protein. It is possible that the interaction with the 3' side of the loop was subsequently lost as the protein continued to evolve, or that this interaction is only seen in SLIV, an RNA that is not bound by this protein in the cell.

In addition to nucleobases that are specifically recognized by the RRMs, there are examples of nucleobases that act as "negative discriminators" (Schimmel 1989) to prevent binding by an RRM. The loop G7 in SLIV is one example; while Urb-V and Urb-Vb are insensitive to the specific nucleobase at this position (and it is variable across metazoan SLIV sequences), Urb-Va binds with significantly weaker affinity when it is present. In vertebrate SLIV, this is universally conserved as a G, consistent with evolution of SLIVs to negatively discriminate against binding by Va proteins. The loop-closing base pair serves as a negative discriminator in SLIV and a positive discriminator in SLII RNAs (for Urb-Va/U1A). In vertebrate RNAs, the LCB nucleotides are universally conserved as CG in SLII and UU in SLIV. While both Urb-V and Urb-Va have similar differences in binding free energy for hairpins with CG vs. UU loop-closing base pairs, Urb-V does not discriminate between CG and UG loop-closing base pairs (high affinity for both), while Urb-Va does (data not shown). Loop-closing base pairs in SLIV RNAs across metazoans are substantially more variable than in vertebrates, reinforcing the distinctiveness of LCB sequence conservation in vertebrates.

Although our experiments are designed to test the affinity and specificity of the RRMs during their evolution, they also report on the evolution of the RNA. Notably, while the RRMs in both Urb-Va and Urb-Vb lineages have acquired

multiple mutations, we have observed that the RNAs have evolved minimally in vertebrates (Fig. 2). Evolutionary pressures are exerted by both the RRMs and the RNA stem-loops to maintain their functional relationships, and we have identified several nucleotides that Urb-V and Urb-Va use for discrimination of SLII from SLIV even as the protein sequences vary. From the perspective of the snRNA, it would appear that SLII and SLIV RNA sequences determine what mutations of Urb-Va and Urb-Vb are evolutionary winners.

Subfunctionalization: different binding modes

Prior to the gnathostome (vertebrate) duplication of the Urb gene, a single Urb-family protein localized to both the U1 and U2 snRNPs by binding both U1 SLII and U2 SLIV. This protein bound SLII RNAs with very high affinity. It also bound to SLIV RNAs, although binding affinity for SLIV was somewhat weaker. However, after the gene duplication, Urb-Va and Urb-Vb mutated to adopt specialized binding mechanisms suited to their *in vivo* targets.

Our data show that Urb-Va gained marginal affinity for SLII while losing substantial affinity for SLIV following the gene duplication by adopting a mode of binding that is almost certainly a modification of the RNA recognition employed by Urb-V, effects that were entirely mediated by mutations to the protein β 2/Loop 3. Recognition of SLII by U1A is complex (Scherly et al. 1991; Oubridge et al. 1994; Williams and Hall 1996; Katsamba et al. 2002; Law et al. 2006a,b) and includes interactions between the β 4/ α 3 junction and the top of the RNA loop (C5A6C7), interactions between the RNP motifs and the 5' side of the loop, and interactions between Loop 3 and the CG loop-closing base pair and the adjacent A1. In U1A, discrimination of both the loop-closing base pair and C7 are implicated in the difference in binding affinity for SLII and SLIV. In comparison, Urb-V specificity for SLII over SLIV is almost entirely mediated by differences in recognition of the LCB.

A valine to leucine mutation of amino acid 41, which is at the junction of β 2 and Loop 3 (V₄₁ALKT), is sufficient to confer both increased affinity for SLII and decreased affinity for SLIV. This one change is important for the protein to distinguish between the two stem-loops, presumably by altering the recognition of RNA C7. The adjacent A42V substitution has a very similar effect. These two amino acids are nominally located in β 2; the V/L41 side chain is solvent-exposed, while the A/V42 side chain is directed toward the core of the protein and is unlikely to bind the RNA directly (see Fig. 2C). The mechanisms by which this amino acid facilitates RNA discrimination could be mediated by propagated effects on Loop 3 structure/dynamics.

Like Urb-V and Urb-Va, optimal binding of Urb-Vb to target RNA sequences requires that the RNA-binding element be present within the context of a loop that is at least 10 nt long. Urb-Vb retains Urb-V's tolerance of either a C or G residue at loop position 7 but has lost recognition

of the LCB and A1. The implication is that changes in other parts of the protein impede interactions with the LCB and A1. It is plausible that in Urb-Vb, specific binding to the stem-loops is mediated predominantly through interactions between the RNP motifs and the 5' side of the loop. Discrimination between SLII and SLIV by Urb-Vb becomes apparent at lower salt concentrations where, in contrast to Urb-V and Urb-Va, Urb-Vb has a higher affinity for SLIV than for SLII.

CONCLUSION

Reconstruction of ancestral U1A/U2B' proteins has provided a means to understand how they altered their RNA-binding function following a gene duplication. The resulting protein subfunctionalization (specialization) is characterized by binding modes that retain ancestral features while adopting specific *in vivo* target recognition that allows segregation to the correct snRNP. While the evolution of Urb-Va (and ultimately U1A) specificity appears to be moderately straightforward, the analogous path of Urb-Vb (and so, U2B') is quite complex and remains to be elucidated.

MATERIALS AND METHODS

Maximum-likelihood ancestral RRM sequences for the intermediate nodes between Urb and modern vertebrate proteins were obtained by analyzing modern proteins and the reconstructed phylogeny, as previously described (Williams et al. 2013). CODEML (Yang 1997, 2007) was used for the final sequence reconstruction. Sequences for Urb-D, Urb-V, Urb-Va, and Urb-Vb were obtained from this analysis; this corresponds to the nodes that were robustly reconstructed in the phylogeny. Posterior probabilities of the reconstructed amino acids were also obtained from CODEML.

Protein expression

Expression of Urb has been previously described (Williams et al. 2013). Each subsequent protein was achieved by repeated Quick-Change mutagenesis (Agilent) and purified in the same manner. Briefly, the protein was expressed under control of the TAC promoter in BL-21(DE3) cells (Invitrogen) in LB media. Cells were grown at 37°C to an optical density of 0.8, then induced with 1 mM IPTG for 4 h at 25°C, spun down, and kept at -80°C. Cells were lysed by French Press in 30 mM sodium acetate (pH 5.3), 200 mM NaCl, 2 mM EDTA, and 8.5% sucrose with Sigma Protease Inhibitor, DNaseII, and PMSF added at the time of lysis. Lysate was spun down and passed over an SP-XL FPLC column (GE) that was pre-equilibrated in 20 mM sodium cacodylate (pH 7). The column was washed with 0 M and 100 mM NaCl, then eluted with a gradient of 100–400 mM NaCl over 3 h. Fractions containing the protein were pooled, concentrated, and buffer-exchanged into 10 mM sodium cacodylate (pH 7), 10 mM KCl, and 2 mM EDTA.

RNA transcription

RNA stem-loops were transcribed by T7 polymerase from DNA oligonucleotides (IDT) using [α -³²P] UTP and [α -³²P] CTP. Tran-

scription products were gel-purified. RNA sequences are all as depicted in respective figures.

RNA binding

Nitrocellulose filter binding experiments were performed as previously described (Williams and Hall 2010). Unless otherwise noted, all experiments were performed in 250 mM KCl, 10 mM cacodylate, and 1 mM MgCl₂ (pH 7) at room temperature. Titrations were fit to a Langmuir isotherm in KaleidaGraph. Experiments were performed in duplicate and repeated at least twice. Reported error is the greater of either the standard deviation or the propagated error.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

Research was supported by the National Institutes of Health R01 GM096444 to K.B.H., and F31 GM089576 to S.G.W.

Received January 9, 2014; accepted March 17, 2014.

REFERENCES

- Baker CR, Hanson-Smith V, Johnson AD. 2013. Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**: 104–108.
- Benitex Y, Baranger AM. 2007. Recognition of essential purines by the U1A protein. *BMC Biochem* **8**: 22.
- Collins L, Penny D. 2005. Complex spliceosomal organization ancestral to extant eukaryotes. *Mol Biol Evol* **22**: 1053–1066.
- Hall KB. 1994. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* **33**: 10076–10088.
- Hall KB, Stump WT. 1992. Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res* **20**: 4283–4290.
- Harms MJ, Thornton JW. 2013. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* **14**: 559–571.
- Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* **14**: 33–38.
- Katsamba PS, Bayramyan M, Haworth IS, Myszkowski DG, Laird-Offringa IA. 2002. Complex role of the β_2 - β_3 loop in the interaction of U1A with U1 hairpin II RNA. *J Biol Chem* **277**: 33267–33274.
- Kranz JK, Hall KB. 1998. RNA binding mediates the local cooperativity between the β -sheet and the C-terminal tail of the human U1A RBD1 protein. *J Mol Biol* **275**: 465–481.
- Kranz JK, Lu J, Hall KB. 1996. Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci* **5**: 1567–1583.
- Laird-Offringa IA, Belasco JG. 1995. Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc Natl Acad Sci* **92**: 11859–11863.
- Law MJ, Linde ME, Chambers EJ, Oubridge C, Katsamba PS, Nilsson L, Haworth IS, Laird-Offringa IA. 2006a. The role of positively charged amino acids and electrostatic interactions in the complex of U1A protein and U1 hairpin II RNA. *Nucleic Acids Res* **34**: 275–285.
- Law MJ, Rice AJ, Lin P, Laird-Offringa IA. 2006b. The role of RNA structure in the interaction of U1A protein with U1 hairpin II RNA. *RNA* **12**: 1168–1178.

- McConnell TS, Lokken RP, Steitz JA. 2003. Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. *RNA* **9**: 193–201.
- Oubridge C, Ito N, Evans PR, Teo CH, 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.
- Polycarpou-Schwarz M, Gunderson SI, Kandels-Lewis S, Seraphin B, Mattaj IW. 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.
- Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**: 645–650.
- Record MT, Lohman TM, de Haseth P. 1976. Ion effects on ligand-nucleic acid interactions. *J Mol Biol* **107**: 145–158.
- Rimmele ME, Belasco JG. 1998. Target discrimination by RNA-binding proteins: role of the ancillary protein U2A' and a critical leucine residue in differentiating the RNA-binding specificity of spliceosomal proteins U1A and U2B''. *RNA* **4**: 1386–1396.
- Saldi T, Wilusz C, MacMorris M, Blumenthal T. 2007. Functional redundancy of worm spliceosomal proteins U1A and U2B''. *Proc Natl Acad Sci* **104**: 9753–9757.
- Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW. 1990a. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. *Nature* **345**: 502–506.
- Scherly D, Dathan NA, Boelens W, van Venrooij WJ, Mattaj IW. 1990b. The U2B'' RNP motif as a site of protein–protein interaction. *EMBO J* **9**: 3675–3681.
- Scherly D, Kambach C, Boelens W, van Venrooij WJ, Mattaj IW. 1991. Conserved amino acid residues within and outside of the N-terminal ribonucleoprotein motif of U1A small nuclear ribonucleoprotein involved in U1 RNA binding. *J Mol Biol* **219**: 577–584.
- Schimmel P. 1989. Parameters for the molecular recognition of transfer RNAs. *Biochemistry* **28**: 2747–2759.
- Stump WT, Hall KB. 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.
- Williams DJ, Hall KB. 1996. RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J Mol Biol* **257**: 265–275.
- Williams SG, Hall KB. 2010. Coevolution of *Drosophila* snf protein and its snRNA targets. *Biochemistry* **49**: 4571–4582.
- Williams SG, Hall KB. 2011. Human U2B'' protein binding to snRNA stemloops. *Biophys Chem* **159**: 82–89.
- Williams SG, Harms MJ, Hall KB. 2013. Resurrection of an Urbilaterian U1A/U2B''/SNF protein. *J Mol Biol* **425**: 3846–3862.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555–556.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.