
Transposition of two amino acids changes a promiscuous RNA binding protein into a sequence-specific RNA binding protein

STEPHEN M. GARREY, DANIELLE M. CASS, ANICA M. WANDLER, MARY S. SCANLAN,
and J. ANDREW BERGLUND

Department of Chemistry and Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA

ABSTRACT

In yeast (*Saccharomyces cerevisiae*), the branchpoint binding protein (BBP) recognizes the conserved yeast branchpoint sequence (UACUAAC) with a high level of specificity and affinity, while the human branchpoint binding protein (SF1) binds the less-conserved consensus branchpoint sequence (CURAY) in human introns with a lower level of specificity and affinity. To determine which amino acids in BBP provide the additional specificity and affinity absent in SF1, a panel of chimeric SF1 proteins was tested in RNA binding assays with wild-type and mutant RNA substrates. This approach revealed that the QUA2 domain of BBP is responsible for the enhanced RNA binding affinity and specificity displayed by BBP compared with SF1. Within the QUA2 domain, a transposition of adjacent arginine and lysine residues is primarily responsible for the switch in RNA binding between BBP and SF1. Alignment of multiple branchpoint binding proteins and the related STAR/GSG proteins suggests that the identity of these two amino acids and the RNA target sequences of all of these proteins are correlated.

Keywords: KH domain; branchpoint binding protein; splicing factor 1; branchpoint sequence; pre-mRNA splicing

INTRODUCTION

There are three essential RNA elements for pre-mRNA splicing, the 5' splice site, the 3' splice site, and the branchpoint sequence. These three elements are necessary for the two transesterification reactions that remove an intron and ligate two exons together (Reed 1996). In the first reaction, the 2' OH of the branchsite adenosine is used as the nucleophile to attack the phosphodiester bond between the 5' exon and 5' splice site, creating the free 5' exon and lariat intermediate. In the second reaction, the 3' OH of the free 5' exon is the nucleophile that attacks the phosphodiester bond between the 3' splice site and downstream exon, creating the ligated exons and excised intron. These steps are catalyzed by the spliceosome, which is composed of five snRNAs and ~150 proteins (Staley and Guthrie 1998; Will and Lührmann 2006).

Although the branchpoint sequence, or more specifically, a branchsite nucleophile, is essential for splicing, this motif is less conserved in comparison with the 5' and 3' splice sites across eukaryotic organisms. In human introns the branchpoint sequence is degenerate with the consensus sequence of CURAY (R=purine, Y=pyrimidine) (Keller and Noon 1984). For some human introns it is difficult to identify a putative branchpoint sequence, and mutating a mapped branchsite often results in the splicing machinery using a cryptic branchsite (Ruskin et al. 1985). Multiple branchsites within a single human intron may be used to create different 3' splice sites, resulting in different mRNAs (Wang et al. 2007). In contrast to the degenerate branchpoint sequence in humans, in yeast (*Saccharomyces cerevisiae*), the branchpoint sequence is almost always UACUAAC (Langford and Gallwitz 1983; Pikielny et al. 1983).

The branchpoint sequence is specifically recognized at least twice in splicing, first by the branchpoint binding protein and then subsequently by U2 snRNP through an RNA–RNA interaction in which the branchsite adenosine is unpaired and bulged out for use as a nucleophile in the first reaction of splicing (Query et al. 1994, 1996). The degeneracy of the branchpoint sequence in humans does not

Reprint requests to: J. Andrew Berglund, Department of Chemistry and Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA; e-mail: aberglund@molbio.uoregon.edu; fax: (541) 346-5891.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.633808>.

correlate with the U2 snRNA sequence, because the GUAGUA sequence of U2 snRNA that base pairs with the branchpoint sequence is highly conserved. Although natural branchpoint sequences in human introns are highly variable, the yeast UACUAAC sequence has been shown to be an optimal branchpoint sequence for splicing in the human system (Zhuang et al. 1989).

The branchpoint binding protein, referred to as branchpoint binding protein (BBP) in yeast (Berglund et al. 1997) and splicing factor 1 (SF1) in human (Kramer 1988) is a highly conserved protein that contains multiple RNA binding domains: a K homology (KH) domain, a QUA2 domain, and either one or two zinc knuckles (Berglund et al. 1998b; Rain et al. 1998). The KH domain is a common RNA binding domain that can be subdivided into two subclasses: types I and II. Type I domains, which include the KH domain of BBP/SF1, fold into a $\beta\alpha\beta\beta\alpha$ topology, while type II KH domains fold into an $\alpha\beta\beta\alpha\alpha\beta$ topology (Grishin 2001; Auweter et al. 2006). KH domains bind RNA in a cleft between two loops: a small loop between the second and third α helices consisting of a conserved GXXG motif and a large “variable loop” of nonconserved sequence between β strands two and three.

A subclass of type I KH proteins called the signal transduction activator of RNA (STAR) family or the GRP33-Sam68-GLD-1 (GSG) family have KH domains that are extended at both ends (Vernet and Artzt 1997; Lukong and Richard 2003). The term STAR/GSG will be used below when referring to this subclass of KH domain containing proteins. The STAR/GSG proteins have extended KH domains with a QUA1 domain at the N-terminal side and a QUA2 domain at the C-terminal side. The QUA1 domain is responsible for homodimerization (Chen et al. 1997; Chen and Richard 1998). BBP/SF1 is unique among STAR/GSG proteins, in that it lacks a QUA1 domain. The NMR structure of SF1 in complex with the yeast branchpoint sequence UACUAAC RNA showed that the QUA2 domain is an α helix that extends the RNA binding site of these KH domains (Liu et al. 2001). Specifically, the QUA2 domain of SF1 interacts with the 5' nucleotides of the branchpoint sequence RNA, the underlined nucleotides in the UACUAAC sequence (Liu et al. 2001).

Another notable feature of the STAR/GSG proteins is that they contain a single KH domain, whereas most other KH domain proteins contain multiple KH domains (Gibson et al. 1993; Burd and Dreyfuss 1994). Homodimerization significantly enhances the RNA binding affinity of the STAR/GSG proteins (Chen and Richard 1998; Chow et al. 2000; Ryder and Williamson 2004). Lacking the QUA1 domain, BBP/SF1 does not homodimerize; instead, it interacts with Mud2 in yeast or U2AF65 in human, and these proteins bind cooperatively to the branchpoint sequence and adjacent polypyrimidine tract at the 3' end of the intron (Abovich and Rosbash 1997; Berglund et al. 1998a). Finally, both BBP and SF1 contain zinc knuckles,

which are C-terminal of the QUA2 domains. BBP contains two zinc knuckles, the first of which enhances RNA binding, while the second zinc knuckle may slightly inhibit binding (Garrey et al. 2006). SF1 contains only one zinc knuckle and its presence enhances RNA binding (Berglund et al. 1998b). Zinc knuckles have been shown to be important for RNA binding in other proteins like the HIV-1 nucleocapsid protein (Dannull et al. 1994; De Guzman et al. 1998) and the human SR protein 9G8 (Cavaloc et al. 1999).

Although BBP and SF1 are orthologous proteins, these two proteins have different affinities for the branchpoint sequence and different specificities. These differences between BBP and SF1 are not surprising, because the yeast branchpoint sequence is the highly conserved UACUAAC (Langford and Gallwitz 1983; Pikielny et al. 1983), while human branchpoints are characterized by a degenerate consensus sequence, CURAY (Keller and Noon 1984). BBP binds RNAs containing the branchpoint sequence of UACUAAC with nanomolar affinity (Garrey et al. 2006). Mutational analysis of the branchpoint sequence along with SELEX has demonstrated the importance of all seven positions of the yeast branchpoint sequence for sequence-specific binding by BBP (Berglund et al. 1997; Garrey et al. 2006). Meanwhile, human SF1 binds the yeast BPS with more than 200-fold weaker affinity and much less specificity than that of BBP (Berglund et al. 1997; Garrey et al. 2006).

In this study, chimeras of BBP and SF1 were generated to determine which amino acids are responsible for their significant differences in branchpoint sequence RNA binding affinity and specificity. We find that the QUA2 domains of BBP and SF1 are primarily responsible for the differences in specificity and affinity between BBP and SF1. Within the QUA2 domain, a transposition of two amino acids confers yeast BBP-like RNA binding affinity and specificity to the human SF1 for the yeast branchpoint RNA sequence.

RESULTS

Difference in RNA binding affinity and specificity between BBP and SF1

Previously we had shown qualitatively that there are significant differences in RNA binding affinity and specificity between SF1 and BBP (Berglund et al. 1997). Here, we have quantified the differences between the two proteins. The wild-type protein constructs used for comparison contained only the RNA binding domains: KH, QUA2, and one zinc knuckle (Fig. 2, below). BBP contains a second C-terminal zinc knuckle, but this domain was excluded because we had previously shown it is not involved in RNA binding (Garrey et al. 2006). The BBP construct consisted of amino acids 147–296 and will be referred to as BBP-296. The SF1 construct consisted of amino acids 134–297 and will be referred to as SF1-297.

The affinities of BBP-296 and SF1-297 for a 22-mer RNA (BP-22, Fig. 1C) containing the yeast consensus branchpoint sequence UACUAAC were measured. BBP-296 bound the BP-22 RNA with a K_d of 3 nM (Fig. 1A, lanes 1–6; Table 1), while SF1-297 bound this RNA with a K_d of only 740 nM (Fig. 1, lanes 13–18; Table 1). Therefore, these proteins have an approximate 250-fold difference in their affinity for the optimal branchpoint sequence.

To compare the specificities of BBP-296 and SF1-297, we tested their binding to four mutant RNAs with different point mutations within the branchpoint sequence. The

positions of the branchpoint sequence are numbers 1–7 and mutants will be referred to in the following format, wild-type nucleotide–position–mutant nucleotide (i.e., U1G). The binding of BBP to these mutant RNAs demonstrates the high-sequence specificity of the yeast protein (Fig. 1A). Whereas BBP-296 binds the wild-type branchpoint sequence with a K_d in the low nanomolar range, its apparent K_d for each of the mutant sequences is in the micromolar range (Fig. 1, lanes 7–12). Mutating any of these four nucleotides results in almost 400-fold to over 7000-fold weaker binding by BBP (Fig. 1A). This is shown graphically in Figure 1B, in which $\Delta\Delta G$ [$\Delta\Delta G = -RT \ln(K_{d1}/K_{d2})$] reflects the difference in affinity of a protein for two different RNA substrates and can be used as a measure of sequence specificity (Fig. 1B). For the human protein, SF1-297, these same RNA mutations reduce binding affinities between approximately twofold and 29-fold, or in terms of $\Delta\Delta G$, from 1.6 kJ/mol to 7.8 kJ/mol (Fig. 1B; Table 1). These results show that BBP-296 is highly sensitive to changes in the branchpoint sequence, whereas SF1-297 has relaxed sequence specificity (Table 1).

Determining the regions of BBP and SF1 that are responsible for the differences in RNA binding affinity and specificity

To identify the region or regions of BBP-296 that give it higher affinity and specificity over SF1-297, we created chimeric proteins in which each of five segments of BBP-296 replaced the corresponding region of SF1-297 (Fig. 2A). The alignment of BBP and SF1, as well as the domain architecture, were used to delineate the regions that were exchanged (Fig. 2B). The first region, which we call the conserved KH (cKH), (SF1 amino acids 134–187) contains the first half of the KH domain. It contains the most conserved part of the KH domain, and includes the GXXG motif and the first two β -strands and α -helices. The second region (SF1 amino acids 188–201) contains the variable loop (VL), which is the least conserved region of the protein. Region three is on the “backside” of the protein, away from the RNA binding site, and contains the third α -helix and third β -sheet (SF1 amino acids 202–232). The last two regions are the QUA2 domain (SF1 amino acids 233–257) and the zinc knuckle domain (ZnK) (amino acids 258–297).

The human SF1-297 construct was exchanged with one of each of the five different regions from yeast BBP-296: creating five different yeast–human chimeras; SF1-ycKH, SF1-yVL, SF1-y β 3 α 3, SF1-yQUA2, and SF1-yZnK (see Table 1). These five chimeric proteins were made and assayed for binding to the wild-type BP-22 RNA (UACUAAC) and the four mutant RNAs as shown in Figure 1. Four of the five chimeric proteins bind RNA, but SF1-y β 3 α 3 did not (data not shown), suggesting that this protein was not properly folded. This result is not surprising, because amino acids from this region are part of the

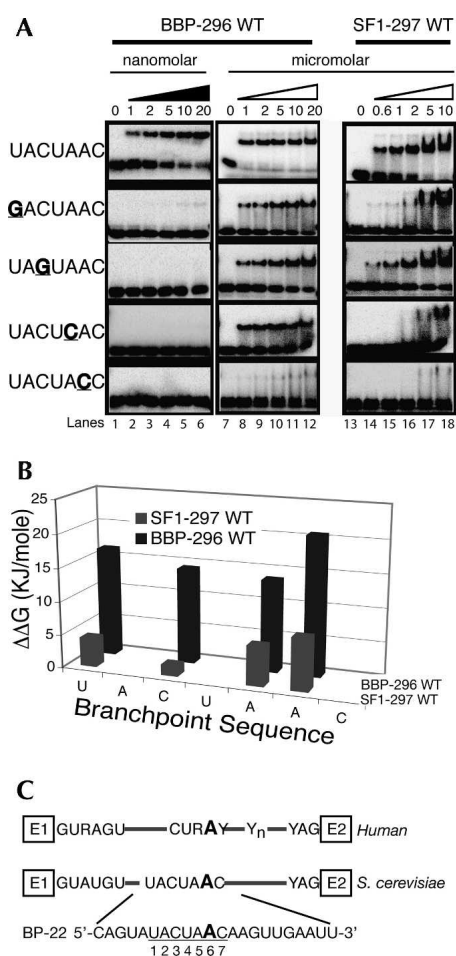

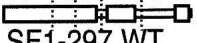
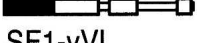
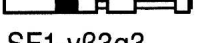

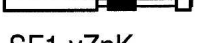

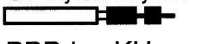


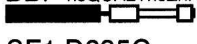
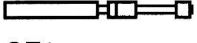
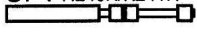
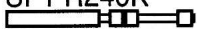
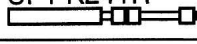


FIGURE 1. A quantitative comparison of the RNA binding specificity of BBP and SF1. (A) Binding to the BP-22 RNA is shown in the top row of the gels. BP-22 RNA contains the wild-type branchpoint sequence and is labeled on the left as UACUAAC. Binding to the mutant RNAs are below the wild-type UACUAAC and labeled on the left. Point mutations are underlined and in bold. Concentrations are listed underneath the triangles. (Lanes 1–6) Binding to BBP-296 in the low nanomolar range, (black triangle). To determine K_d s for the mutant RNAs, binding in the micromolar range (white triangle) was required (lanes 7–12). (B) A bar graph comparing the specificities measured in $\Delta\Delta G$ for the point mutations tested in A with BBP and SF1. (C) A schematic of yeast and human introns are shown with the BP-22 RNA sequence used in these studies shown at the bottom. The branchsite adenosine used as the nucleophile in the first step of chemistry is underlined and in a larger font.

TABLE 1. Disassociation constants (K_d s) and $\Delta\Delta$ Gs for BBP, SF1, and chimeric proteins

	<i>UACU</i> AAC K_d (μ M)	<i>GACU</i> AAC K_d (μ M) $\Delta\Delta$ G	<i>UCCU</i> AAC K_d (μ M) $\Delta\Delta$ G	<i>UAGU</i> AAC K_d (μ M) $\Delta\Delta$ G	<i>UACU</i> CAC K_d (μ M) $\Delta\Delta$ G	<i>UACU</i> ACC K_d (μ M) $\Delta\Delta$ G
BBP-296 WT 	0.003(\pm 0.0002)	3.7 (\pm 3.7) 16 KJ/mole	1.53 (\pm 0.83) 14.3 KJ/mole	1.5 (\pm 1.5) 14 KJ/mole	1.2 (\pm 0.77) 14 KJ/mole	24 (\pm 23) 21 KJ/mole
SF1-297 WT 	0.74 (\pm 0.31)	4.8 (\pm 1.2) 4.3 KJ/mole	0.32 (\pm 0.11) -1.9 KJ/mole	1.5 (\pm 1.9) 1.6 KJ/mole	9.4 (\pm 3.3) 5.9 KJ/mole	22 (\pm 7.9) 7.8 KJ/mole
SF1-ycKH 	2.3 (\pm 1.4)	11 (\pm 3.7) 3.7 KJ/mole	ND	1.9 (\pm 0.73) -0.4 KJ/mole	23 (\pm 9.3) 5.4 KJ/mole	12 (\pm 8.5) 3.8 KJ/mole
SF1-yVL 	1.2 (\pm 0.52)	10 (\pm 2.1) 4.8 KJ/mole	ND	1.4 (\pm 0.50) 0.4 KJ/mole	26 (\pm 5.8) 7.1 KJ/mole	50 (\pm NA) 8.6 KJ/mole
SF1-yβ3α3 	ND	ND	ND	ND	ND	ND
SF1-yQUA2 	0.036 (\pm 0.021)	3.0 (\pm 0.91) 10 KJ/mole	1.6 (\pm NA) 8.7 KJ/mole	3.7 (\pm 2.4) 11 KJ/mole	7.0 (\pm 4.8) 12 KJ/mole	19 (\pm 4.7) 14 KJ/mole
SF1-yZnK 	0.35 (\pm 0.08)	4.7 (\pm 0.18) 6.0 KJ/mole	ND	0.69 (\pm 0.66) 1.6 KJ/mole	8.3 (\pm 1.8) 7.3 KJ/mole	18 (\pm 12) 9.1 KJ/mole
SF1-yQUA2+yZnK 	0.014 (\pm 0.003)	0.74 (\pm 0.20) 9.1 KJ/mole	ND	0.49 (\pm 0.16) 8.2 KJ/mole	0.84 (\pm 0.19) 9.4 KJ/mole	5.0 (\pm 0.98) 14 KJ/mole
BBP-hscKH 	0.031 (\pm 0.018)	>30 (\pm NA) >16 KJ/mole	ND	22 (\pm NA) 15 KJ/mole	>30 (\pm NA) >16 KJ/mole	>30 (\pm NA) >16 KJ/mole
BBP-hsQUA2 	0.50 (\pm 0.060)	5.7 (\pm 0.97) 5.6 KJ/mole	1.1 (\pm NA) 1.8 KJ/mole	0.66 (\pm 0.080) 0.7 KJ/mole	9.2 (\pm 1.2) 6.7 KJ/mole	9.3 (\pm 2.5) 6.8 KJ/mole
BBP-hsQUA2+hsZnK 	1.2 (\pm 0.14)	10 (\pm 2.5) 5.0 KJ/mole	ND	2.4 (\pm 0.20) 1.6 KJ/mole	21 (\pm 7.2) 6.7 KJ/mole	14 (\pm 2.8) 5.7 KJ/mole
SF1 D235G 	0.21 (\pm 0.020)	2.8 (\pm 1.4) 6.0 KJ/mole	ND	0.41 (\pm 0.17) 1.5 KJ/mole	4.2 (\pm 0.12) 6.9 KJ/mole	8.3 (\pm 2.5) 8.5 KJ/mole
SF1 R240K/K241R 	0.040 (\pm 0.01)	1.0 (\pm 0.33) 7.4 KJ/mole	0.36 (\pm NA) 5.1 KJ/mole	1.2 (\pm 0.72) 7.8 KJ/mole	2.1 (\pm 0.37) 9.1 KJ/mole	12 (\pm 4.1) 13 KJ/mole
SF1 R240K 	0.070 (\pm 0.002)	0.95 (\pm 0.48) 6.0 KJ/mole	0.55 (\pm 0.07) 4.7 KJ/mole	1.1 (\pm 0.3) 6.4 KJ/mole	2.0 (\pm 0.6) 7.7 KJ/mole	7.3 (\pm 2.1) 11 KJ/mole
SF1 K241R 	0.20 (\pm 0.04)	2.0 (\pm 0.42) 5.3 KJ/mole	0.62 (\pm 0.06) 2.6 KJ/mole	0.41 (\pm 0.14) 1.7 KJ/mole	2.19 (\pm 0.24) 5.5 KJ/mole	8.0 (\pm 2.2) 8.5 KJ/mole

The protein constructs and their domain structures are shown in the far left column, each with three rectangles representing the KH, QUA2, and zinc knuckle domains. Black indicates yeast BBP regions and white indicates human SF1 regions. Each construct has two rows associated with it. The first row shows the K_d and the standard deviation in parentheses for the corresponding RNA listed at the top of the table. The second row lists the $\Delta\Delta$ G. (NA) Very little or no binding when referring to a K_d or when used to represent a standard deviation in which only two repeats were done.

core of the protein (Liu et al. 2001). The RNA binding data for SF1-ycKH, SF1-yVL, and SF1-yZnK are shown in Figure 3. The RNA binding specificity and affinity for these three chimeras are similar to SF1-297 (Table 1). The subtle differences between SF1-297 and the chimeras included modestly reduced sequence specificity of one chimera and a general enhancement of RNA binding affinity for another (Fig. 3B). The sequence specificity of SF1-ycKH is reduced compared with SF1-297, specifically at C3 (1.6 KJ/mole for

SF1-297 versus -0.4 KJ/mole for SF1-ycKH) and A6 (7.8 KJ/mole for SF1-297 versus 3.8 KJ/mole for SF1-ycKH) of the branchpoint sequence (Fig. 3; Table 1). Meanwhile, SF1-yZnK binds the wild-type RNA with a twofold lower K_d in comparison with the human SF1-297 (Fig. 3; Table 1). The general RNA binding affinity was enhanced for SF1-yZnK, and only minimal changes in specificity were observed when the zinc knuckles were exchanged between the yeast and human proteins (Table 1, cf. $\Delta\Delta$ Gs for

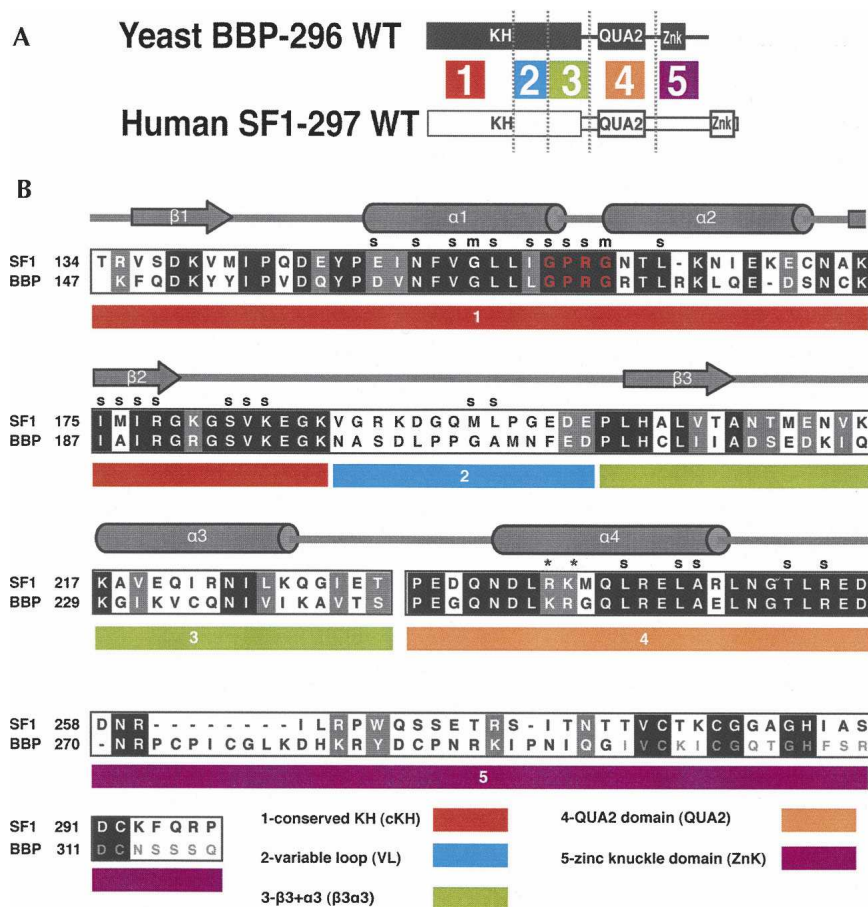


FIGURE 2. A map of the five regions used to make the BBP and SF1 chimeras. (A) The domain structure of BBP-296 (black) and SF1-297 (white), and placement of the five different regions. (B) An alignment of BBP-296 and SF1-297 and detailed map of the five regions underneath. A key of the five regions and their names are listed at the *bottom* of the alignment. In the alignment, black boxes show identical amino acids between BBP and SF1, and gray boxes represent similar amino acids between the two proteins. The GXXG motif is shown in red letters. Secondary structure and contacts to the RNA are marked (s for side chain and m for main chain RNA contacts, cylinders for α -helices, and arrows for β -strands) based on the human SF1 NMR structure (Liu et al. 2001). An asterisk (*) above the alignment marks the important arginine (R) and lysine (K) discussed in Figure 6. The chimeras were named according to the region of the protein that was replaced with the yeast sequence and a y placed before the region name to represent yeast. For example, in the construct SF1-yKH, the yeast-conserved KH (cKH) region was exchanged for the homologous human cKH region. The second zinc knuckle of BBP was not used in the protein constructs for binding studies, and those amino acids are shown in a smaller light gray in the alignment.

SF1-297 and SF1-yZnK for mutant RNAs), consistent with previous research (Berglund et al. 1998b; Garrey et al. 2006).

The QUA2 domain is primarily responsible for the differences in specificity and affinity between BBP and SF1

In contrast to the other chimeras, SF1-yQUA2 binds the wild-type RNA and the mutant RNAs with dramatically altered affinity and specificity compared with SF1-297. The RNA binding properties of SF1-yQUA2 are very similar to those of BBP-296. Figure 4 (lanes 1–6) shows that SF1-

yQUA2 binds BP-22 with a K_d of 36 nM, which is more than 20-fold tighter than wild-type SF1-297 and only 12-fold weaker than BBP-296 (Fig. 1, lanes 1–6). Again, similar to BBP-296, affinity of this protein for the mutant RNAs is drastically reduced (Fig. 4A, lanes 7–12). Comparing the $\Delta\Delta G$ s for SF1-yQUA2 and BBP-296 (Fig. 4B; Table 1) shows that much of the enhanced binding by SF1-yQUA2 can be attributed to interactions with the C3 of the branchpoint sequence (11 KJ/mole for SF1-yQUA2 versus 1.6 KJ/mole for SF1-297) and U1, but other positions are affected as well (Table 1). These results clearly show that the QUA2 domain is important in modulating the specificity of these two proteins. The remaining differences in $\Delta\Delta G$ between SF1-yQUA2 and BBP-296 for the different RNAs are likely the result of subtle effects caused by amino acid differences in other regions of the protein.

To determine whether or not the addition of the yeast zinc knuckle (region 5) to SF1-yQUA2 would enhance affinity but not alter the specificity, the human zinc knuckle was replaced by the yeast zinc knuckle domain (SF1-yQUA2+yZnK). This change does indeed enhance binding and the K_d for BP-22 drops from 36 to 14 nM (data not shown; Table 1). The specificity of SF1-yQUA2 and SF1-yQUA2+yZnK for BP-22 RNA and the mutant RNAs remain approximately the same as shown by the similar $\Delta\Delta G$ s for these chimeras (Table 1), indicating that the zinc knuckle only affects the general affinity of the protein–RNA interaction.

Given the observations above, we expected that replacing the QUA2 domain in BBP with the QUA2 domain from SF1 should drastically reduce the affinity and specificity of this chimera (BBP-hsQUA2) for RNA. As expected, the affinity of BBP-hsQUA2 changed to a K_d of 496 nM for the BP-22 RNA, a 165-fold decrease in affinity, and the specificity of this protein was significantly reduced as well (Fig. 4A, lanes 13–18). Figure 4B shows that BBP-hsQUA2 displays the same relatively small differences in affinities for the variant branchpoint sequence RNAs as that of SF1-297 (Table 1), indicating similarly low-sequence specificity. These two chimeric proteins (SF1-yQUA2 and BBP-hsQUA2) demonstrate that the QUA2 domain plays a significant role in

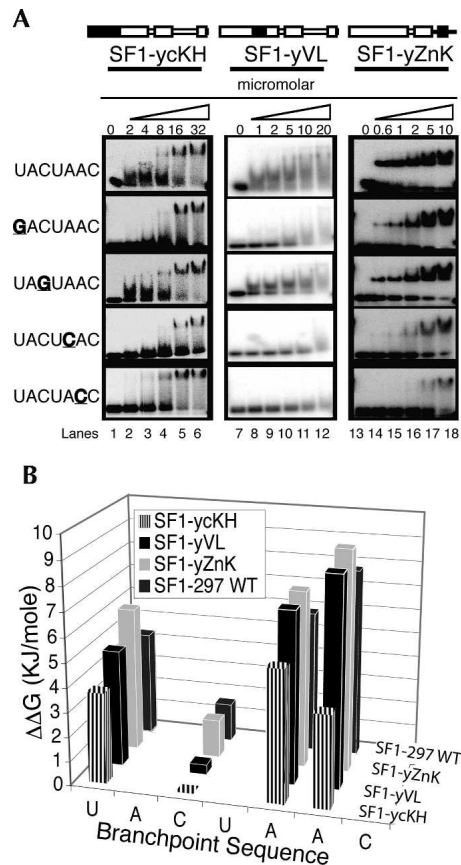


FIGURE 3. RNA binding specificities for three BBP-SF1 chimeras. (A) All three have a different yeast BBP region exchanged into the SF1-297 protein, the conserved KH (cKH) (lanes 1–6), the variable loop (VL) (lanes 7–12), and the yeast first zinc knuckle (ZnK) (lanes 13–18). RNA binding is in the micromolar range with concentrations listed above the gels. (B) A bar graph comparing the binding specificities ($\Delta\Delta G$) of SF1-297 wild-type protein and the three chimeras analyzed in A.

modulating the RNA binding affinity and specificity of both SF1 and BBP.

To again test the modular nature of the zinc knuckles, we replaced the first yeast zinc knuckle from BBP-hsQUA2 protein with the human zinc knuckle (BBP-hsQUA2+hsZnK). This change further reduced the binding affinity of the protein to all five RNAs from 1.5-fold to threefold (Table 1). The specificity ($\Delta\Delta G$ s) of this protein is no different compared with BBP-hsQUA2, indicating that the general affinity has been weakened but specificity has not been affected.

Transposing an arginine and lysine in the QUA2 domain endows SF1 with similar RNA binding characteristics to that of BBP

There are only five amino acid differences between the yeast and human QUA2 domains (see Fig. 2). To determine which of these amino acids are modulating the RNA binding we made individual (D235G, this numbering is

based on SF1 sequence) and double mutations (D235G/M242G and R240K/K241R) in human SF1. These mutations covered four of the five amino acid differences between the SF1 and BBP QUA2 domains (Fig. 2). The D235G mutation removed a negative charge. We hypothesized that this would enhance the ability of the QUA2 domain to fully engage in binding due to the possible removal of charge repulsion with the phosphate backbone of the RNA. However, our results showed that the SF1-D235G mutation bound RNA with only threefold greater affinity than SF1-297 (Table 1). Specificity changes for this mutant were very modest, with only the U1G mutation exhibiting an effect compared with wild-type SF1 ($\Delta\Delta G$ s in Table 1). The double mutation D235G/M242G behaved

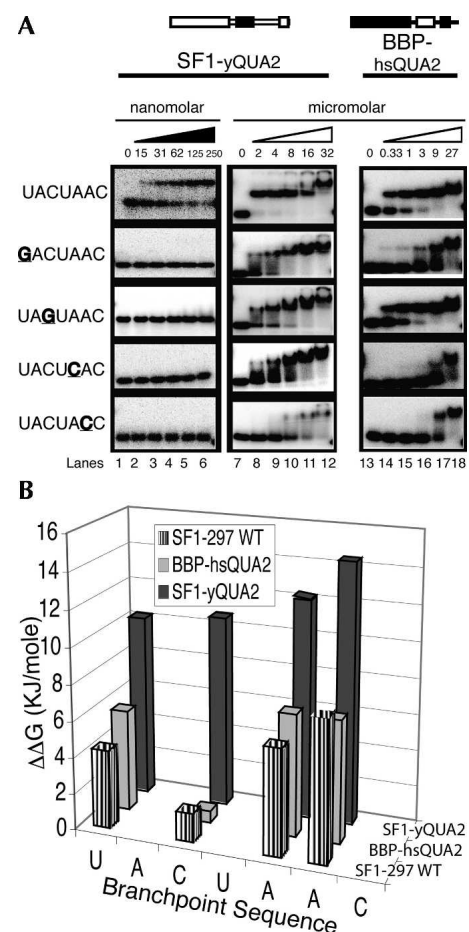


FIGURE 4. Exchanging the QUA2 domains of SF1 and BBP result in transformed RNA binding specificities and affinities. (A) Lanes 1–6 show gel-shift assays for SF1-yQUA2 in a low nanomolar concentration range (black triangle). For this construct, the yeast QUA2 domain replaces the human QUA2. Lanes 7–12 show SF1-QUA2 binding the RNAs with protein in the micromolar range. In lanes 13–18, BBP-hsQUA2 protein binds the RNAs with protein concentrations in the micromolar range. BBP-hsQUA2 is a BBP-296 protein in which the human QUA2 replaces the yeast QUA2. (B) A bar graph comparing the binding specificity ($\Delta\Delta G$) of SF1-297 wild-type protein and the two chimeras, SF1-yQUA2 and BBP-hsQUA2.

the same as the single mutation D235G, indicating that M242 does not play a role in RNA affinity or specificity (data not shown).

The R240K/K241R double mutation in human SF1-297 transposes adjacent arginine and lysine residues. Surprisingly, the exchange of these two similar amino acids has a significant effect on RNA binding specificity and affinity (Fig. 5). Similar to BBP-296 and SF1-yQUA2, the SF1-R240K/K241R protein binds with a K_d in the low nanomolar range to wild-type branchpoint RNA, and binds with drastically reduced affinity to the mutant RNAs (Fig. 5A, lanes 1–6). This indicates that R240 and K241 in human SF1 (or K252 and R253 in yeast) QUA2 domain are responsible for much of the differences in specificity and affinity between BBP and SF1.

Because the QUA2 domain makes significant contacts with A2 (Liu et al. 2001), this position was mutated to a cytosine to determine its effect on SF1-R240K/K241R binding. As shown in Figure 5, A and B, the A2C mutation does affect SF1-R240K/K241R binding (K_d changes from 40 to 360 nM), but not nearly as significantly as this mutation

disrupts BBP-296 binding (Fig. 5B, cf. $\Delta\Delta G$ s). While for SF1-297, the A2C mutation enhances binding slightly (K_d changes from 740 to 320 nM).

Single mutations in the QUA2 domain alter affinity and specificity of SF1

To determine the individual roles of R240 and K241 in the RNA binding of SF1 and BBP, these amino acids in SF1 were mutated individually to their yeast counterparts. Changing R240 to a lysine enhances binding 10-fold (K_d changes from 740 to 70 nM) (Fig. 6A), while changing K241 to an arginine enhances binding a little more than threefold (K_d changes from 740 to 200 nM) (Fig. 6A). With regard to specificity, Figure 6B shows that mutating A2 and C3 have a significant impact on the binding of SF1-R240K, while these RNA mutations have much less of an effect on the binding of SF1-K241R and SF1-297. These results suggest that both amino acids play a role in RNA binding and that the presence of an arginine at position 240 inhibits binding.

Fluorescence anisotropy

Fluorescent anisotropy measurements were used to confirm our gel-shift assay measurements, the BP-22 RNA and the A2C mutant RNAs were fluorescein labeled and used to measure the binding affinities of the SF1-297, BBP-296, and SF1-R240K/K241R proteins. Figure 7A shows that the binding affinity of the SF1-R240K/K241R protein is shifted closely to that of BBP-296, confirming that this transposed protein is binding BP-22 RNA much more like BBP-296 than SF1-297. The K_d s derived from these curves are 11 nM for BBP-296, 30 nM for SF1-R240K/K241R, and 600 nM for SF1-297, while the K_d s derived from the gel-shift experiments are 3 nM for BBP-296, 40 nM for SF1-R240K/K241R, and 740 nM for SF1-297. The measurements with the mutant RNA (A2C) show that all three proteins have reduced levels of binding; BBP-296 binds with a K_d of 250 nM, SF1-297 binds with a K_d of 620 nM, and SF1-R240K/K241R binds with a K_d of 270 nM (Fig. 7B). The measurements determined using the gel-shift assay and fluorescence anisotropy are in good agreement except for BBP-296, for which the K_d s are shifted over threefold for BP-22 RNA and fivefold for the A2C mutant RNA. The use of two different methods to measure the RNA binding affinities of the various proteins and the demonstration that all of the proteins are between 40% and 100% active (Supplemental Fig. 1) shows that R240 and K241 play an important role in determining the RNA binding affinity and specificities of BBP and SF1.

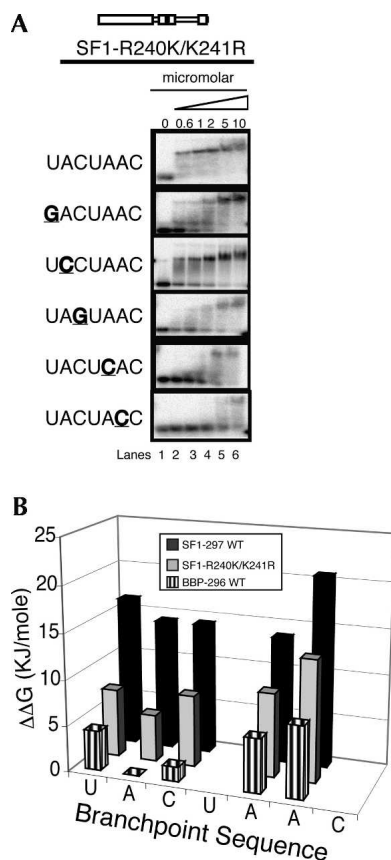


FIGURE 5. The transposition chimera, SF1-R240K/K241R binds the branchpoint sequence RNA like BBP. (A) Lanes 1–6 show binding in the micromolar concentration range for the mutant SF1 protein (triangle) to BP-22 RNA and five mutant RNAs. (B) A bar graph of the binding specificities ($\Delta\Delta G$) of BBP-296, SF1-297, and SF1-R240K/K241R.

DISCUSSION

Our results show that the QUA2 and zinc knuckle domains in the yeast and human branchpoint binding proteins can

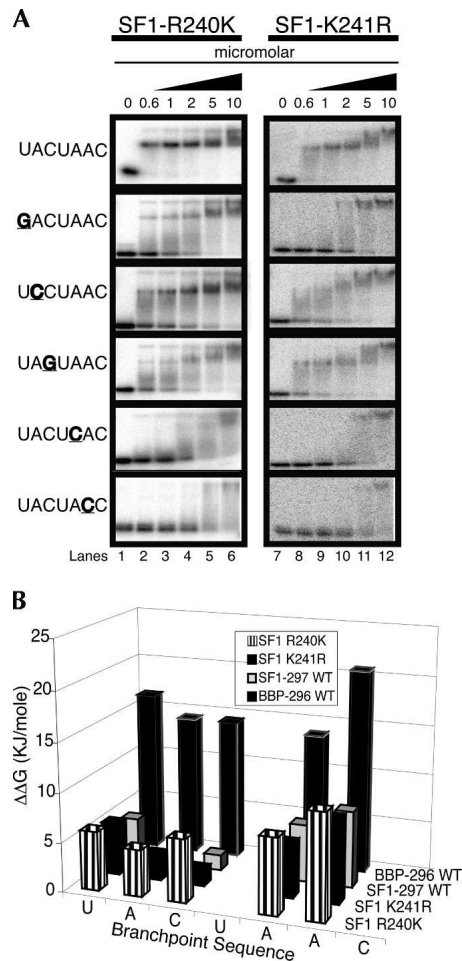


FIGURE 6. Binding of arginine-lysine 240 (R240K) and lysine-arginine 241 (K241R) mutant SF1 proteins to BP-22 RNA and RNA mutants. (A) Lanes 1–6 show binding in the micromolar concentration range for the mutant SF1 proteins (black triangle) to BP-22 RNA and five mutant RNAs. (B) A bar graph of the binding specificities ($\Delta\Delta G$) of BBP-296, SF1-297, SF1-R240K, and SF1-K241R.

be used as modular domains to achieve different levels of RNA binding affinities and specificities. Replacing the zinc knuckle of human SF1 with that of yeast BBP mildly increases the overall RNA binding affinity, but does not significantly alter RNA binding specificity, indicating that these zinc knuckles are general RNA binding modules. Exchanging the QUA2 domains of SF1 and BBP significantly alters the RNA binding affinity and specificity of these proteins so that they mimic those of the protein from which the QUA2 domain was derived. This shows that the QUA2 domain plays an important role in modulating the RNA binding of these two proteins. Furthermore, a transposition of two adjacent basic amino acids (R240 and K241) within the QUA2 domain is sufficient to transform a RNA binding protein with low affinity and specificity into a protein that binds RNA with high affinity and specificity.

The NMR structure of SF1 in complex with the yeast branchpoint sequence shows that the QUA2 domain interacts with the 5' end of the RNA (Fig. 8A), specifically A2, C3, and U4 (Liu et al. 2001), and that K241 sits on top of A2 (Fig. 8B). As one might predict, changing the lysine at position 241 to an arginine enhances binding, because favorable stacking interactions likely occur between the more planar side chain of arginine and the adenosine base, while lysine's conformations do not allow stacking interactions to occur (Flocco and Mowbray 1994). In the STAR/GSG family protein, GLD-1, it was found that the equivalent amino acid to K241 (K314) is important for RNA binding (Lehmann-Blount and Williamson 2005). Interestingly, it is R240 that has the most significant affect on binding, but R240 is distant from A2 and any other nucleotides based on

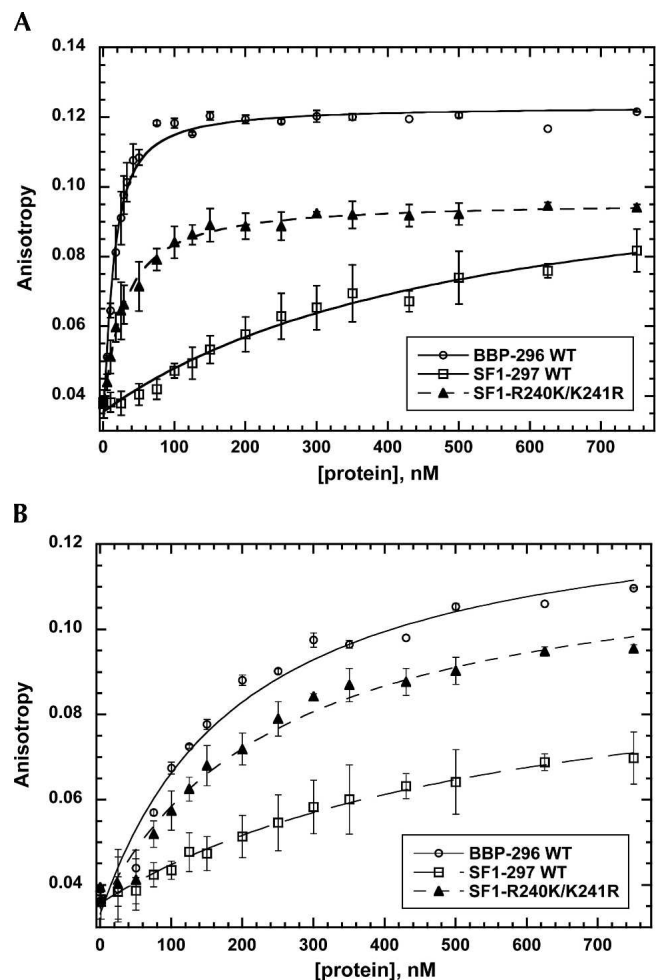


FIGURE 7. Fluorescence anisotropy of BP-22 RNA (A) and a mutant RNA (B) upon addition of BBP-296, SF1-297, or SF1-R240K/K241R. (A) As BP-22 RNA is titrated with BBP-296 (○), SF1-297 (□), or SF1-R240K/K241R (▲), increased anisotropy is observed, indicating protein–RNA binding. (B) Titration of BBP-296, SF1-297, or SF1-R240K/K241R with mutant RNA required higher concentrations of protein to achieve binding. No error bars are shown for points on the titration curve where a single measurement was taken.

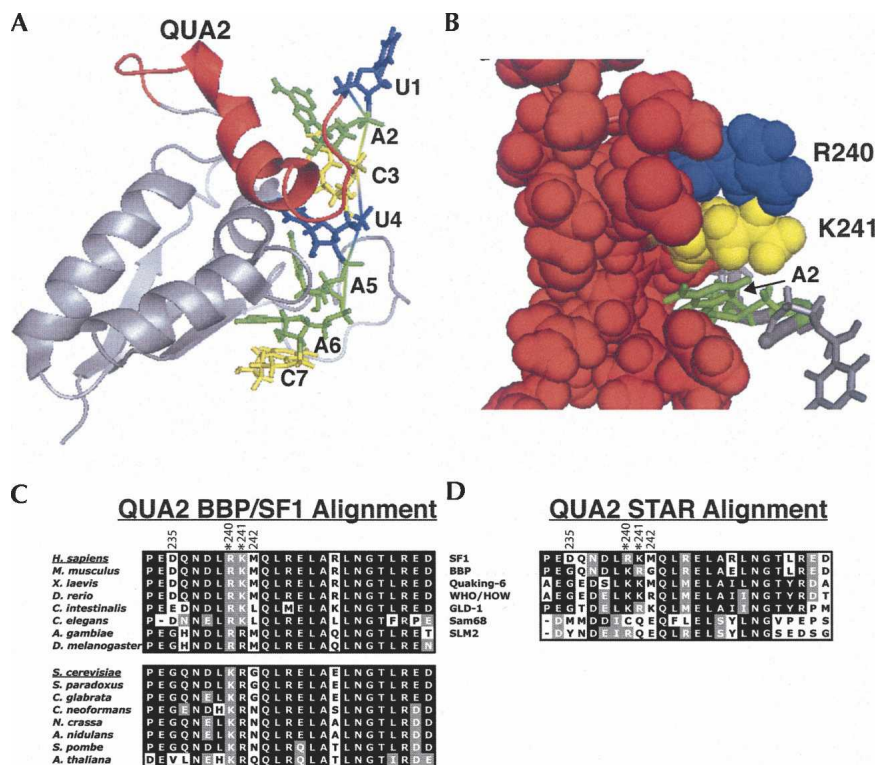


FIGURE 8. Analyzing the QUA2 domain structure from the NMR structure of SF1 in complex with the branchpoint sequence RNA (PDB #1K1G) (Liu et al. 2001) and alignments of branchpoint binding proteins from multiple organisms and the related STAR/GSG proteins. (A) The QUA2 domain (red) relative to the KH domain and branchpoint sequence. The nucleotides of the branchpoint sequence are numbered 1 through 7. (B) A spaced-filled view of R240 (blue) and K241 (yellow) and the rest of the QUA2 domain (red) binding RNA. The A2 nucleotide (green) fits in a pocket formed by the QUA2 domain. (C) An alignment of QUA2 domains of SF1, BBP, and branchpoint binding proteins from different organisms. (D) A QUA2 domain alignment of SF1, BBP, and other STAR/GSG proteins. Accession numbers for these proteins can be found in Supplemental Table 1.

the NMR structure (Fig. 8B). One possible interpretation of our data is that when R240 is mutated to a lysine, the QUA2 adopts a conformation facilitating additional interactions between the QUA2 and the RNA, thus, the observed enhanced binding affinity and specificity.

The U2 snRNA sequence used to base pair with the branchpoint sequence is highly invariant, yet the branchpoint sequence varies amongst organisms as do the RNA binding affinities and specificities of the branchpoint binding proteins, raising the intriguing possibility that the evolution of this protein and the branchpoint sequence are linked. Specifically, we propose that the identity of amino acids 240 and 241 coevolved with the branchpoint sequences. In general, fungal introns have branchpoint sequences that are more similar to the highly conserved UACUAAC branchpoint sequence of *Saccharomyces cerevisiae* introns (Bon et al. 2003), and the fungal branchpoint binding proteins all have lysine at position 240 and arginine at 241 (Fig. 8C). In contrast, the branchpoint sequences in vertebrate introns are less conserved (Keller and Noon 1984), and

the vertebrate branchpoint binding proteins have arginine at position 240 and lysine at 241 (Fig. 8C). This analysis suggests that having K240 and R241 in the branchpoint binding protein correlates to the UACUAAC branchpoint sequence, while the transposition (R240 and K241) correlates with the CURAY degenerate branchpoint sequence. The consensus branchpoint sequence in *Drosophila* is CUAUU (Mount et al. 1992), which might be considered an intermediate consensus sequence compared with the yeast and human branchpoint consensus sequences, suggesting that the presence of arginines at positions 240 and 241 in the insect branchpoint binding proteins might modulate the binding to match this CUAUU consensus sequence.

SF1 and BBP are members of the STAR/GSG family of RNA binding proteins, and analysis of the QUA2 domains from the other members of this group supports the model that one or both of these amino acids play an important role in RNA binding specificity. In the alignment of the QUA2 domains of STAR/GSG proteins (Fig. 8D), except for Sam68 and Slm2, these two amino acids are lysines and arginines. GLD-1 and BBP have the lysine and arginine in the same positions, and GLD-1, like BBP, binds the UACUAAC sequence with high affinity and specificity (Ryder and Williamson 2004; Lehmann-Blount and Williamson 2005). The quaking (QKI) protein binds YACUAAY (Ryder and Williamson 2004; Galarneau and Richard 2005), which is very similar to the UACUAAC site for BBP and has two lysines at positions 240 and 241. Sam68 binds to the consensus sequence of UAAA, which lacks the 5' nucleotides recognized by the other members of the STAR/GSG family of proteins (Lin et al. 1997). Sam68 does not contain a basic amino acid at either position 240 or 241 (Fig. 8D). These correlations between RNA binding sites and the identity of the amino acids at positions 240 and 241 suggest that these two amino acids may play an important role in tuning the affinity and specificity of the STAR/GSG proteins for their specific RNA targets.

MATERIALS AND METHODS

Plasmid construction

Our previously described SF1, here known as SF1-297, was used as a template to make all point mutants. Primers and templates used

to make each of the BBP/SF1 chimeras are shown in Supplemental Table 1. Those chimeras that took multiple PCR steps are divided into A, B, and C. For some chimeras with multiple PCR steps, the product of the first PCR reaction (A) was used as a primer for the second PCR reaction (B), or the product of the second PCR reaction (B) was used as a primer for the third PCR reaction (C). These are indicated under the primer sequence column as Product of A or Product of B. All final PCR fragments, once the last PCR step was completed, were cut with BamHI at the N terminus and EcoRI at the C terminus and cloned into pGEX-6P-1 (Pharmacia).

Protein purification and expression

BBP, SF1, and all other chimeras and mutant constructs were transformed into BL21(STAR) cells (Novagen). Cells were resuspended in 25 mM Tris (pH 7.5), 300 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.1% Triton X-100 (Sigma). Cells were sonicated and spun to pellet insoluble debris at 19,800g for 30 min.

The supernatant was added to glutathione-agarose beads (Sigma product# G 4510) for 15 min and washed three times with resuspension buffer (above). Beads were washed once with 1X cleavage buffer (50 mM Tris at pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, and 0.1% Triton X-100). The protein of interest was cleaved from GST using Precision protease (Pharmacia). All constructs were purified over an ion exchange column (Mono S HR 10/10) and dialyzed into protein storage buffer (25 mM Tris at pH 7.5, 150 mM NaCl, 1 mM DTT, and 15% glycerol). A concentration step was not needed.

The activities of all of the proteins were determined by titrating RNA against a constant amount of protein. We found that all of the proteins except SF1- $\gamma\beta 3\alpha 3$, which did not bind RNA to begin with, had activities in the 40%–100% range (Supplemental Fig. 1). Also, to compare the overall secondary structures of SF1-297 and BBP-296, as well as three of the mutants (SF1-R240K/K241R, SF1-R240K, and SF1-K241R), we used circular dichroism and found that all five of these proteins contained approximately the same percentage of secondary structure (data not shown).

Gel-shift assays

RNAs were labeled with [γ - 32 P]ATP using T4 polynucleotide kinase, purified on denaturing polyacrylamide gels, and cleaned up using Bio-Rad P6 spin columns. Binding buffer consisted of 25 mM Tris (pH 7.5), 50 mM NaCl, 2 mM DTT, 0.1% Triton X-100, and 0.07 mg/mL tRNA (Bakers yeast from Sigma). The radio-labeled oligonucleotide was at an approximate concentration of 0.1 nM. Free RNA and protein/RNA complexes were separated on 0.5X TBE 6% polyacrylamide gels run at 150 V for 1 h at 4°C. The gel-shift assay was repeated three times for each RNA substrate except for those indicated in Table 1. Gels were dried and exposed on a Storm 860 PhosphorImager (Amersham). Quantification of the fraction of RNA bound was done using ImageQuant. Data was plotted in Kaleidagraph and data points fit to the following equation to calculate an approximate K_d : Fraction RNA bound = $\frac{([K_d + [RNA_{Total}] + [SF1_{Total}]] - \sqrt{([K_d + [RNA_{Total}] + [SF1_{Total}]]^2 - 4([SF1_{Total}][RNA_{Total}])})}{2[RNA_{Total}]}$. To determine ΔG and $\Delta\Delta G$, the following formulas were used: $\Delta G = -RT \ln(1/K_d)$ and $\Delta\Delta G = -RT \ln(K_{d1}/K_{d2})$, K_{d1} is protein binding to the wild-type BP-22 RNA, and K_{d2} is protein binding to the mutant RNA in that column. For example, the $\Delta\Delta G$ in Table 1 for

BBP-296 WT and the UACUAAC mutant was calculated using the K_d for BBP-296 with UACUAAC and the K_d for BBP-296 with the mutant RNA, UACUACC.

Fluorescence anisotropy assays

Titration were performed using a Jobin Yvon Horiba FluoroMax-3 fluorescence spectrometer. 5'-Fluorescently labeled RNA oligonucleotides were obtained from IDT. The sequence of the wt RNA was 5'-fluorescein-CAGUAUACUACAAGUUGAAUU-3'. The sequence of the mutant RNA was 5'-fluorescein-CAGUAUA GUAACAA GUUGAAUU-3'. Excitation and emission wavelengths were 495 and 525 nm, respectively. The excitation and emission bandpasses were set to 3.5 nm, and the instrument was set at 1 nm resolution. Samples were maintained at 25°C with a recirculating water bath. Proteins were dialyzed into storage buffer (25 mM Tris at pH 7.5, 150 mM NaCl, 1 mM TCEP [Tris(2-carboxyethyl) phosphine hydrochloride, Fluka], 15% glycerol) containing TCEP instead of DTT, as DTT caused significant light scattering. RNA samples were prepared in a buffer containing 25 mM Tris (pH 7.5), 50 mM NaCl, and 0.07 mg/mL tRNA (Sigma). The fluorescent RNA oligonucleotide concentration was 10 nM. Protein was added to a final protein concentration of 750 nM. Data was plotted in Kaleidagraph and fit to the following equation to calculate an approximate K_d : Anisotropy = $A_0 + (A - A_0) \frac{([Protein_{Total}] + [RNA_{Total}] + K_d) - \sqrt{([Protein_{Total}] + [RNA_{Total}] + K_d)^2 - 4[RNA_{Total}][Protein_{Total}]}}{2[RNA_{Total}]}$. For each combination of BBP-296 WT or SF1-297 WT and RNA, three independent experiments were performed. The approximate K_d values were determined for each experiment, and the K_d values were averaged. The error reported is the standard deviation of the three measurements. For SF1-R240K/K241R, two experiments were performed and the approximate K_d values were averaged. The error reported is the standard deviation of the two measurements.

SUPPLEMENTAL DATA

Supplemental Materials can be obtained via email: aberglund@molbio.uoregon.edu.

ACKNOWLEDGMENTS

We thank Alice Barkan, Vickie DeRose, Rodger Voelker, and members of the Berglund laboratory for experimental advice and comments on the manuscript. This work was supported by a grant from NSF (0616264-MCB) to J.A.B.

Received May 14, 2007; accepted October 3, 2007.

REFERENCES

- Abovich, N. and Rosbash, M. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**: 403–412.
- Auweter, S.D., Oberstrass, F.C., and Allain, F.H. 2006. Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res.* **34**: 4943–4959. doi: 10.1093/nar/gkl620.
- Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**: 781–787.

- Berglund, J.A., Abovich, N., and Rosbash, M. 1998a. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branch-point region recognition. *Genes & Dev.* **12**: 858–867.
- Berglund, J.A., Fleming, M.L., and Rosbash, M. 1998b. The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA* **4**: 998–1006.
- Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuveglise, C., Munsterkotter, M., Guldener, U., Mewes, H.W., Van Helden, J., Dujon, B., et al. 2003. Molecular evolution of eukaryotic genomes: Hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* **31**: 1121–1135. doi: 10.1093/nar/gkg213.
- Burd, C.G. and Dreyfuss, G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**: 615–621.
- Cavaloc, Y., Bourgeois, C.F., Kister, L., and Stevenin, J. 1999. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**: 468–483.
- Chen, T. and Richard, S. 1998. Structure-function analysis of Qk1: A lethal point mutation in mouse quaking prevents homodimerization. *Mol. Cell. Biol.* **18**: 4863–4871.
- Chen, T., Damaj, B.B., Herrera, C., Lasko, P., and Richard, S. 1997. Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: Role of the KH domain. *Mol. Cell. Biol.* **17**: 5707–5718.
- Chow, L.C., Gelinis, R.E., Broker, T.R., and Roberts, R.J. 2000. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. 1977. *Rev. Med. Virol.* **10**: 362–371; discussion 355–366.
- Dannull, J., Surovoy, A., Jung, G., and Moelling, K. 1994. Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues. *EMBO J.* **13**: 1525–1533.
- De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**: 384–388.
- Flocco, M.M. and Mowbray, S.L. 1994. Planar stacking interactions of arginine and aromatic side-chains in proteins. *J. Mol. Biol.* **235**: 709–717.
- Galarneau, A. and Richard, S. 2005. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* **12**: 691–698.
- Garrey, S.M., Voelker, R., and Berglund, J.A. 2006. An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding. *J. Biol. Chem.* **281**: 27443–27453.
- Gibson, T.J., Thompson, J.D., and Heringa, J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. *FEBS Lett.* **324**: 361–366.
- Grishin, N.V. 2001. KH domain: One motif, two folds. *Nucleic Acids Res.* **29**: 638–643. doi: 10.1093/nar/29.3.638.
- Keller, E.B. and Noon, W.A. 1984. Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci.* **81**: 7417–7420.
- Kramer, A. 1988. Presplicing complex formation requires two proteins and U2 snRNP. *Genes & Dev.* **2**: 1155–1167.
- Langford, C.J. and Gallwitz, D. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* **33**: 519–527.
- Lehmann-Blount, K.A. and Williamson, J.R. 2005. Shape-specific nucleotide binding of single-stranded RNA by the GLD-1 STAR domain. *J. Mol. Biol.* **346**: 91–104.
- Lin, Q., Taylor, S.J., and Shalloway, D. 1997. Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains. *J. Biol. Chem.* **272**: 27274–27280.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**: 1098–1102.
- Lukong, K.E. and Richard, S. 2003. Sam68, the KH domain-containing superSTAR. *Biochim. Biophys. Acta* **1653**: 73–86.
- Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O., and Fields, C. 1992. Splicing signals in *Drosophila*: Intron size, information content, and consensus sequences. *Nucleic Acids Res.* **20**: 4255–4262. doi: 10.1093/nar/20.16.4255.
- Pikielny, C.W., Teem, J.L., and Rosbash, M. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: Implications for U1 RNA and metazoan mRNA splicing. *Cell* **34**: 395–403.
- Query, C.C., Moore, M.J., and Sharp, P.A. 1994. Branch nucleophile selection in pre-mRNA splicing: Evidence for the bulged duplex model. *Genes & Dev.* **8**: 587–597.
- Query, C.C., Strobel, S.A., and Sharp, P.A. 1996. Three recognition events at the branch-site adenine. *EMBO J.* **15**: 1392–1402.
- Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P., and Kramer, A. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* **4**: 551–565.
- Reed, R. 1996. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* **6**: 215–220.
- Ruskin, B., Greene, J.M., and Green, M.R. 1985. Cryptic branch point activation allows accurate in vitro splicing of human β -globin intron mutants. *Cell* **41**: 833–844.
- Ryder, S.P. and Williamson, J.R. 2004. Specificity of the STAR/GSG domain protein Qk1: Implications for the regulation of myelination. *RNA* **10**: 1449–1458.
- Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92**: 315–326.
- Vernet, C. and Artzt, K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**: 479–484.
- Wang, J., Tse, S.W., and Andreadis, A. 2007. Tau exon 6 is regulated by an intricate interplay of trans factors and cis elements, including multiple branch points. *J. Neurochem.* **100**: 437–445.
- Will, C.L. and Lührmann, R. 2006. Spliceosome structure and function. In *The RNA world*. (eds. R. Gesteland et al.), pp. 369–400. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Zhuang, Y.A., Goldstein, A.M., and Weiner, A.M. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc. Natl. Acad. Sci.* **86**: 2752–2756.