The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence

J. ANDREW BERGLUND,¹ MARGARET L. FLEMING, and MICHAEL ROSBASH

Howard Hughes Medical Institute, Departments of Biochemistry and Biology, Brandeis University, Waltham, Massachusetts 02254, USA

ABSTRACT

The yeast and mammalian branchpoint sequence binding proteins (BBP and mBBP/SF1) contain both KH domain and Zn knuckle RNA-binding motifs. The single KH domain of these proteins is sufficient for specific recognition of the pre-mRNA branchpoint sequence (BPS). However, an interaction is only apparent if one or more accessory modules are present to increase binding affinity. The Zn knuckles of BBP/mBBP can be replaced by an RNA-binding peptide derived from the HIV-1 nucleocapsid protein or by an arginine-serine (RS)₇ peptide, without loss of specificity. Only the seven-nucleotide branchpoint sequence and two nucleotides to either side are necessary for RNA binding to the chimeric proteins. Therefore, we propose that all three of these accessory RNA-binding modules bind the phosphate backbone, whereas the KH domain interacts specifically with the bases of the BPS. Proteins and protein complexes with multiple RNA-binding motifs are frequent, suggesting that an intimate collaboration between two or more motifs will be a general theme in RNA–protein interactions.

Keywords: BBP; BPS; mBBP/SF1; RNA-binding proteins; splicing; Zn knuckle

INTRODUCTION

The branchpoint sequence (BPS) is important for the chemistry of pre-mRNA splicing. The 2'OH of the underlined adenosine (see BPS below) attacks the 5' splice site to form the free 5' exon and lariat intermediate during the first step of splicing. The second step consists of attack by the free 5' exon at the 3' splice site to form the ligated exon products. These reactions take place within the spliceosome, a complex containing several small nuclear ribonucleoproteins plus approximately 50–100 additional protein factors (for reviews, see Moore et al., 1993; Madhani & Guthrie, 1994; Kramer, 1996).

In both yeast and mammals, spliceosome formation begins with intron recognition, which requires only U1 snRNP and a small number of proteins (review: Reed, 1996). One of these is the branchpoint sequence binding protein (yBBP), which is an essential yeast (*Saccharomyces cerevisiae*) splicing factor. yBBP recognizes

the 3' end of introns through specific binding to the BPS and through interactions with other proteins (Abovich & Rosbash, 1997; Berglund et al., 1997b). In yeast, the BPS is an almost invariant seven-nucleotide intronic sequence (UACUAAC), which is necessary for intron recognition and splicing (review: Rymond & Rosbash, 1992). The mammalian orthologue (mBBP) recognizes the more degenerate mammalian BPS (YNCURAY, from Keller & Noon, 1984), and achieves optimal binding through both RNA–protein and protein–protein interactions (Abovich & Rosbash, 1997; Berglund et al., 1997b, 1998). mBBP, also known as SF1 (Splicing Factor 1), was shown previously to be required for formation of the first ATP-dependent splicing complex (Kramer, 1992; Arning et al., 1996). Both yBBP and mBBP contain two types of RNA-binding motifs, a K homologous (KH) domain and one or two Zn knuckles. The KH domain (Fig. 1) was identified originally in the human heterogeneous ribonucleoprotein (hnRNP) K protein (Gibson et al., 1993; Siomi et al., 1993a). NMR structures reveal a $\beta\alpha\alpha\beta\beta\alpha$ secondary structure that adopts an $\alpha\beta$ fold (Musco et al., 1996, 1997). There are examples of KH domain-containing proteins interacting with RNA (Siomi et al., 1994; Chen et al., 1997; Lin et al., 1997), but it is not yet known which regions within the KH domain are involved in RNA binding.

Reprint requests to: Michael Rosbash, Howard Hughes Medical Institute, Departments of Biochemistry and Biology, Brandeis University, Waltham, Massachusetts 02254, USA; e-mail: rosbash@binah.cc.brandeis.edu.

¹Present address: Department of Chemistry & Biochemistry, University of Colorado, Boulder, Colorado 80309, USA.

FIGURE 1. Yeast and mammalian branchpoint sequence binding proteins and their RNA substrates. Schematic representation of the domains involved in RNA binding. Nucleotides in bold are important for binding; those in large type are the most important.

yBBP and mBBP also contain regions to either side of the KH domain that are conserved within a subset of KH-containing proteins (Arning et al., 1996; Ebersole et al., 1996; Vernet & Artzt, 1997). This family, termed STAR for signal transduction and activator of RNA, includes human Sam-68 and *Caenorhabditis elegans* Gld-1, both possibly involved in cell cycle progression (Fumagalli et al., 1994; Jones & Schedl, 1995; Barlat et al., 1997); the Qkl protein, which is responsible for the quaking phenotype in mice (Ebersole et al., 1996); and an hnRNP protein, Grp33, from *Artemia salina* (Cruz-Alvarez & Pellicer, 1987).

BBP contains a second RNA-binding motif. It is a Zn knuckle and is C-terminal to the KH domain, just after the conserved region. yBBP has two Zn knuckles, whereas mBBP has only one (Figs. 1, 2). The Zn knuckle, also known as a Zn finger in the retrovirus field, contains the sequence CX₂CX₄HX₄C. (C is cysteine, H is histidine, and X represents any amino acid.) The prototype for the Zn knuckle motif is the retroviral nucleocapsid (NC) protein (Rein, 1994; Darlix et al., 1995). Zn knuckles are also found in the splicing factors, yeast Slu7p, and the human SR protein 9G8 (Frank & Guthrie, 1992; Cavaloc et al., 1997). In the case of the HIV-1 NC protein, there are two Zn knuckle motifs. They are important for specific recognition of RNA packaging signals during virus encapsidation (Berkowitz et al., 1993, 1995; Berkowitz & Goff, 1994; Dannull et al., 1994; Clever et al., 1995; Berglund et al., 1997a). An NMR structure of an HIV-1-NC packaging signal protein–RNA complex reveals both base-specific and nonspecific (phosphate) backbone interactions (De Guzman et al., 1998).

We have shown previously that subregions of yBBP and mBBP containing the KH domain and Zn knuckles specifically bind the BPS, as summarized in Figure 1 (Berglund et al., 1997b). To determine if BPS specificity requires only one or multiple domains of yBBP, we tested different domains of yBBP with different accessory RNA-binding modules: two different forms of the KH domain, a fusion between the KH domain and a (nonspecific) basic peptide, a fusion between the KH domain and an arginine-serine (RS)₇ peptide, and a fusion of the Zn knuckles with the basic peptide. The results indicate that the KH domain contains all of the sequence specificity and that the Zn knuckles supply nonspecific general binding. Similar experiments were performed with mBBP, demonstrating that the KH domain of mBBP is also responsible for BPS specificity. Furthermore, we present evidence in support of a model showing that the KH domain interacts specifically with the BPS bases and the Zn knuckles with the BPS phosphate backbone.

RESULTS

Using a gel-shift assay, we showed that a region of yBBP containing all three RNA-binding domains (yBBP190) specifically recognizes the BPS in the context of a 22-nt RNA substrate (BP-22, caguaUACUAA Caaguugaauu) derived from the rp51a yeast intron (Berglund et al., 1997b). The strong homology between yBBP and mBBP within the KH domain (42% identity, 60% similarity, Fig. 2) suggested that it alone might be responsible for specific binding. To test this hypothesis, we expressed and purified a fragment of yBBP containing only the yKH domain (yBBP95, boxed in Fig. 1) and tested its ability to bind BP-22. yBBP95 was not able to bind RNA at the concentrations tested (up to 10 μM, data not shown). Because yBBP95 did not express to high levels in *Escherichia coli*, it seemed possible that yBBP95 might be missing a region involved in KH domain stabilization.

The region C-terminal to the KH domain is quite similar between yBBP and mBBP; a comparison of the next 30 amino acids reveals a similarity of 83% between the two proteins (Fig. 2). This conservation was noted previously within the STAR family of KH-containing proteins (Ebersole et al., 1996; Vernet & Artzt, 1997). A fragment containing these additional amino acids, yBBP125, was made and purified to homogeneity (Materials and Methods). yBBP125 expressed to much higher levels than yBBP95, suggesting that the addition of these amino acids does indeed stabilize the KH domain, at least in *E. coli*. yBBP125 still did not bind RNA in a gel shift assay (highest concentration tested was 50 μM). Although it was positive with a filter binding assay, there was only a small difference in affinity between BP-22 and a mutant form of BP-22, in which the branchpoint adenosine is changed to cytidine

FIGURE 2. Sequence alignment of yBBP and mBBP. Only the regions involved in RNA binding are shown. Conserved amino acids between the two proteins are in black. The defined RNA-binding domains (KH and Zn knuckles) are boxed and labeled. Numbering to the right is derived from the full-length yBBP. All KH-containing clones begin at the start of the KH domain as marked by the star. The region between the stars for both yBBP and mBBP is responsible for specificity and corresponds to yBBP125 and mBBP125. yBBP190 is derived from amino acids 145–330, which starts at the beginning of the yKH domain and ends after the second Zn knuckle. Alignment was done using the GCG software package (Devereux et al., 1984).

(UACUAAC to UACUAcC, change is in lowercase; data not shown). This same mutation reduced binding 125-fold for yBBP190 (Berglund et al., 1997b). Because yBBP125 binds RNA weakly, it is difficult to distinguish between specific and nonspecific binding. The weak binding of yBBP125 indicates that the Zn knuckles contribute to the affinity of BBP for the BPS.

There are three possible roles for the Zn knuckles: (1) specific recognition of the BPS; (2) improvement of binding through nonspecific RNA interactions; and (3) specific protein–protein interactions with the KH domain. In an attempt to distinguish between these possibilities, we substituted for the Zn knuckles in yBBP190 a completely different motif, a basic peptide derived from the HIV-1 nucleocapsid (NC) protein (see Fig. 4 legend for the sequence). This basic peptide can function as an RNA chaperone and apparently recognizes RNA nonspecifically (de Rocquigny et al., 1992). The fusion protein, yBBP125-NC, contains the KH domain, the highly conserved subsequent amino acids and then

29 amino acids from the NC protein (Materials and Methods, see Fig. 4 for a schematic representation).

yBBP125-NC specifically bound the BPS and did so with similar specificity and only slightly weaker affinity than yBBP190. This is demonstrated by strong binding of yBBP125-NC to BP-22 with little or no binding to BP-22 mutants in which the UACUAAC is mutated (Fig. 3A and data not shown). Point mutations at every position within the BPS either decreased or eliminated yBBP125-NC binding. Mutating the branchpoint adenosine to any other nucleotide completely eliminated binding, as did changing the uridine at the fourth position to an adenosine. Whenever one of the other five positions within the BPS was mutated, yBBP125-NC binding was reduced 40–100-fold, depending on the position. Thus, the overall specificity of yBBP125-NC agrees well with published data for yBBP190, which contains both the KH domain and Zn knuckles (Berglund et al., 1997b). However, the specificity of the two proteins is not identical. For yBBP190, different mutations within the BPS

FIGURE 3. Binding of various fusion proteins demonstrates that BPS specificity is within the KH domain. Panels **A–D** show the different proteins binding to the wild-type BPS (BP-22) and to point mutations within the BPS. Mutations are in bold and underlined. Protein concentrations are shown above each autoradiograph and decrease in threefold increments.

affect binding from 5- to 125-fold (Berglund et al., 1997b), whereas for yBBP125-NC, the same mutations affect binding by at least 40-fold, i.e., mutations with weak effects on yBBP190 binding have stronger effects on yBBP125-NC binding (see Discussion).

To test if another very different basic peptide could also rescue the RNA-binding affinity of the yBBP KH domain, we made a fusion of the KH domain to an arginine–serine peptide of 14 amino acids (RS)₇, termed yBBP125-RS. We choose this basic peptide because it has been shown to interact in a nonspecific but functional manner with the mammalian BPS, when attached to the mammalian splicing factor U2AF65 (Valcárcel et al., 1996). yBBP125-RS binds RNA well, and mutating any one of the seven positions within the BPS either greatly reduces or eliminates yBBP125-RS binding (Fig. 3B and

data not shown). The specificity of yBBP125-RS is therefore quite similar to that of yBBP125-NC. Although we cannot rule out the possibility that the KH domain of yBBP125 is misfolded when expressed alone, this appears unlikely based on the ability of either the NC or RS domain to rescue yBBP-125 RNA binding.

To be certain that the Zn knuckles of yBBP do not contribute measurably to specific binding, the region of yBBP containing the Zn knuckles was fused to the NC basic peptide (yBBPZn₂-NC), expressed, purified, and tested for specific binding. yBBPZn₂-NC bound RNA but without detectable specificity (data not shown). The results of the different protein constructs (summarized in Fig. 4) indicate that the KH domain of yBBP contains all of the specificity for the BPS, and the Zn knuckles supply nonspecific binding affinity.

For the mammalian orthologue mBBP, we had previously shown that a subregion containing the KH domain and Zn knuckle (mBBP182) is sufficient for specific binding to the BPS (Berglund et al., 1997b). To determine if the mammalian KH domain is also solely responsible for specificity, we used the same approach:

		Binding	Specificity
yBBP KH			
yBBP190		++	++ (y)
yBBP95		-	-
yBBP125		+	-/+
yBBP125-NC		++	++ (y)
yBBP125-RS		++	++ (y)
yBBP Zn₂			
yBBPZn ₂ -NC		+	-
mBBP KH			
mBBP182		++	++ (m)
mBBP110-yZn ₂ (chimera)		++	++ (m)
mBBP125-NC		++	++ (m)

FIGURE 4. Summary of the RNA-binding properties of the different yeast and mammalian proteins. Two pluses [++] in the binding column represents binding observed only in a gel-shift assay, one plus [+] represents binding observed in a filter binding assay, and a minus [–] represents no binding. Two pluses in the specificity column represents specific binding and is marked with either a (y) or (m) indicating specificity similar to the yBBP190 or mBBP182, the [–/+] represents a decreased level of specific binding, and the [–] represents no specific binding. Data for yBBP190 and mBBP182 are from Berglund et al. (1997b). The amino acid sequence of the NC peptide is VKGGRAPRKKGGERQANFLGKIWPYKGR (basic amino acids are in bold).

we constructed a fusion of the NC peptide to the KH domain of mBBP (mBBP125-NC), as well as a fusion containing the mammalian KH domain and the yeast Zn knuckles (mBBP110-yZn₂). Binding experiments were performed using the yeast BPS RNA substrate, because it has been shown to be an optimal BPS for mammalian splicing (Zhuang et al., 1989). This substrate was also used for the binding studies with mBBP182 (Berglund et al., 1997b). Both mBBP125-NC and mBBP110-yZn₂ bind with similar specificity to that of mBBP182: mutations at the branchpoint adenosine or adjacent uridine affect binding, whereas mutations at other positions within the BPS have little or no effect (Fig. 3C,D and data not shown). We conclude that the KH domain of mBBP also contains all of the information for specific BPS recognition.

The fact that different protein structures can complement the Zn knuckle deficiency indicates that the latter probably does not function through specific protein–protein contacts with the KH domain. The Zn knuckle probably helps binding through nonspecific interactions with the RNA, and this could be accomplished in different ways. The Zn knuckles could interact with the BPS-adjacent nucleotides in a sequence-independent manner, either to the phosphate backbone or to the bases. Alternatively, the Zn knuckles could bind to the phosphate backbone of the BPS itself (Fig. 6A). Supporting this second possibility is a footprinting experiment with yBBP190, showing that only the BPS and one or two adjacent nucleotides are protected from nuclease cleavage (Berglund et al., 1997b). This suggests that yBBP might bind well to a short RNA containing only the BPS and one or two nucleotides to either side. Figure 5 shows that an 11-nt RNA (uaUAC UAACaa; the BPS is in uppercase) binds yBBP190 almost as strongly as the 22-nt RNA used in the previous experiments. Taken together with the footprinting (Berglund et al., 1997b), the results suggest that yBBP binds the BPS as shown in the model (Fig. 6A). This model is not limited to this protein, because the other constructs also bind the 11-nt RNA substrate with nearly the same affinity as the 22-nt substrate (Fig. 5B,C and data not shown). Of course, this highly reductionist model ignores many other protein–protein and protein–RNA (and perhaps RNA–RNA) contacts that are likely relevant to BPS binding within the biological context of complete splicing complexes.

DISCUSSION

Both yBBP and mBBP have multiple RNA-binding modules. The single KH domains contribute most and perhaps all of the sequence specificity. Although this region of the yeast and mammalian proteins manifests a difference in specificity, we have not been able to identify regions within the KH domain that are responsible for this difference or more generally for base-specific rec-

FIGURE 5. An 11-nt RNA substrate binds with an affinity similar to that of a 22-nt substrate. Protein concentrations are shown above each autoradiograph and decrease in threefold increments.

ognition (data not shown). The Zn knuckles, in contrast, appear to be nonspecific RNA-binding modules (Fig. 6). There is no *in vitro* RNA binding without the Zn knuckles, indicating that both proteins require more than a single KH domain. Although this could reflect *in vitro* protein folding difficulties, many other RNA-binding proteins require multiple RNA-binding domains or bind to RNA as members of multiprotein complexes. For the STAR family of KH domain-containing proteins, homo-

FIGURE 6. Model for the interactions between yBBP (and mBBP) and its cognate RNA substrate. See text for details.

dimerization and oligomerization appear to be important for RNA binding (Chen et al., 1997; Zorn & Krieg, 1997). For the fragile X mental retardation gene product (FMR1), as well as the hnRNP K protein, deletion or mutational inactivation of a single domain strongly inhibits RNA binding, suggesting that the interaction normally requires two or more modules (Siomi et al., 1993b, 1994). Our results and others therefore suggest that single KH domains bind only weakly to RNA. The third KH domain of neuronal RNA-binding protein, Nova-1, may be an exception; it has been shown to bind an *in vitro*-selected RNA with a K_D in the low nanomolar range (Buckanovich & Darnell, 1997).

Although the Zn knuckles of yBBP and mBBP appear to promote RNA binding by providing nonspecific affinity, other Zn knuckles have been shown to interact with RNA in a sequence-specific manner. The HIV-1 NC protein is well-studied and binds several specific elements within the HIV RNA packaging element with similar affinity (Clever et al., 1995), suggesting that the sequence specificity is compatible with multiple RNA target interactions. An NMR structure of NC with one of the packaging elements shows how NC is able to bind RNA, through base-specific contacts and several phosphate-backbone interactions, respectively (De Guzman et al., 1998). In the case of yBBP and mBBP,

our results are best explained by proposing comparable Zn knuckle-backbone nonspecific interactions. Indeed, binding of yBBP125-NC and yBBP125-RS to the BBP-22 RNA substrate is salt-sensitive (data not shown). However, we cannot exclude that the Zn knuckles may make some contribution to specificity, for example, by aiding the positioning of the BPS bases and enhancing the KH domain-BPS interaction. This is similar to the proposal of Valcárcel et al. (1996), in which the RS-domain of U2AF65 aids U2 snRNP-BPS base pairing. A Zn knuckle contribution to specificity might be relevant to the somewhat different levels of sequence specificity of yBBP190 versus yBBP125-NC/yBBP125-RS (5–125-fold versus 40- to >125-fold, respectively), which could also reflect an intramolecular Zn knuckle-KH domain protein-protein interaction. But it is equally likely that these modest differences in specificity reflect the artificial nature and somewhat arbitrary design of the yBBP125-NC and yBBP125-RS fusion proteins.

We have shown that it is possible to increase RNA-binding affinity by adding nonnative amino acids to the KH domain. In the case of the well-studied U1A N-terminal RBD or RRM, the addition of seven native C-terminal amino acids increased binding to its cognate RNA stem-loop 30-fold without affecting specificity (Nagai et al., 1990; Hall 1994). Two of the additional amino acids are lysines, and it was proposed that they increase binding through electrostatic interactions (Nagai et al., 1990; Hall, 1994). The co-crystal structure revealed that the two lysines do not make any direct contacts, but probably help position the RNA (Nagai et al., 1994). Therefore, it is entirely possible that the Zn knuckles, NC and RS domains bind to RNA through interactions that are not strictly ionic. These include an effect on positioning, hydrogen bonding, or arginine base stacking interactions, as has been observed in a non-specific protein-ssDNA complex (Bochkarev et al., 1997).

yBBP binds RNA with better affinity than mBBP (Berglund et al., 1997b), 0.5 mM compared to 30 mM. This difference is probably due, at least in part, to the presence of two Zn knuckles in yBBP; this is indicated by the stronger binding of the artificial mBBP110-yZn₂ (approximate K_D = 8 mM) compared to 30 mM for the natural mBBP182 with its single Zn knuckle (Berglund et al., 1997b). Yet, all of these values are high, indicating that both yBBP and mBBP bind RNA with weak affinity compared to many other RNA-binding proteins, i.e., mM instead of nM. Weak binding might be important to facilitate replacement of BBP with U2 snRNP at the next step of the spliceosome assembly process.

Binding *in vivo* is almost certainly stronger than what is measured *in vitro*, because additional factors within the spliceosome probably help mBBP and yBBP bind to their BPS substrates. For mBBP, one of these factors is U2AF65. It interacts with mBBP, and the two proteins bind cooperatively to an RNA substrate con-

taining the two adjacent sites: a BPS and a polypyrimidine (PY) tract (Fig. 6B; Berglund et al., 1998). U2AF65 is an essential mammalian splicing factor and contains four domains, an amino-terminal RS domain and three RRM or RBD RNA-binding motifs. The RBDs interact with the PY tract (Zamore et al., 1992), and the RS domain with the BPS (Valcárcel et al., 1996). It is therefore tempting to speculate that the U2AF65 RS domain interacts predominantly with the BPS phosphate backbone and serves as a *trans*-acting affinity module for mBBP binding, similar to the observed enhancement with the covalently attached Zn knuckles or RS domain.

The formation of multiprotein RNA-binding complexes more generally reduces the distinction between covalent and noncovalent associations with affinity modules. In this context, features of the two comparable yeast proteins (Mud2p and yBBP) are relevant. They also act at the BPS region early in yeast intron recognition. Mud2p is a possible yeast orthologue of U2AF65, and the C-terminal RBDs of the two proteins are conserved. It is through this domain that Mud2p and U2AF65 interact with yBBP and mBBP, respectively (Berglund et al., 1998; Rain et al., 1998). Mud2p does not contain the conserved RS domain characteristic of metazoan and even *Schizosaccharomyces pombe* U2AF65. But yBBP contains two Zn knuckles rather than the single motif present on mBBP, providing each two-protein system with two affinity modules. Moreover, Mud2p lacks the first two RNA-binding domains present on U2AF65 (Abovich et al., 1994). If Mud2p binds RNA more weakly than U2AF65, stronger RNA binding by yBBP with its two Zn knuckles might be necessary to achieve *in vivo* BPS recognition.

Not only do most KH domains appear to function in collaboration with other KH domains or other RNA-binding modules, but many RNA-binding proteins with the RBD or RRM motif also contain either multiple copies of this motif or other RNA-binding modules (Birney et al., 1993). Like yBBP and mBBP, one domain may contribute most of the sequence specificity. Alternatively, some RNA-binding proteins may use their multiple domains to bind multiple RNA targets, at the same time or sequentially. If these RNA targets are close in space, it might be quite difficult to detect *in vitro* RNA binding of an individual domain to a single target, similar to our observations with the BBP KH domain. It is therefore likely that other single-motif protein–RNA interactions will also be too weak to study easily. These considerations suggest that the use of a covalently added basic peptide (NC29) or its equivalent will be of general utility.

MATERIALS AND METHODS

RNA synthesis, labeling, and purification

All RNA substrates were made using Perseptive Expedite RNA amidites on an Expedite 8909 Oligonucleotide Synthe-

sizer. RNA oligonucleotides were deprotected as recommended by the manufacturer. 5' End-labeling was performed with [γ - 32 P]ATP and T4 polynucleotide kinase. Radiolabeled RNAs were purified on 20% denaturing acrylamide gels, followed by Bio-Rad P6 spin columns.

Plasmid construction and protein purification

All protein constructs (yBBP190, yBBP95, yBBP125, yBBP125-NC, yBBP125-RS, yBBPZn₂-NC, mBBP110-yZn₂, and mBBP125-NC) were made with the pGEX-6P-1 vector (Pharmacia), a GST-fusion construct. Each protein was cleaved from GST while bound to glutathione sepharose with Precision protease under conditions recommended by Pharmacia. After cleavage and elution from the beads with 25 mM Tris, pH 7.5, and 200 mM NaCl, proteins were dialyzed against 25 mM Tris, pH 7.5, and 25 mM NaCl. Protein was then bound to CM sepharose and eluted with a gradient of NaCl, 25–500 mM. Peak fractions were concentrated using an Amicon ultrafiltration cell and then dialyzed against 25 mM Tris, pH 7.5, 50 mM NaCl, 1 mM DTT, and 15% glycerol and stored at -20°C . This purification scheme resulted in better than 95% pure protein for all constructs. All constructs have four additional amino acids from the GST cleavage site. The primers used to make the different constructs: for yBBP190 (amino acids 145–330), 5'-GGGGGATCCAAATTCAGGACAA GTATTAT (KH common oligo) and 5'-GGGGAATTCTCAAG AATTATTTACCGTGGC; for yBBP95 (amino acids 145–240), the KH common oligo and 5'-GGGGAATTCTCATTTGATAA CAATGTTTTG; for yBBP125 (amino acids 145–269), the KH common oligo and 5'-GGGGAATTCTCAATCTTCTCTTAA GTACC. Construction of yBBP125-NC was more complicated because of the 29-amino acid addition to the carboxy terminus of KH-128 (amino acids 145–271 of BBP plus the 29-amino acid basic peptide). This was done with two PCR reactions, the first added the template for the first 15 amino acids of the basic peptide, and the second reaction added the other 14 amino acids. The first PCR reaction used the KH common oligo and 5'-TTGTCTCTCGCCGCCTTTTTTCT TGGTGCTCTACCGCCTTTTACCCTGTTATCTTCTCTTAA AGT; the second PCR reaction used the product from the first reaction as template, and used again the KH common oligo (sequence as above) and 5'-GGGGAATTCTCATCTTCTCTT GTAGCTTGCCAGATTTTGCCTAAAAAATTAGCTTGCT CTGCGCCGCTTT (NC common oligo) to generate the coding region for BBP-NC. To construct yBBP125-RS, we used the following oligos: the KH common oligo (see above sequence) and 5'-GGGAGATTCTCAAGAGCGGGAACGGCTGCCA GAACGAGAGCGGGAGCGGCTACGCCTGTTATCTTCTC TTAA. The construction of yBBPZn₂-NC was similar to YBBP125-NC, in that two PCR reactions were again performed to attach the NC peptide to the region of BBP containing the Zn knuckles (amino acids 266–330 of BBP plus the 29 amino acid basic peptide). In the first PCR reaction, the oligos were 5'-GGGGGATCCACTTTAAGAGAAGATAA CAGG (Zn common oligo) and 5'-TTGTCTCTCGCCGCCT TTTTTCTTGGTGCTCTACCGCCTTTTACAGAATTATTAC GCTGGCGTT; the second PCR reaction used the Zn common oligo and NC common oligo to generate the coding region for yBBPZn₂-NC. To construct mBBP125-NC, we also used two PCR reactions. The first consisted of 5'-GGGGGATCCACAC GTGTGAGTGATAAAGTC and 5'-TTGTCTCTCGCCGCCTT

TTTTCTTGGTGCTCTACCGCCTTTTACCCTGTTATCGTC TTCCCGAAG. The second PCR reaction (template from the first reaction) used 5'-GGGGGATCCACACGTGTGAGTGAT AAAGTC and the NC common oligo (sequence as above). The template for mBBP110-yZn₂ was a chimeric protein of mBBP and yBBP (SF1-BBP/SST chimera, gift of Nadja Abovich). The SST site is between the KH domain and Zn knuckles at amino acid 246 in mBBP and is in the highly conserved region C-terminal to the KH domain. The oligos 5'-GGGGGA TCCCGTGTGAGTGATAAAGTCATG and 5'-GGGGAATTCT CAAGAATTATTTACCGTGGC were used to generate the PCR product for mBBP110-yZn₂. All PCR products were digested with *Bam*H I and *Eco*R I and inserted into pGEX-6P-1.

Gel-shift assay

Purified protein and RNA were incubated in binding buffer (25 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, and 30 mM yeast tRNA) for 30 min before being separated in a 7.5% 0.5× TBE polyacrylamide gel. This was done at 4 °C and run at 100 volts for approximately 2.5 h. Several experiments used binding buffer without EDTA and no apparent difference in affinity or specificity was observed.

ACKNOWLEDGMENTS

We thank members of the Rosbash and Schultz laboratories for advice and discussion, Melissa Moore and Tom Cech for reviewing the manuscript, and Nadja Abovich for supplying the DNA template (SF1-BBP/SST chimera) for the construction of the plasmid overexpressing mBBP110-yZn₂. We also thank L.A. Monaghan and E. Dougherty for help with manuscript and figure preparation. This work was supported by a grant from the NIH (GM-23549) to M.R.

Received March 27, 1998; returned for revision April 10, 1998; revised manuscript received June 1, 1998

REFERENCES

- Abovich N, Liao XC, Rosbash M. 1994. The yeast MUD2 protein: An interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes & Dev* 8:843–854.
- Abovich N, Rosbash M. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89:403–412.
- Arning S, Gruter P, Bilbe G, Kramer A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 2:794–810.
- Barlat I, Maurier F, Duchesne M, Guitard E, Tocque B, Schweighoffer F. 1997. A role for Sam68 in cell cycle progression antagonized by a spliced variant within the KH domain. *J Biol Chem* 272:3129–3132.
- Berglund JA, Abovich N, Rosbash M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes & Dev* 12:858–867.
- Berglund JA, Charpentier B, Rosbash M. 1997a. A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res* 25:1042–1049.
- Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. 1997b. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89:781–787.
- Berkowitz RD, Goff SP. 1994. Analysis of binding elements in the human immunodeficiency virus type 1 genomic RNA and nucleocapsid protein. *Virology* 202:233–246.
- Berkowitz R, Luban J, Goff S. 1993. Specific binding of HIV type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J Virol* 67:7190–7200.
- Berkowitz RD, Ohagen A, Hoglund S, Goff SP. 1995. Retroviral nucleocapsid domains mediate the specific recognition of genomic viral RNAs by chimeric gag polyproteins during RNA packaging in vivo. *J Virol* 69:6445–6456.
- Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 21:5803–5816.
- Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L. 1997. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* 385:176–181.
- Buckanovich RJ, Darnell RB. 1997. The neuronal RNA binding protein nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol Cell Biol* 17:3194–3201.
- Cavaloc Y, Popielarz M, Fuchs JP, Gattoni R, Stevenin J. 1997. Characterization and cloning of the human splicing factor 9G8: A novel 35 kDa factor of the serine/arginine protein family. *EMBO J* 13:2639–2649.
- Chen T, Damaj BB, Herrera C, Lasko P, 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.
- Clever J, Sassetti C, Parslow TG. 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type1. *J Virol* 69:2101–2109.
- Cruz-Alvarez M, Pellicer A. 1987. Cloning of a full-length complementary DNA for an *Artemia salina* glycine-rich protein. Structural relationship with RNA binding proteins. *J Biol Chem* 262:13377–13380.
- Dannull J, Surovoy A, Jung G, 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.
- Darlix JL, Lapadat-Tapolsky M, de Rocquigny H, Roques BP. 1995. First glimpses at structure–function relationships of the nucleocapsid protein of retroviruses. *J Mol Biol* 254:523–537.
- De Guzman RN, Wu ZR, Stalling CC, Pappalardo L, Borer PN, Summers MF. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 ψ -RNA recognition element. *Science* 279:384–388.
- de Rocquigny H, Gabus C, Vincent A, Fournie-Zalusky MC, Roque B, Darlix JL. 1992. Viral RNA annealing activities of HIV-1 nucleocapsid protein require only peptide domains outside the zinc fingers. *Proc Natl Acad Sci USA* 89:6472–6476.
- Devereux J, Haeblerli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Ebersole TA, Chen Q, Justice MJ, Artzt K. 1996. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genetics* 12:260–265.
- Frank D, Guthrie C. 1992. An essential splicing factor, *SLU7*, mediates 3' splice-site choice in yeast. *Genes & Dev* 6:2112–2124.
- Fumagalli S, Totty NF, Hsuan JJ, Courtneidge SA. 1994. A target for Src in mitosis. *Nature* 368:871–874.
- Gibson TJ, Thompson JD, Heringa J. 1993. The KH domain occurs in a diverse set of RNA-binding proteins that include the anti-terminator NusA and is probably involved in binding to nucleic acids. *FEBS Lett* 324:361–366.
- Hall KB. 1994. Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochemistry* 33:10076–10088.
- Jones AR, Schedl T. 1995. Mutations in *gld-1*, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in src-associated protein Sam68. *Genes & Dev* 9:1491–1504.
- Keller EB, Noon WA. 1984. Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc Natl Acad Sci USA* 81:7417.
- Kramer A. 1992. Purification of splicing factor SF1, a heat-stable protein that functions in the assembly of a presplicing complex. *Mol Cell Biol* 12(10):4545–4552.
- Kramer A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* 65:367–409.
- Lin Q, Taylor SJ, 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.
- Madhani HD, Guthrie C. 1994. Dynamic RNA–RNA interactions in the spliceosome. *Annu Rev Genet* 28:1–26.
- Moore MJ, Query CC, Sharp PA. 1993. Splicing of precursors to mRNAs by the spliceosome. In: Gesteland RF, Atkins JF. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 303–357.
- Musco G, Kharrat A, Stier G, Faternali F, Gibson TJ, Nilges M, Pastore A. 1997. The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nature Struct Biol* 4:712–716.
- Musco G, Stier G, Joseph C, Antonietta M, Morelli C, Nilges M, Gibson TJ, Pastore A. 1996. Three-dimensional structure and stability of the KH domain: Molecular insights into the fragile X syndrome. *Cell* 85:237–245.
- Rain JC, Rafi Z, Rhani Z, Legrain P, 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.
- Rein A. 1994. Retroviral RNA packaging: A review. *Arch Virol* 9:513–522.
- Rymond BC, Rosbash M. 1992. Yeast pre-mRNA splicing. In: Jones EW, Pringle JR, Broach JR. *The molecular and cellular biology of the yeast Saccharomyces: Gene expression*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 143–192.
- Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G. 1994. Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* 77:33–39.
- Siomi H, Matunis MJ, Michael WM, Dreyfuss G. 1993a. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* 21:1193–1198.
- Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G. 1993b. The protein product of the fragile X gene, FMR-1, has characteristics of an RNA binding protein. *Cell* 74(2):291–298.
- Valcárcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF⁶⁵ RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. *Science* 273:1706–1709.
- Vernet C, Artzt K. 1997. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet* 13:479–484.
- Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355:609–614.
- Zhuang Y, Goldstein M, Weiner AM. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc Natl Acad Sci USA* 86:2752–2756.
- Zorn AM, Krieg PA. 1997. The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in *Xenopus* embryos. *Genes & Dev* 11:2176–2190.