Nucleic Acids Research, 1998, Vol. 26, No. 11 **2625–2637**

Activities of the Sex-lethal protein in RNA binding and protein:protein interactions

Mark Samuels[+], Girish Deshpande* and Paul Schedl

Department of Molecular Biology, Princeton University, Princeton, NJ 08540, USA

Received February 16, 1998; Revised and Accepted April 20, 1998

ABSTRACT

The *Drosophila* sex determination gene *Sex-lethal* (*Sxl*) controls its own expression, and the expression of downstream target genes such as *transformer*, by regulating pre-mRNA splicing and mRNA translation. *Sxl* codes an RNA-binding protein that consists of an N-terminus of ~100 amino acids, two 90 amino acid RRM domains, R1 and R2, and an 80 amino acid C-terminus. In the studies reported here we have examined the functional properties of the different Sxl protein domains in RNA binding and in protein:protein interactions. The two RRM domains are responsible for RNA binding. Specificity in the recognition of target RNAs requires both RRM domains, and proteins which consist of the single domains or duplicated domains have anomalous RNA recognition properties. Moreover, the length of the linker between domains can affect RNA recognition properties. Our results indicate that the two RRM domains mediate Sxl:Sxl protein interactions, and that these interactions probably occur both in *cis* and *trans*. We speculate that *cis* interactions between R1 and R2 play a role in RNA recognition by the Sxl protein, while *trans* interactions stabilize complex formation on target RNAs that contain two or more closely spaced binding sites. Finally, we show that the interaction of Sxl with the snRNP protein Snf is mediated by the R1 RRM domain.

INTRODUCTION

In *Drosophila melanogaster*, the initial choice of somatic sexual identity is made early in development in response to the primary sex-determining signal, the X chromosome to autosome ratio (1–6). The female developmental pathway is selected when the X/A ratio is one (2X/2A), while the male pathway is selected when the ratio is one half (1–3,7). The binary switch gene *Sex-lethal* (*Sxl*) is the target for the X/A signaling system; it is turned on in chromosomally female cells, while it remains off in chromosomally male cells (4,6,8). Once the pathway has been selected, *Sxl* plays a central role in subsequent sexual development by directing both determination and differentiation (1,5,9–12). *Sxl* controls these facets of development by post-transcriptional regulatory mechanisms that operate at the level of alternative splicing and mRNA translation (10,13–17). The determined state in females is maintained by a *Sxl* autoregulatory feedback loop (10,11,18,19). In this feedback loop, Sxl proteins promote their own expression by directing the splicing machinery to exclude the male-specific third exon, L3, from the mature *Sxl* mRNAs, joining exons L2 and L4. Sxl proteins activate the *transformer* (*tra*)→*doublesex* female differentiation pathway by directing the splicing machinery to skip the default 3′ splice site of the *tra* pre-mRNA, joining the first exon to the female 3′ splice site (12,16,20,21). In the absence of Sxl proteins, the male determined state is maintained by the default splicing of *Sxl* pre-mRNAs, while the default splicing of *tra* pre-mRNA results in male differentiation. The *Sxl* gene also negatively regulates the X-chromosome dosage compensation system by preventing the expression of the *msl-2* gene. However, the regulation of *msl-2* is not at the level of splicing; instead it appears to be at the level of translation (13,15,22–24).

Some progress has been made in elucidating how *Sxl* directs the splicing of *tra* and *Sxl* pre-mRNAs. Studies on *tra* have suggested that a blockage mechanism is used to promote the female-specific splice (14,16,17,25). In the currently favored model, Sxl protein binds target sites in the polypyrimidine tract at the default 3′ splice site. This prevents the general splicing factor U2AF from binding to the default polypyrimidine tract, and instead U2AF assembles a splicing complex at the weaker, downstream female-specific acceptor site (25). The mechanism employed in *Sxl* autoregulation is clearly distinct from that proposed for *tra* (26,27). First, there are multiple *cis*-acting elements for Sxl protein both upstream and downstream of the male exon L3. Of these elements, those located downstream seem to be most critical for male exon skipping. Second, the *cis*-acting elements are located at a considerable distance from the male exon 3′ and 5′ splice sites. Hence, if Sxl protein prevents general splicing factors from binding to the male exon splice sites, it must do so indirectly, and not by competing for overlapping binding sites (26–28).

Sxl transcripts are subject to a number of alternative splices that are not sex specific but may be tissue or stage specific (29,30). While these alternatively spliced mRNAs produce Sxl proteins that have slightly different amino acid sequences, all of the proteins contain two copies of the 90 amino acid RNA recognition motif (RRM, also known as the RNA-binding domain or RBD), RRM1 (R1) and RRM2 (R2) (11,31). As expected from this homology,

*To whom correspondence should be addressed. Tel: +1 609 258 5003; Fax: +1 609 258 1028; Email: gdeshpande@watson.princeton.edu

[+]Present address: Myriad Genetics, Salt Lake City Utah, UT, USA

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

bacterially expressed Sxl protein binds to RNA, and can regulate splicing *in vitro* (14,25,31–34). The sequence specificity of the Sxl protein has been determined. Putative target sites were initially identified by molecular genetic analysis of the sequences required for the alternative splicing of *tra* (16) and *Sxl* (26,27). A sequence of eight contiguous uridine residues found in both *tra* and *Sxl* was suggested as the likely target for Sxl protein. Direct binding of Sxl protein to RNAs containing this sequence was subsequently observed, and the strength of binding was found to decrease when the number of contiguous U residues was reduced. Conversely, synthetic poly-uridylate is a potent inhibitor of binding to the specific target sites. *In vitro* selection (SELEX) experiments with RNA substrates have confirmed *Sxl*'s preference for poly(U) runs (33,35). In both the *Sxl* and *tra* pre-mRNAs the Sxl protein binding sites are often arranged as pairs of closely spaced poly U runs. For example, the default *tra* polypyrimidine tract has a canonical U_8 run, plus a second U_5GUUG tract just upstream. The Sxl protein binds with high affinity to the U_8 run, and with lower affinity to the U_5GUUG sequence (31). Moreover, the binding of a Sxl protein to this lower affinity site appears to be stabilized by contacts with protein bound to the high affinity site. It is thought that similar protein:protein interactions stabilize Sxl protein binding to the paired sites in the *Sxl* pre-mRNAs (34).

Genetic studies have implicated three other genes in *Sxl* splicing regulation: *sans-fille* (*snf*), (9,36,37), *fl(2)d* (38) and *virilizer* (39). While little is known about the molecular nature of either *fl(2)d* or *virilizer*, the *snf* gene has been cloned, and it encodes a protein homologous to the vertebrate proteins U1A and U2B″ (36,40). These two proteins are components, respectively, of the vertebrate U1 and U2 snRNPs. Mutations in *snf* can have female-lethal interactions with *Sxl* mutations, and this genetic interaction appears to be due to a failure in properly establishing or maintaining the female specific splicing pattern of *Sxl* pre-mRNAs (9,41). A plausible hypothesis to explain this interaction is that *Sxl* may regulate the splicing of the male exon L3 by modulating the activity of the U1 and/or U2 snRNP through contacts with the Snf protein. Support for this hypothesis has come from the finding that Sxl and Snf proteins physically interact both *in vitro* and *in vivo* (28).

The aim of the present work is to understand why the Sxl protein has two similar but not identical RRM domains. We have focused our attention on two of the known activities of the Sxl protein, the recognition of target RNAs and protein:protein interactions. Besides examining the RNA binding activity of the individual RRM domains, R1 and R2, we have asked whether it is possible to reconstitute the RNA recognition properties of the 'wild type' Sxl protein by duplicating either R1 or R2. We have found that the affinity and specificity of proteins containing duplicated RRM domains is not equivalent to that of the wild type protein. Analysis of protein:protein interactions shows that R1 and R2 can interact with each other and with themselves. Taken together with the unusual RNA binding activity of an Sxl protein in which the order of the R1 and R2 domains is reversed, our results suggest that *cis* interactions between R1 and R2 may be important in the high affinity binding and sequence selectivity of the wild type Sxl protein. Additionally it appears that *trans* interactions between two Sxl proteins bound to adjacent sites on target mRNAs may be mediated by contacts between the Sxl RRM domains, rather than the N-terminal domain. Finally, we show that interactions between Sxl and Snf are mediated primarily by the first Sxl RRM domain, R1.

MATERIALS AND METHODS

Construction of recombinant Sxl subclones

The individual RRMs of Sxl (R1, R2), or both RRMs together (R12), were amplified from cDNA clone MS3 (31) using specific oligonucleotide primers that included novel *Bam*HI sites at each end for subcloning into pBluescript. The primers were designed such that any combination of these RRM fragments generated using *Bam*HI and in the correct orientation would automatically allow retention of the Sxl open reading frame. This property of the subclones simplified the construction of the duplicate RRM fusions. The PCR clones were all verified by sequencing.

The R1, R2 and R12 fragments were excised from pBluescript using *Xba*I and *Sal*I, and cloned into pGEX-KG between these sites to generate GST fusion clones G-R1, G-R2 and G-R12 in the correct reading frame. For the G-R1R1, G-R2R2, G-R1-R2 and G-R2-R1 clones, ligations were performed with various mixtures of the individual RRM fragments generated by *Bam*HI cleavage of the pBluescript clones, followed by recloning into pBluescript. Among the many possible ligation products, the appropriate clones were identified by diagnostic restriction digestions, but orientations of the inserts and junction sequences crucial to maintain reading frames were all verified by sequencing. The combined RRM fragments were recloned in pGEX-KG between *Xba*I and *Sal*I as previously (31). G-R1ΔR1 and G-R2ΔR2 were generated by single-stranded oligonucleotide-directed mutagenesis in pBluescript using deletion oligos designed to remove specific sequences in the linker regions between the duplicated RRMs.

Sequences of Sxl subclones

The general structure of the GST-Sxl fusion proteins is as follows [all amino acids except those labeled (SXL RRMs) derive from the pGEX-KG vector; (LVPRGS) constitutes the thrombin cleavage site, with cleavage between the R and G residues]:

(GST functional gene)-(LVPRGS)-PGISGGGGGILELVDP-(SXL RRMs)-DPPGCRNSISSLSIPSTRAQA.

The precise Sxl-encoded sequences are as follows: β-sheet residues are in large type, those for R2 are as presented by Lee *et al.* (42); those for R1 from a survey of alignments in the literature (there is some disagreement as to the predicted location of β-sheet 4 in RRM1). In these sequences, vertical bars have been inserted at the junctions of RRMs or duplications. The DP or D residues between parallel vertical bars are not encoded by Sxl but are residual from the *Bam*HI sites derived from the PCR primers.

G-R1. SLGSGGSDDLMNDPRASNT**NLIVNY**LPQDMTDRE-LYALFRAIGP**INTCRIM**RDYKTGYSFG**YAFVDF**TSEMDSQ-RAIKVLNGITVRNKRLKV**SYAR**PGGE.

G-R2. LKVSYARPGGESIKDT**NLYVT**NLPRTITDDQLDTIFG-KYGS**IVQKNIL**RDKLTGRPRG**VAFVRY**NKREEAQEAISA-LNNVIPEGGSQPL**SVRL**AEE.

G-R12. SLGSGGSDDLMNDPRASNT**NLIVNY**LPQDMTD-RELYALFRAIGP**INTCRIM**RDYKTGYSFG**YAFVDF**TSE-MDSQRAIKVLNGITVRNKRLKV**SYAR**PGGE|SIKDT**NLY-VT**NLPRTITDDQLDTIFGKYGS**IVQKNIL**RDKLTGRPRG-**VAFVRY**NKREEAQEAISALNNVIPEGGSQPL**SVRL**AEE.

G-R1R1. SLGSGGSDDLMNDPRASNT**NLIVNY**LPQDMT-DRELYALFRAIGP**INTCRIM**RDYKTGYSFG**YAFVDF**TSE-MDSQRAIKVLNGITVRNKRLKV**SYAR**PGGE|DP|SLGSG-

GSDDLMNDPRASNT**NLIVNY**LPQDMTDRELYALFRAIG-P**INTCRIM**RDYKTGYSFG**YAFVDF**TSEMDSQRAIKVLN-GITVRNKRLKV**SYAR**PGGE.

G-R1ΔR1. SLGSGGSDDLMNDPRASNT**NLIVNY**LPQDMTD-RELYALF-RAIGP**INTCRIM**RDYKTGYSFG**YAFVDF**TSEM-DSQRAIK-VLNGITVRNKRLKV**SYAR**PGGE|D|RASNT**NLI-VNY**LPQDMTDRELYALFRAIGP**INTCRIM**RDYKTGYSFG-**YAFVDF**TSEMDSQRAIKVLNGITVRNKRLKV**SYAR**PGGE.

G-R2R2. LKVSYARPGGESIKDT**NLYVT**NLPRTITDDQLDTI-FGKYG-S**IVQKNIL**RDKLTGRPRG**VAFVRY**NKREEAQEAI-SALNNVIPEGGSQPL**SVRL**AEE|DP|LKVSYARPGGESIKDT-**NLYVT**NLPRTITDDQLDTIFGKYGS**IVQKNIL**RDKLTGRP-RG**VAFVRY**NKREEAQEAISALNNVIPEGGSQPL**SVRL**AEE.

G-R2ΔR2. LKVSYARPGGESIKDT**NLYVT**NLPRTITDDQLD-TIFGKYG-S**IVQKNIL**RDKLTGRPRG**VAFVRY**NKREEAQE-AISALNNVIPEGGSQPL**SVRL**AEE|D|GGESIKDT**NLYVT**NL-PRTITDDQLDTIFGKYGS**IVQKNIL**RDKLTGRPRG**VAFVR-Y**NKREEAQEAISALNNVIPEGGSQPL**SVRL**AEE.

G-R1-R2. SLGSGGSDDLMNDPRASNT**NLIVNY**LPQDMTD-RELYALFRAIGP**INTCRIM**RDYKTGYSFG**YAFVDF**TSEMD-SQRAIKVLNGITVRNKRLKV**SYAR**PGGE|DP|LKVSYARPG-GESIKDT-**NLYVT**NLPRTITDDQLDTIFGKYGS**IVQKNIL**R-DKLTGRP-RG**VAFVRY**NKREEAQEAISALNNVIPEGGSQP-L**SVRL**AEE.

G-R2-R1. LKVSYARPGGESIKDT**NLYVT**NLPRTITDDQLD-TIFGKYGS**IVQKNIL**RDKLTGRPRG**VAFVRY**NKREEAQE-AISALNNVIPEGGSQPL**SVRL**AEE|DP|SLGSGGSDDLMND-PRASNT**NLIVNY**LPQDMTDRELYALFRAIGP**INTCRIM**RD-YKTGYSFG**YAFVDF**TSEMDSQRAIKVLNGITVRNKRLKV-**SYAR**PGGE.

The GST-Sxl-NT construct was generated by resecting a full-length pGEX-Sxl plasmid at the *Bal*I site in Sxl and religating at the *Hin*dIII site of pGEX-KG, thereby removing the C-terminal 2/3 of Sxl coding sequence. The resulting fusion protein contains approximately the first 90 amino acids of Sxl based on the MS3 sequence, and stops 15 amino acids before the beginning of the Sxl RRM1-containing fragment (33 amino acids before the actual beginning of RRM1 based on the consensus sequence).

Construction of recombinant Snf subclones

Fragments of the Snf coding region were amplified from cDNA D25 (40) using appropriate oligonucleotide primers including restriction sites for cloning the PCR products directly into pGEX-KG. GST-Snf-NC contains the entire Snf protein coding region. The Snf coding region was split between Snf-N and Snf-C at the position of the single intron in the Snf gene. GST-Snf-N contains the first 118 amino acids of Snf, including the entire RRM1. GST-Snf-C contains the last 98 amino acids of Snf, including the entire RRM2.

Expression and purification of recombinant Sxl and Snf proteins

MBP-Sxl protein, unfused Sxl protein and unfused Snf protein were purified as previously described (28,31,40). pGEX derivatives were grown in DH5α bacteria. For protein overexpression, fresh overnight cultures were diluted 1/100 into 250–500 ml of LB+amp medium, grown for several hours, induced with IPTG for 3 h, then harvested, washed and frozen as cell pellets at –70°C. Soluble protein extracts were prepared as previously described (31). GST fusion proteins were purified by binding of soluble cell extracts to glutathione–Sepharose (Pharmacia) equilibrated in 50 mM HEPES–NaOH, pH 7.9, 0.1 mM EDTA, 0.5 mM DTT (HED) plus 0.1 M NaCl. Binding was for 45 min to 1 h at room temperature. Beads were washed three times batch-wise with 10 bed vol of HED + 0.6 M NaCl + 0.1% NP-40, followed by three washes with HED + 0.1 M NaCl + 0.1% NP-40, followed by three washes with HED + 0.1 M NaCl. Fusion proteins were eluted from the resin batch-wise with 0.1 M Tris–HCl, pH 7.9, 0.1 mM EDTA, 0.5 mM DTT, 0.01 M NaCl, 10 mM glutathione (Sigma). Experimentation indicated that after purification NP-40 was generally not necessary to maintain protein solubility or stability.

Protein concentrations were determined by soluble Coomassie blue binding (BioRad) and absorbance at 595 nM with a bovine γ-globulin standard. Although the dye binding assay performs variably with different purified proteins, we considered this method sufficient given that the various GST-Sxl fusion proteins were so similar in design. Proteins were stored frozen at –70°C with little apparent loss of activity over time.

Thrombin cleavage of GST fusion proteins

Typically, 20 mg aliquots of fusions were treated with thrombin (Sigma) at a ratio of 10 U of protease per mg of protein, from several hours to overnight at room temperature. No attempt was made to inactivate the protease; however, cleavage products were stored frozen at –70°C to minimize over-digestion.

RNA binding assays

Radioactive RNA probes (200 nt long) were synthesized from linearized PIP4 derivative plasmids as described in detail in Samuels *et al*. (31). The PIP4 insert is a wholly synthetic sequence developed to assay general splicing activity. The derivatives employed in the current work differ only at their 3′ acceptor sites; the specific sequences are as follows:

Sxl-WT	CCCAUUUUUUUUCACAG
Sxl-Mut	CCCAUUCUCUCUCACAG
PIP4	CCCUUUUUUUUCCACAG
PIP4-U_7	CCCCUUUUUUUCCACAG
PIP4-U_7(5′A)	CCCAUUUUUUUCCACAG
Tra-WT	CCCUCUUUUUGUUGUUUUUUUUCUAG
Tra-Mut	CCCUCUUUUUGUUGUUCUCUCUCUAG

The Sxl-WT sequence corresponds to the *D.melanogaster Sxl* male exon acceptor; the Tra-WT sequence corresponds to the *D.melanogaster tra* non-specific acceptor site.

Three oligonucleotides were also used for the RNA binding assays. The sequence of the oligos are 1: CAUAUUUUUUUUC-AC, 2: CAUAUCUCUUUUCAC, and 3: UAUUUUUUUUG-AUUUUUUUUCA. Number 1, or 'mono', contains a single canonical Sxl binding site. This canonical binding site is mutated by the introduction of two C residues in oligo number 2. Finally, oligo number 3, or 'di', has two contiguous canonical Sxl binding sites. The oligos were 5′ end labeled with ^{32}P using radioactive ATP and T4 polynucleotide kinase.

Gel mobility shift and UV cross-linking assays were performed essentially as previously described (31).

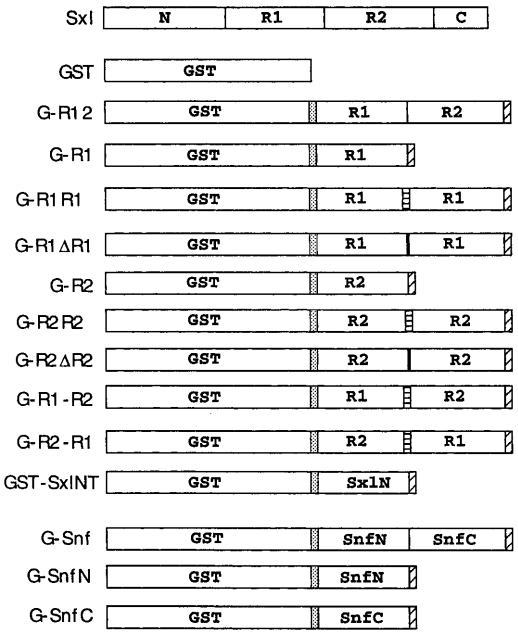


Figure 1. Schematic diagram of fusion proteins. Structures of the various GST-Sxl and GST-Snf fusion proteins are shown. The unfused Sxl and GST shown at the top are for comparison. Shaded and diagonally or horizontally hatched boxes represent novel linker sequences between domains or at the C-terminus. Thick vertical lines between duplicated Sxl R1 or R2 domains in G-R1ΔR1 and G-R2ΔR2 represent shortened novel linker sequences. Only G-R12 and G-Snf retain unmodified wild type linker sequences between the RRM domains. The Sxl N-terminal domain has amino acids 1–92; the R1 protein has amino acids 107–206 and the R2 protein has amino acids 203–291. For all specific sequences see Materials and Methods.

Protein blotting assays

Western blots were performed with anti-Sxl antibodies as previously described (2,28,31). Far western blots with biotinylated Sxl or MBP-Sxl were performed as previously described (28).

Co-immunoprecipitation assays

In vitro co-immunoprecipitations were performed essentially as described previously (43,44). Chemical cross-linking of protein-A Sepharose beads was performed as described in Harlow and Lane (44).

RESULTS

RNA binding of the Sxl RRM domains

We have previously shown that a full-length Sxl protein, either alone or when fused to the maltose-binding protein (MBP-Sxl), is capable of high affinity sequence-specific RNA binding (31). In order to learn more about the contributions of the different Sxl protein domains to RNA binding, we expressed and purified a series of GST-Sxl fusion proteins (Fig. 1). We focused our attention primarily on the two RRMs of Sxl, R1 and R2.

The first fusion protein examined was G-R12 which contains both Sxl RRM domains but lacks the N- and C-terminal domains (Fig. 1). Gel shift and UV cross-linking assays indicate that the RNA recognition properties of G-R12 are quite similar to those of the unfused Sxl protein or the MBP-Sxl protein (31). The R12 protein with (Fig. 2A), and without the GST moiety (not shown), binds with high affinity to Sxl-Wt RNA containing a canonical U_8 Sxl recognition sequence, but poorly to the Sxl-Mut RNA. When assayed with a spectrum of binding site variants, the G-R12 protein showed steadily decreasing binding in the series Sxl-Wt > PIP4-U_7(5′A) > PIP4 > PIP4-U_7 >> Sxl-Mut (see Materials and Methods for sequence), exactly the same order as for MBP-Sxl (data not shown; see 31). Finally, like MBP-Sxl, the G-R12 protein forms two successive complexes on an RNA containing the Tra-Wt default acceptor site; the first on the canonical U_8 sequence, and the second on the weaker upstream binding site (not shown; see 31). However, the yield of the slower migrating (presumably) dimeric complex was reduced compared to MBP-Sxl, suggesting that G-R12 may have lost some protein:protein interaction capabilities. These results, as well as studies by others (32,45), indicate that the R12 protein retains most of the RNA recognition properties of intact Sxl.

We next examined the fusion proteins G-R1 and G-R2 which contain only a single RRM domain (Fig. 1). While we did not detect RNA:protein complexes with G-R1 or G-R2 in gel shift assays, UV cross-linking experiments indicate that each domain alone is capable of sequence specific interactions. In the experiment shown in Figure 3, the Sxl Wt oligo #1 was used as a probe for UV cross-linking. Both G-R1 and G-R2 could be cross-linked to oligo #1; however, based on the relative signal in this and other experiments, the binding activity of the individual domains is greatly reduced compared to G-R12 (or MBP-Sxl). We estimate the binding activity of the R1 to be ~1/60–1/80 that of G-R12, while R2 is ~1/100. Comparable results were obtained with oligo #3 (which contains two canonical Sxl binding sites) and with longer RNA probes.

Like G-R12, the single domains showed little or no cross-linking to the Sxl Mut oligo (#2) suggesting that they are capable of discriminating between different polypyrimidine sequences (data not shown). On the other hand, the individual domains seem to be less specific than G-R12. This is suggested by experiments in which the binding of the individual RRMs to oligo #1 (Sxl Wt), is competed with cold oligos #1 or #2 (Sxl Mut) (Fig. 3A) or various homo and mixed sequences polymers (not shown). Binding of G-R1 and G-R2 to the Wt oligo could be competed not only with cold Wt oligo but also with cold Mut oligo. In contrast, for G-R12, cold Sxl Wt oligo is a much more efficient competitor than the Sxl Mut oligo.

Binding activity of duplicated Sxl RRMs

The results described in the previous section indicate that the single RRM domains have RNA recognition properties that are quantitatively and qualitatively different from those of G-R12. An obvious question is whether it is possible to generate a protein with the affinity and specificity of R12 by simply duplicating either the R1 or R2 domain. To answer this question, we generated the fusion proteins G-R1R1 and G-R2R2 (Fig. 1). An intrinsic problem in designing these duplicates is setting the length (and nature) of the amino acid sequence that serves as the linker between the tandem RRM domains. This is a potentially important issue for two reasons. First, theoretical calculations predict that as the linker length decreases (assuming a flexible non-interactive linker), the binding affinity of a multi-RRM domain protein should increase, approaching the product of the

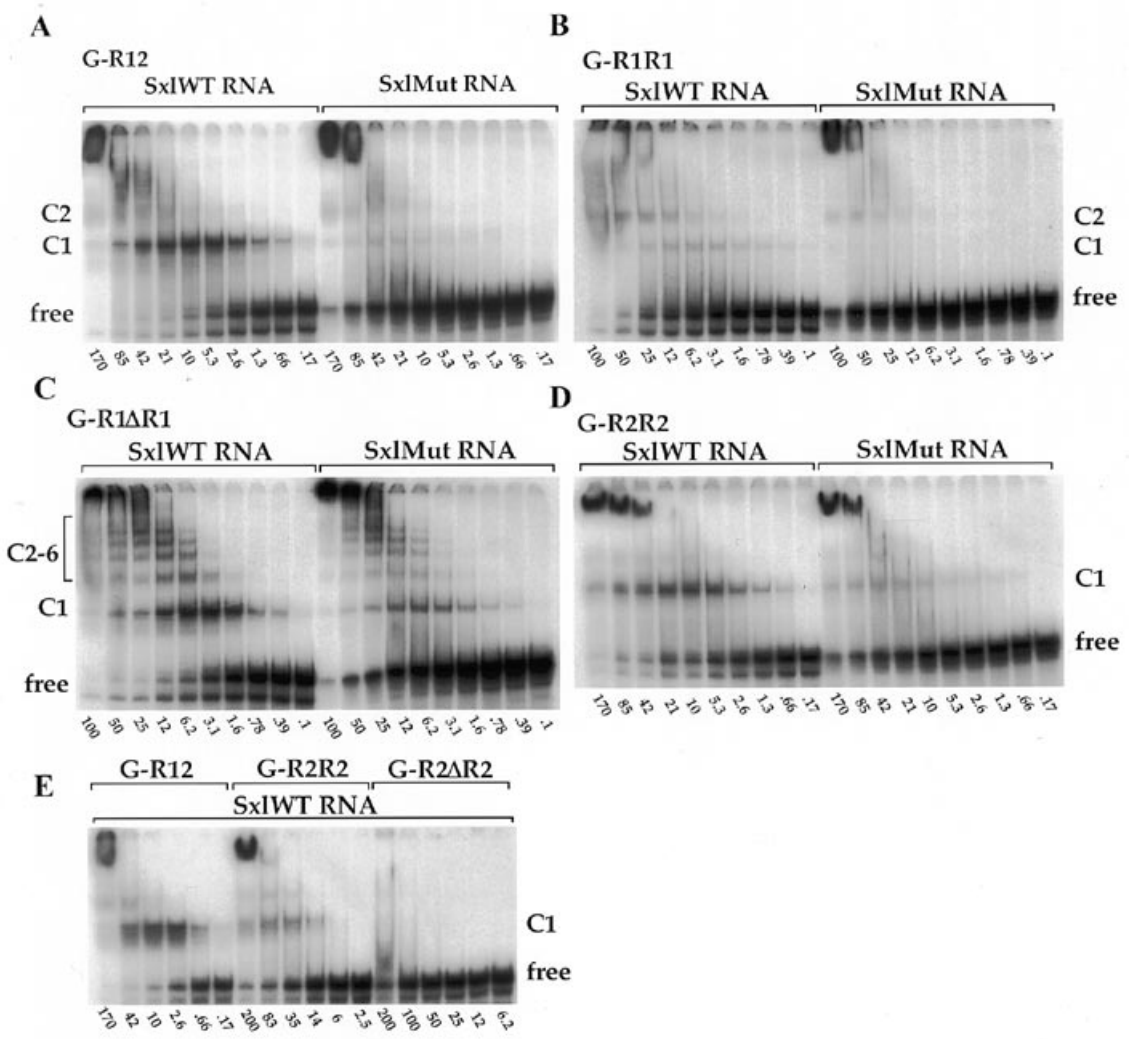


Figure 2. Gel mobility shift assays with duplicated R1 and R2 fusions. Varying concentrations of fusion proteins were incubated with Sxl RNAs: Sxl-WT and Sxl-Mut, and electrophoresed on native acrylamide gels as described in Materials and Methods. The unbound RNA species is indicated as 'free' on the right side of each panel. Bound species are indicated either as 'bound' or as complex 'C1' or 'C2' as appropriate. All protein concentrations are given in nM final concentrations in the binding reactions. (**A**) G-R12 at 170, 85, 42.5, 21.2, 10.6, 5.3, 2.6, 1.3, 0.66, 0.17 nM with SxlWT RNA; G-R12 at the same concentrations with SxlMut RNA. (**B**) G-R1R1 at 100, 50, 25, 12.5, 6.2, 3.1, 1.6, 0.78, 0.39, 0.1 nM with SxlWT RNA; G-R1R1 at the same concentrations with SxlMut RNA. (**C**) G-R1ΔR1 at 100, 50, 25, 12.5, 6.2, 3.1, 1.6, 0.78, 0.39, 0.1 nM with SxlWT RNA; G-R1ΔR1 at the same concentrations with SxlMut RNA. (**D**) G-R2R2 at 170, 85, 42.5, 21.2, 10.6, 5.3, 2.6, 1.3, 0.66, 0.17 nM with SxlWT RNA; G-R2R2 at the same concentrations with SxlMut RNA; (**E**) G-R12 at 170, 42.5, 10.6, 2.6, 0.66, 0.17 nM with SxlWT RNA; G-R2R2 at 200, 83, 35, 14.5, 6.0, 2.5 nM with SxlWT RNA; G-R2ΔR2 at 200, 100, 50, 25, 12.5, 6.2 nM with SxlWT RNA.

binding affinities of the individual RRMs (46). Conversely, the binding affinity should decrease as the linker length increases, ultimately equalling only twice the affinity of the individual domains. Second, *cis* (and *trans*) protein:protein interactions between the homologous (or, in the case of R12, heterologous) RRM domains could play a critical role in the RNA recognition capabilities of the duplicate fusion proteins (see 47 and below). Such interactions may be dependent upon the length and/or nature of the linker. The G-R1R1 linker is 25 amino acids, while the G-R2R2 is 21 amino acids (Materials and Methods). Both of these linkers are significantly longer than the native R12 Sxl linker sequence which is only 9 amino acids. For these reasons we generated a second set of duplicates, G-R1ΔR1 and G-R2ΔR2, whose linker sequence (10 and 12 amino acids respectively) is closer to that of native Sxl. The RNA binding properties of these duplicates are summarized here.

G-R1R1. Unlike G-R1, the duplicated RRM1 fusion, G-R1R1, formed complexes with wild type Sxl RNA that could be detected using the gel shift assay. However, the binding activity of the G-R1R1 duplicate in both gel shift and UV cross-linking assays is reduced compared to the G-R12 fusion protein. In the gel shift assay using wild type Sxl RNA, higher concentrations of input protein were required to detect levels of shifted complex equivalent to that formed with G-R12 (compare Fig. 2B with A). In addition, the G-R1R1 protein is unusual in that it forms both a primary complex and a second more slowly migrating complex (presumably a dimer) on Sxl-Wt RNA. In contrast, G-R12 does

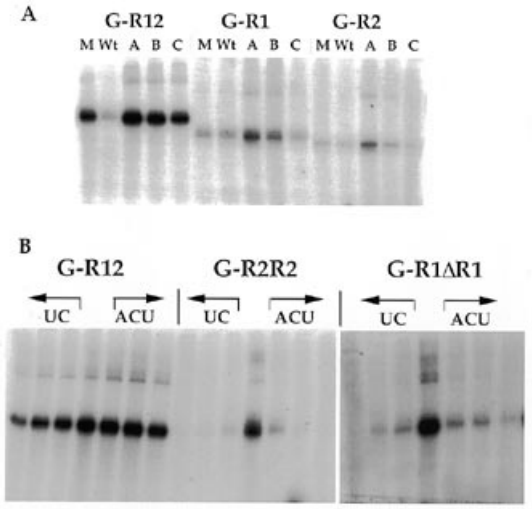


Figure 3. RNA binding and competition assays for fusion proteins containing the single RRM domains, R1 and R2. In the experiment shown in this figure (**A**), the Sxl fusion proteins G-R12, G-R1 and G-R2 were incubated with oligo #1 ('Mono'). This oligo is 15 nt in length and contains a single canonical Sxl protein binding site. RNA binding was detected by UV cross-linking the oligo to the protein. The samples were then analyzed by SDS–PAGE. For G-R12, the initial input of protein in lane A is 30 nM, while it is 600 nM for G-R1 and G-R2. In lanes B and C, the amount of input fusion protein is successively reduced 2-fold. Cold competitor RNA, oligo #1 ('Wt') and oligo #2 ('M') were added as indicated. Oligo #2 is 15 nt in length and contains a mutated Sxl protein binding site in which two of the U residues were replaced by C residues (Materials and Methods). Protein concentrations were the same as in lane A. Final concentration of each cold competitor was ~10 µM. (**B**) Duplicate fusion proteins. UV-cross-linking and competition assays with the random co-polymers, poly UC and poly ACU. The fusion proteins G-R12, G-R2R2 and G-R1ΔR1 (at 0.03, 0.07 and 0.07 µM) were incubated with radioactively labeled oligo #1 either without (central lane for each fusion protein) or with increasing concentrations of the random co-polymers, poly UC (left arrow) or poly ACU (right arrow). Concentrations of the random co-polymers were 0.004, 0.01 and 0.02 mg/ml. The binding of the fusion proteins to labeled oligo #1 was assayed by UV cross-linking. Oligo #1 is 15 nt in length and contains a single canonical Sxl protein binding site.

not usually form dimers with Sxl Wt RNA (Fig. 2A) though dimers are observed with substrates that have two closely spaced Sxl binding sites.

Differences between G-R1R1 and G-R12 are also evident using oligonucleotides containing either one or two Sxl Wt binding sites. While G-R12 gel shifts the 15 nt oligo #1 (one binding site), G-R1R1 complexes with this oligo do not survive gel electrophoresis. G-R1R1 does shift the 22 nt oligo #3 (two binding sites); however, the binding activity of the duplicate fusion protein is substantially less (>50-fold) than that of G-R12 (not shown). Though the difference in binding activity of the two proteins is not as pronounced in UV cross-linking assays (only~15–20-fold for the oligo #1 and ~8–10-fold for the oligo #3; not shown), G-R1R1 still binds less well to the shorter single binding site substrate. This apparent preference for larger substrates containing two binding sites could mean that G-R1R1:RNA complexes are stabilized by *trans* protein:protein interactions. This possibility is supported by gel shifts using the Sxl-Mut RNA. Although G-R12 binding is significantly reduced by mutations in the poly U run, this protein does form primary complexes with Sxl-Mut RNA at high concentrations (Fig. 2A). However, this is not the case for G-R1R1; even at very high concentrations no primary complexes are detected (Fig. 2B). On the other hand, we do detect significant amounts of a more slowly migrating secondary complex, presumably a G-R1R1 dimer. Moreover, the yield of this secondary complex is nearly equivalent to that observed with the Sxl-Wt RNA (Fig. 2B). Similar results were obtained for the Tra-Wt and Tra-Mut RNAs (not shown).

These findings suggest that the G-R1R1 protein is capable of discriminating between canonical and non-canonical Sxl binding sequences; however, it also appears that some types of *trans* protein:protein interaction (see also G-R1ΔR1 below) can stabilize inappropriate complexes with non-canonical sequences. Additional evidence that the G-R1R1 protein is less able to discriminate between canonical and non-canonical sequences than G-R12 comes from experiments in which cross-linking to Sxl oligos is competed with poly U and UC and ACU random co-polymers. While the cross linking of both proteins to Sxl oligos is strongly competed by poly U, the two proteins differ substantially in their sensitivity to the UC and ACU co-polymers. In the case of the G-R12 protein, the UC polymer is a moderate to weak competitor, while the ACU polymer is a very poor competitor (Fig. 3B). In contrast, the binding to G-R1R1 to the Sxl oligos is readily competed by both the UC and the ACU co-polymer (data not shown).

G-R1ΔR1. The reduction in the length of the linker separating the R1 domains in the G-R1ΔR1 protein has two effects on RNA binding. First, as predicted (46), shortening the linker increases the overall binding activity of the fusion protein in both gel shift and UV cross-linking assays. Second, it markedly stimulates the formation of oligomeric protein:RNA complexes. Both of these effects are illustrated in the gel shift assay using Sxl WT and Sxl Mut RNAs shown in Figure 2C.

G-R1ΔR1 complexes with Sxl Wt RNA can be detected at lower protein concentrations than those required for the G-R1R1 fusions (compare Fig. 2B and C). In addition, unlike G-R1R1, the G-R1ΔR1 duplicate is able to gel shift an oligo containing only a single Sxl protein binding site (not shown). However, the binding activity of G-R1ΔR1 is still reduced compared to G-R12 (>10-fold). The difference in the binding activity of G-R1ΔR1 and G-R12 in the UV cross-linking assay is smaller, only about 3–6-fold for both the single and double site oligos.

The most remarkable property of the G-R1ΔR1 protein is its propensity to generate a ladder of discrete shifted bands. As illustrated in Figure 2C, at least six shifted products can be seen in the ladder formed with the Sxl Wt RNA. Multiple bands are even observed in gel shifts with the oligos containing one and two Sxl binding sites (not shown). As evident from a comparison of the G-R1ΔR1 gel shifts with the Sxl WT and Sxl Mut RNA substrates (see Fig. 2C), the production of both monomeric and multimeric complexes is only slightly sequence specific. The gel shift patterns with these two RNA substrates are remarkably similar and there is only a small decrease in shifted band intensity with the mutant RNA. Similar results were obtained using Tra-WT and Tra-Mut RNA.

The inability of the G-R1ΔR1 fusion protein to discriminate between different RNA sequences is also evident in competition experiments. In the UV cross-linking experiment shown in Figure 3B, the binding of the G-R12 and G-R1ΔR1 to the oligo containing two Sxl protein binding sites was competed with increasing concentrations of either UC or ACU random co-

polymers. The cross-linking of the G-R12 fusion is reduced by the UC polymer only at the highest concentrations, while even at high concentrations the ACU polymer has little effect. A quite different result is obtained for G-R1ΔR1; both the UC and ACU co-polymers are quite efficient competitors, and there is a substantial drop in the yield of cross-linked Sxl oligo even at the lowest polymer concentrations. Hence while the reduced linker length of the G-R1ΔR1 protein facilitates cooperative binding to RNA substrates, a consequence of this enhanced cooperativity is a substantial reduction in recognition specificity.

G-R2R2. Of the four duplicate RRM domain fusion proteins, the G-R2R2 fusion most closely resembles G-R12. The G-R2R2 fusion protein has relatively strong RNA binding activity in gel shift assays with the Sxl-Wt RNA probe (Fig. 2D), with PIP4 and Tra RNA and with the oligos (not shown). However, it is clear from side-by-side comparison that the binding of G-R2R2 to these different substrates is still weaker than that of the intact G-R12 fusion (>15-fold). The differences in binding activity estimated from the UV cross-linking assay are slightly smaller: ~5-fold for the two oligo substrates.

The G-R2R2 protein also exhibits sequence specificity in the gel shift assay, and, as shown in Figure 2D, binding to the Sxl-Mut RNA probe is clearly weaker than to the Sxl-Wt probe. However, the specificity of this protein for the Sxl Wt probe is not equivalent to G-R12 (compare Fig. 2D with A). Competition experiments also indicate that G-R2R2 protein discriminates rather poorly between canonical Sxl protein binding sites and other RNA sequences. Like both of the RRM1 duplicates, the binding of G-R2R2 to Sxl oligos is readily competed by both of the non-specific random polymers, UC and ACU (Fig. 3B).

G-R2ΔR2. While shortening the linker length in the R1 duplicate had the expected effect of increasing binding affinity (46), this was not the case for the R2 duplicate, G-R2ΔR2. As illustrated in Figure 2E, we found that G-R2ΔR2 showed substantially weaker RNA binding activity than G-R2R2. Similar results were obtained in UV cross-linking assays. With the oligo containing a single Sxl site the RNA binding activity of the G-R2ΔR2 protein in the cross-linking assay is as much as 25–50-fold less than that of G-R12 (not shown). This level of activity is only about twice that of the single R2 fusion protein, and suggests that the reduction in the length of the linker sequence between the two RRM domains may interfere with the ability of the two R2 domains to simultaneously bind RNA.

Linker effects on combined RRM1 and RRM2

Since changes in the length of the linker region separating the RRM domains altered the RNA binding properties of the duplicate fusion proteins, we decided to examine the effect of spacing on the G-R12 protein. For this purpose we generated a fusion protein, G-R1-R2, in which the length of the linker sequence between the two domains is increased compared to that in G-R12 (see Materials and Methods for construction, and details of the structure). Except for the extra amino acids between the two domains, G-R1-R2 is identical to the initial G-R12 fusion.

In spite of the increased spacing between R1 and R2, G-R1-R2 bound RNA strongly (Fig. 4A). In side-by-side comparisons using the Sxl Wt probe, the binding activity of G-R1-R2 in the gel shift assay using the Sxl-WT RNA probe was only slightly weaker than that of G-R12 (Fig. 4A). There was also little difference in RNA binding activity in the UV cross-linking assay with oligos containing either one or two Sxl binding sites (Fig. 4D). Moreover, like G-R12, G-R1-R2 retained good sequence specificity, and was able to discriminate between Sxl-Wt and Sxl-Mut RNAs (Fig. 5B). Thus, altering the linker sequences between R1 and R2 in the RRM1-RRM2 protein had a much less profound effect on its RNA recognition properties than was observed for the duplicated RRM proteins.

Reversing the order of the RRM domains gives a protein with unusual RNA binding properties

We also generated a novel fusion protein, G-R2-R1, in which we reversed the order of the two RRMs. In gel shift assays using the Sxl Wt probe, the G-R2-R1 fusion protein showed only very weak binding to Sxl-Wt RNA, much less than either G-R12 or G-R1-R2. As can be seen in Figure 4C, virtually no primary complex is formed even at quite high protein input. On the other hand, we do observe a fairly prominent band corresponding to the secondary complex. This unusual behavior is also evident in UV cross-linking assays. As shown in Figure 4D, the G-R2-R1 fusion protein shows substantially less cross-linking to oligo #1 which contains a single binding site than either G-R12 or G-R1-R2. A quite different result is obtained when the oligo containing two Sxl binding sites (#3) is used as a probe. As can be seen in the second part of Figure 4D, the RNA binding activity of the G-R2-R1 protein is essentially equivalent to G-R12 (or G-R1-R2) when there are two Sxl binding sites. One plausible hypothesis to explain the unusual behavior of G-R2-R1 is that *cis* protein:protein interactions between R1 and R2 increase the affinity and/or specificity of Sxl protein. These *cis* R1:R2 interactions cannot occur when a single G-R2-R1 protein moiety is bound to RNA; however, when there are two G-R2-R1 proteins bound to closely spaced sites on an RNA *trans* R1:R2 protein:protein interactions can occur which reproduce the normal R1:R2 *cis* interactions.

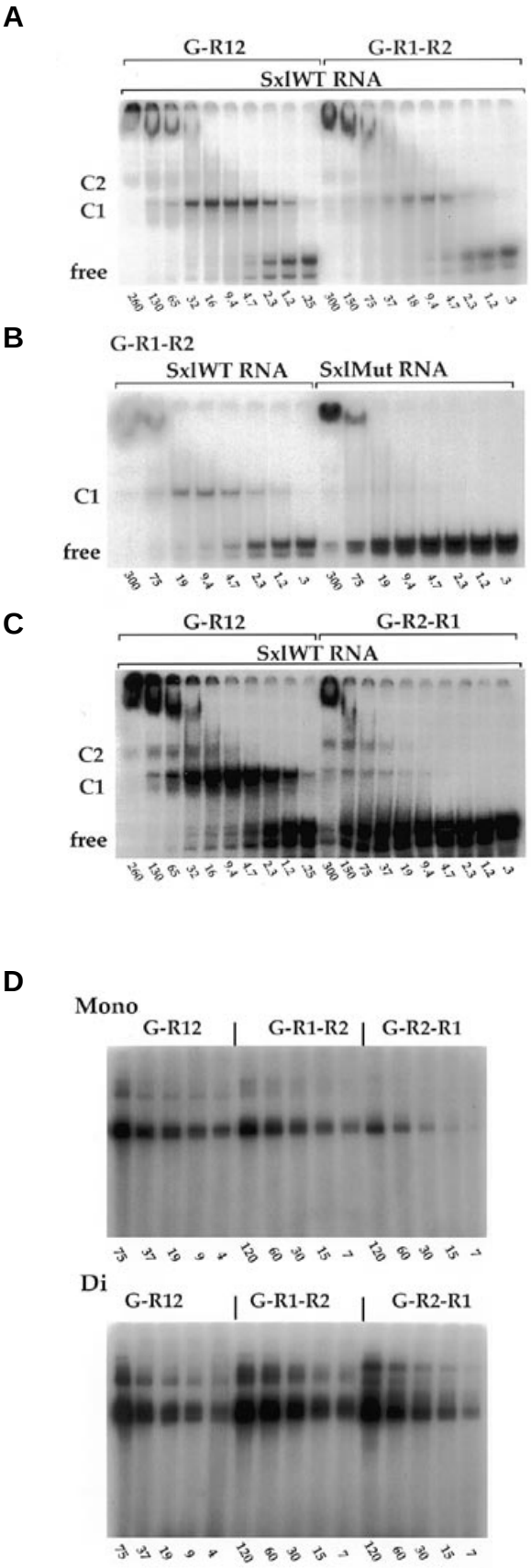
Sxl protein:protein interactions

The results described above (see also 31,34,45) indicate that the binding of Sxl protein to RNA can be stabilized by protein:protein interactions. Two types of protein:protein interactions are suggested. The first are *trans* interactions between the RRM domains of proteins bound to nearby sites. The second are *cis* interactions between two RRM domains within the same protein. Since these interactions were detected in the presence of target RNAs, it is not clear whether they are dependent upon RNA binding or can occur in the absence of exogenous RNA. To investigate this question we used biotinylated Sxl protein (either unfused or MBP-Sxl) to probe western blots containing full length Sxl and various GST fusion proteins.

We found that full length (unfused) Sxl protein labels both unfused Sxl and MBP-Sxl but not GST (Fig. 5A) or MBP (data not shown). Similar results were obtained when the MBP-Sxl fusion protein was used as a probe. These findings indicate that the Sxl:Sxl interactions can occur in the absence of added exogenous RNA.

Mapping the Sxl interaction domains

Previous studies by Wang and Bell (34) suggested that the *trans* Sxl:Sxl protein interactions observed in RNA binding experiments are mediated by the N-terminal domain. Although a comparision of



the RNA binding properties of the G-R12 protein with full length Sxl (MBP-Sxl or unfused Sxl) also suggested that the N-terminal domain may enhance cooperative *trans* interactions (data not shown), the contributions of the N-terminus appeared to be relatively minor. In contrast, our analysis of the RNA binding properties of proteins containing duplicated or reversed RRM domains indicates that RNA dependent *trans* (and *cis*) protein:protein interactions are likely to be mediated by the Sxl RRM domains themselves. To further explore the role of the different Sxl protein domains in protein:protein interactions, we probed western blots of GST fused to different Sxl domains with either the full length (unfused) Sxl protein (Fig. 5A) or the MBP-Sxl protein (not shown).

Unlike either the unfused Sxl protein or the MBP-Sxl protein, the Sxl N-terminus fusion protein, GST-NT, is not labeled by the biotinylated Sxl protein probe (Fig. 5A). Although this is a negative result, it would suggest that the N-terminus is probably not responsible for the Sxl protein:protein interactions we detected in western blots. This suggestion is supported by experiments in which we used biotinylated GST-NT to probe western blots: the Sxl N-terminus did not label any of the Sxl proteins (not shown).

A quite different result is obtained for the RRM domains. As can be seen in Figure 5A, all of the GST fusions containing the RRM domains are labeled with the Sxl probe. Additionally, the signal observed for the R2 fusions (either the single: G-R2 or the duplicate: G-R2R2) is stronger than that observed for the RRM1 fusions (single: G-R1 and duplicate: G-R1R1). Though the difference in signal between the two RRM domains is small, it is consistently observed in the far western experiments.

To provide additional evidence that the individual RRM domains are capable of protein:protein interactions we probed western blots containing different Sxl proteins with either G-R1 or G-R2. In the example shown in Figure 5B, we probed a blot containing increasing concentrations of G-R1 and G-R2 with labelled G-R2 protein. This experiment shows that G-R2 recognizes itself and G-R1 but does not label GST. Similar results are obtained when G-R1 is used as the probe. These findings indicate that the two RRM domains are capable of both homotypic and heterotypic interactions.

Protein:protein interactions detected by immunoprecipitation

The results described in the previous section map the Sxl:Sxl interaction domains to the two RRM domains, R1 and R2. To confirm this finding we asked whether the two RRM domains can

Figure 4. Gel mobility shift and UV cross-linking assays with novel combined R1 and R2 fusions. Gel mobility shift assays in (A)–(C) were performed as described in Figure 2. (**A**) G-R12 at 260, 130, 65, 32.5, 16.2, 8.1, 4.1, 2.0, 1.0, 0.25 nM with SxlWT RNA; G-R1-R2 at 300, 150, 75, 37.5, 18.8, 9.4, 4.7, 2.3, 1.2, 0.3 nM with SxlWT RNA. (**B**) G-R1R2 at 300, 75, 18.8, 9.4, 4.7, 2.3, 1.2, 0.3 nM with SxlWT RNA; G-R1-R2 at the same concentrations with SxlMut RNA. (**C**) G-R12 as in (A) with SxlWT RNA; G-R2-R1 at 300, 150, 75, 37.5, 18.8, 9.4, 4.7, 2.3, 1.2, 0.3 nM with SxlWT RNA. (**D**) shows the UV cross linking of the three Sxl fusion proteins, G-R12, G-R1-R2 and G-R2-R1, to oligo #1: 'Mono' (Materials and Methods) and oligo #3: 'Di'. Oligo #1 (Mono) is 15 nt in length and contains a single canonical Sxl protein binding site while oligo #3 (Di) is 22 nt in length and contains two contiguous Sxl protein binding sites. As indicated by the arrows for each fusion protein, the highest protein concentration is in the reaction mix loaded in the left-most lane. The input fusion protein is then diluted 2-fold in each successive lane. The amount of input G-R12 beginning with the first lane of the series is 75, 37.5, 18.7, 9.4, 4.5 nM, while for G-R1-R2 and G-R2-R1 the input protein is 120, 60, 30, 15, 7.5 nM.

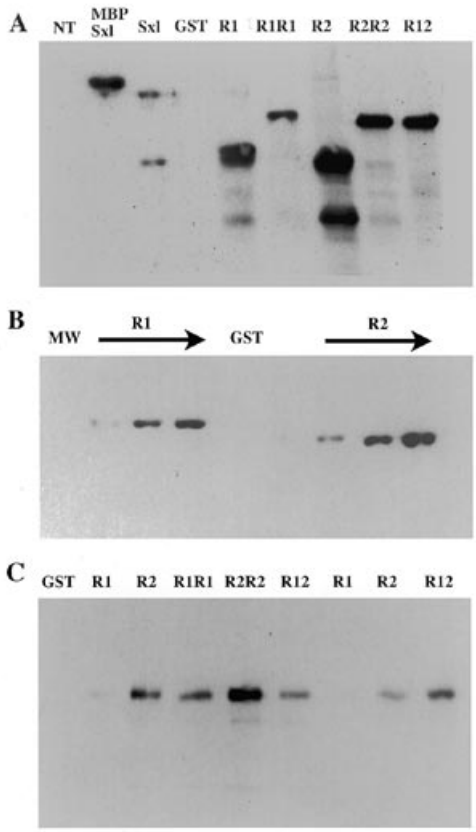


Figure 5. Sxl protein interacts with itself *in vitro*. (**A**) Detection of Sxl-Sxl interactions by far western blots. Various Sxl fusion proteins were expressed and purified as described in the Materials and Methods. These fusion proteins were run on SDS–PAGE and transferred onto nitrocellulose. Biotinylated, unfused Sxl fusion protein was used to probe western blots of purified, recombinant Sxl sub-fragments. Binding of biotinylated Sxl was visualized by chemiluminescence. Samples as loaded from left to right: GST-NT (Sxl N-terminus fused to GST), MBP-Sxl, unfused Sxl, GST alone, G-R1, G-R1R1, G-R2, G-R2R2 and G-R12. (**B**) Individual RRMs can interact with each other in far western analysis. Increasing concentrations of G-R1 and G-R2 were run on SDS–PAGE and transferred onto nitrocellulose. Biotinylated, G-R2 fusion protein was then used to probe the western blots. Binding was visualized by chemi-luminescence. Note that no binding is observed in the lane containing the GST protein control. (**C**) Detection of Sxl-Sxl interactions by immunoprecipitation. For *in vitro* imunoprecipitations, different purified GST Sxl fusion proteins were preincubated with GST antibody beads for 2 h at room temperature. The loaded beads were then incubated in the presence of equal quantitites of biotinylated Sxl for 2 h at room temperature. The beads were washed with immunoprecipitation buffer and eluted by boiling them in SDS–PAGE sample buffer. Samples were run on gels and transferred onto nitrocellulose. The binding of biotinylated Sxl was visualized by chemiluminescence. The samples in the lanes from left to right were obtained from GST beads loaded with: GST alone, G-R1, G-R2, G-R1R1, GT-R2R2, G-R12, G-R1, G-R2 and G-R12.

co-immunoprecipitate biotinylated Sxl protein. For this purpose, equal quantities of the different GST-Sxl fusion proteins where bound to Sepharose beads by covalently coupled anti-GST antibody. Each preparation of beads was then incubated with biotinylated Sxl protein. After collecting and extensively washing the beads, the bound biotinylated Sxl protein was released by boiling and analyzed by gel electrophoresis and blotting.

As can be seen in Figure 5C, Sxl protein does not bind to the control beads containing unfused GST protein. In contrast, Sxl protein does bind to beads containing the GST-Sxl fusion G-R12 (Fig. 5C). As was observed in the far western experiments, the interaction capabilities of the two RRM domains in this binding assay are not entirely equivalent. In general, the amount of Sxl protein bound by G-R1 is somewhat less than that bound by G-R2 (see Fig. 5C). Similar differences are evident for the duplicate RRM fusion proteins (compare the R1 duplicate G-R1R1with the R2 duplicate G-R2R2). We also asked whether the interactions between Sxl proteins could be stimulated by the addition of exogenous RNA; however, no effects were observed (not shown). The yield of Sxl protein in the immunoprecipitates was also unaffected by the addition of RNase (not shown).

Protein:protein interactions between Sxl and Snf

One of the genes thought to play a key role in *Sxl* splicing regulation is *snf*. The first *snf* mutation identified was the anti-morphic female sterile allele, *1621*. snf^{1621} females are sterile because their germ cells express *Sxl* in the male rather than the female mode (9,37,41). This germline defect appears to be due to a failure to properly activate and/or maintain the *Sxl* autoregulatory feedback loop, and the sterility of snf^{1621} females can be completely rescued by gain-of-function *Sxl* mutations which constitutively splice *Sxl* transcripts in the female mode (41). Defects in *Sxl* autoregulation are also evident in the soma. In contrast to wild type, the soma of snf^{1621} females has small but readily detectable amounts of male spliced *Sxl* mRNA (41). In addition, snf^{1621} shows synergistic female-lethal interactions with loss-of-function *Sxl* mutations (9,36,41). As in the germ line, these lethal interactions appear to be due to a failure in the *Sxl* autoregulatory feedback loop in early embryos.

The *snf* gene encodes a protein which consists of two closely spaced RRM domains. The Snf RRM domains show extensive homology to the RRM domains of two related vertebrate splicing factors, U1A and U2B″ (36,40). U1A is a component of the U1 snRNP, while U2B″ is a component of the U2 snRNP. Snf appears to be the fly equivalent of these two vertebrate proteins and is found in both the U1 and the U2 snRNPs (28,40). Studies on the RNA binding properties of the vertebrate U1A and U2B″ proteins indicate that the N-terminal RRM domain of these proteins is responsible for recognition of the cognate snRNAs, while the C-terminal RRM domain is not (20). This also seems to be true for the *D.melanogaster* Snf protein (unpublished data).

In addition to the anti-morphic *1621* allele other *snf* mutations have been recovered (36). One of these, snf^{JA2}, deletes the C-terminal RRM domain. Remarkably, flies hemizgyous or homozygous for this hypomorphic allele are viable and fertile. By contrast, a deletion, snf^{J210}, that removes both the N- and C-terminal RRM domains is lethal. Like the antimorphic *snf* allele, snf^{1621}, both snf^{JA2} and snf^{J210} show synergistic female-lethal interactions with *Sxl* mutations; however, these interactions are less severe than those observed for snf^{1621} (36). The snf^{JA2}/snf^{J210} heteroallelic combination also exhibits defects in oogenesis similar to those found in snf^{1621} females.

In previous work, we have shown that Sxl and Snf are part of a large immunoprecipitable complex *in vivo*, and can interact directly with each other *in vitro*. (28). Since the experiments of Flickinger and Salz (36) suggest that the N-terminal RRM domain of *snf* is essential for both viability and *Sxl* autoregulation, we would expect that this domain mediates not only Snf-snRNA interactions but also Snf-Sxl interactions. On the other hand, though deletions of the

C-terminal *snf* RRM domain are viable, they are not completely wild type with respect to *Sxl* autoregulation. This raises the possibility that the C-terminal domain might also participate in Snf-Sxl interactions. To investigate these questions, we isolated GST fusion proteins containing either the N- or the C-terminal Snf RRM protein domains, and as a positive control, the full length Snf protein. These Snf proteins were then tested for interactions with Sxl using either a far western or co-immunoprecipitation assay.

In the experiment presented in Figure 6A, Sxl protein was bound to Sepharose beads by covalently coupled Sxl antibody. Aliquots of the Sxl:Sepharose beads were incubated with equal quantities of the three different GST-Snf fusion proteins. Any bound GST-Snf fusion protein was then visualized by western blotting using anti-GST antibody. As expected from previous studies (28), the GST-full length Snf protein but not GST alone (not shown) is co-immunoprecipitated by Sxl protein. Also detected in the Sxl immunoprecipitates are the N- and the C-terminal Snf fusion proteins. Usually the yield of the C-terminal Snf fusion protein is somewhat less than the N-terminal fusion protein; however the difference is small. Similar results were obtained when western blots of the three GST-Snf fusion proteins were probed with biotinylated Sxl protein (data not shown).

Sxl-Snf interaction is mediated primarily by the Sxl RRM1 domain

We next asked which domains of the Sxl protein are responsible for the Sxl:Snf protein interactions. For this purpose, partially purified full length Snf protein was bound to Sepharose beads by covalently cross-linked Snf antibody. After incubating the different GST-Sxl fusion proteins with the Snf beads, the Sxl protein associated with the Snf beads was visualized by western blotting using antibody directed against GST. The results of this experiment are presented in Figure 6B. Like full length Sxl (not shown; see 28) the GST fusion protein containing the two Sxl RRM domains, G-R12, can be detected in the Snf immunoprecipitates. In contrast, the unfused GST protein control is not found in the Snf immunoprecipitates. Of the GST-Sxl fusion proteins tested, only those containing the first RRM domain, R1, are immunoprecipitated by Snf. As can be seen in Figure 6B, both G-R1 and G-R1R1 are found in the Snf immunoprecipitates, while G-R2 and G-R2R2 are not. We also tested whether the Sxl N-terminus is able to interact with Snf; however, no interactions were observed (data not shown). The specificity for the R1 RRM domain evident in the Snf co-immunoprecipitation experiments was also observed in far western experiments when we probed blots of Snf protein with different biotinylated GST-Sxl fusion proteins. Only Sxl fusion proteins containing the R1 domain labeled Snf protein (data not shown).

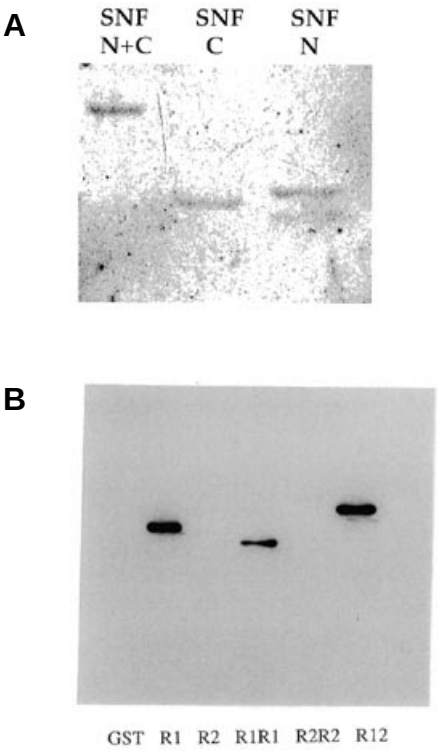


Figure 6. (**A**) Snf subdomains interact with Sxl. Purified Sxl protein was biotinylated as described in Materials and Methods. Sepharose beads containing covalently coupled Sxl antibody were incubated with Sxl protein. The Sxl antibody beads were then incubated with equal quantities of the different GST-Snf fusions. The beads from each incubation mixture were collected, washed and eluted by boiling in the presence of SDS–PAGE sample buffer. Samples were run on SDS–PAGE gels, transferred to nitrocellulose and detected by probing the western blots with anti-GST antibody. Samples as loaded from left to right: GST-Snf NC, GST-Snf C, GST-Snf N. Two bands are observed in the N-terminal RRM fusion protein lane. Both bands are detected in Coomassie stained gels. The lower band appears to correspond to a proteolytic breakdown product. (**B**) RRM R1 of Sxl recognises Snf. Partially purified Snf protein was bound to Sepharose beads containing covalently coupled Snf antibody . The Snf Sepharose beads were incubated with different Sxl sub-fragment GST fusions in the presence of immunoprecipitation buffer. The beads were washed, boiled with sample buffer and subsequently run on SDS–PAGE. The gels were transferred and blots probed with anti-GST antibody. Samples as loaded from left to right: GST alone, GST-R1, GST-R2, GST-R1R1, GST-R2R2, GST-R12.

DISCUSSION

Critical to properly executing its different regulatory functions, the Sxl protein must be able to recognize and bind with high affinity to appropriate target RNAs. Previous studies have shown that the preferred Sxl binding site corresponds to a stretch of eight or more contiguous U residues (27,31,46). Moreover, the protein exhibits a modest cooperativity in its interactions with RNA when the substrate contains two closely spaced binding sites (31,34). This cooperativity is likely to be relevant to the selection of the correct RNA species as many of the known *in vivo* regulatory targets for the protein contain multiple poly U runs. Of the four Sxl protein domains, RNA binding is mediated by the two internal domains RRM domains, R1 and R2. These domains are homologous in sequence and structure to similar domains found in a large family of RNA binding proteins (11,14,32,42,46,48). The function of the Sxl N-terminal and C-terminal domains is not known, though there have been reports that the N-terminus may play some role in the cooperative binding of the protein to RNA (34).

The presence of two RRM domains in Sxl is not unusual; many proteins in this family contain two or more RRM domains (49). For several of these proteins, the functional properties of the individual RRMs have been analyzed. To a first approximation, these multi-RRM proteins can be divided into two classes. One class is represented by the snRNP proteins, U1A, U2B″ and Snf (36,40,49–54). Only the N-terminal RRM domain of these proteins is required for specific binding to cognate snRNAs

(51,52). The C-terminal RRM is not involved in the recognition of snRNAs, and, at least in the case of U1A, it appears to bind to a completely different RNA (55). The second class of RRM proteins is represented by the general splicing factor U2AF which contains three RRM domains (56,57) and by poly A binding protein, PABP, which contains four RRM domains (58,59). Unlike the snRNP proteins, recognition of substrate RNAs by U2AF and PABP is not accomplished by a single RRM domain. For U2AF, maximal binding requires all three domains, though the individual RRMs appear to exhibit some limited specificity for RNAs containing a polypyrimidine tract (60). In the case of PABP, RNA binding activity can only be reconstituted by combining two or more RRM domains (36).

Our results and previous reports (31,45) indicate that Sxl falls into the second class of RRM proteins. Like the individual RRM domains of U2AF and PABP, the two RRM domains of Sxl, R1 and R2, show greatly reduced RNA binding activity on their own. Although the individual Sxl RRMs do not form protein:RNA complexes that are stable under our gel electrophoresis conditions, UV cross-linking experiments indicate that they bind RNA. Both are ~100-fold less active than the G-R12 protein which contains the normal combination of RRM domains. In addition, although each domain exhibits preferential binding to substrates containing the wild type U_8 site, competition experiments reveal that both have a reduced ability to discriminate between canonical and non-canonical RNA sequences.

Although these findings argue that two RRM domains are required for specific, high affinity binding, it is not clear why Sxl has two different domains, especially since the RNA recognition properties of R1 and R2 evident in the UV cross-linking assay seem nearly indistinguishable. One possibility is that the amino acid sequence differences between the two RRM domains are of no functional significance. Alternatively, the two RRM domains may have acquired specialized functions. To address this question, we generated a series of fusion proteins containing duplicated RRM domains separated by linkers of different lengths.

To a first approximation, duplication of the RRM domains reconstitutes much of the RNA affinity exhibited by the fusion protein, G-R12, containing both RRM domains. For three of the four duplicate fusion proteins (G-R1R1, G-R1ΔR1 and G-R2R2), the RNA binding activity as assayed by both gel shift and UV cross-linking approaches that of G-R12. The fourth duplicate, G-R2ΔR2, which has a shorter linker than its sister G-R2R2, is an exception; the binding activity of this duplicate is not much greater than the single R2 RRM. We presume that because of the shortened linker length, steric 'clashes' in the G-R2ΔR2 duplicate interfere with the simultaneous binding of the two R2 domains to the RNA substrate.

While our results indicate that duplication of the RRM domains can substantially increase RNA binding activity, the RNA recognition properties of the duplicate fusion proteins are clearly anomalous. Moreover, analysis of the duplicates reveals some differences between R1 and R2 that were not evident in the experiments with the single domain proteins. Although all four of the duplicates recognize and (to varying extents) favor the canonical Sxl protein binding sequence, they do not appear to discriminate between this sequence and other sequences as well as the wild type protein. Based on gel shift assays with RNAs containing Sxl-WT and Sxl-Mut binding sites it would appear that the substrate specificity of G-R2R2 is closest to that of G-R12, and this duplicate would be followed by the G-R1R1 fusion protein. However, unlike G-R12, the binding of both G-R2R2 and G-R1R1 to RNAs containing Sxl binding sites can be competed not only by the specific competitor poly U, but also by two non-specific competitors, poly UC and poly ACU.

In the case of G-R1R1, the efficacy of the non-specific competitors may be due at least in part to the tendency of this protein to form (what are presumably) dimeric complexes even on substrates that lack a canonical Sxl protein binding site. The formation of these non-specific dimeric complexes appears to be facilitated by interactions between G-R1R1 proteins bound to adjacent sites on the RNA. That such *trans* interactions are important in stabilizing protein:RNA complexes is supported by the finding that the G-R1R1 protein is unable to gel shift a short oligo containing a single Sxl binding site, but can gel shift an oligo containing two contiguous Sxl binding sites.

The detrimental effects of *trans* protein:protein interactions on RNA binding specificity are even more evident for G-R1ΔR1. When this duplicate binds to an RNA molecule it assembles into a multimeric protein array. The formation of this array does not seem to require a specific RNA sequence for nucleation, and arrays are assembled with nearly equal efficiency on RNA substrates that have or do not have canonical Sxl binding sites. It is curious that the reduced linker length in G-R1ΔR1 seems to enhance rather than suppress *trans* protein:protein interactions. One would have expected that a longer linker would impose fewer constraints on the spatial relationship between the two R1 domains in *cis* and hence be more permissive for *trans* interactions. Perhaps, the shorter linker in G-R1ΔR1 locks the two R1 domains into a configuration that promotes *trans* interactions.

Unlike the R1 duplicates, neither of the R2 duplicates exhibits a tendency to form non-specific dimeric or multimeric protein:RNA complexes. This is somewhat surprising since our assays for protein:protein interactions in the absence of RNA suggested that R2 may play a more important role than R1. One possibility is that when the R2 duplicates bind to RNA the configuration of the RRM domains is not especially favorable for the establishment of contacts with adjacent fusion proteins. Another possibility is that the R1 domain differs from R2 in that it undergoes a conformational change upon RNA binding that substantially enhances its potential for protein:protein interactions. Although RNA did not seem to be required for protein:protein interactions in either our immunoprecipitation or far western assays, Sakashita and Sakamoto (45) have described RNA-dependent interactions between Sxl fusion proteins. Finally, if (as argued below) the Sxl RRM domains are in some type of parallel (or head-to-tail), rather than an anti-parallel (or a head-to-head/tail-to-tail) configuration, the ladder of complexes observed for G-R1R1 would imply that the R1 domain has (at least) two surfaces that can participate in Sxl protein:protein interactions. It is possible that R2 has only one interaction surface.

Why are the RNA recognition properties of the duplicated RRMs not equivalent to those of the proteins containing both RRM domains? Two factors are probably important. First, the individual RRM domains have somewhat different activities in RNA binding and protein:protein interactions. Second, the two domains appear to function in concert in a manner that cannot be reproduced by proteins that consist of duplicated domains. One plausible speculation is that RNA recognition depends not only upon the intrinsic specificity of each RRM domain, but also upon a specificity that is imposed or generated by *cis* protein:protein interactions between the two domains. For example, *cis* protein:protein

interactions might align the RNA binding surface of the R1 and R2 domains in a manner which would maximize specific contacts with the string of uridine residues that define the canonical binding sequence. This alignment of the two RRMs would at the same time disfavor tight binding to RNA sequences that do not closely conform to the canonical sequence. Further specificity might be achieved if the R1 domain undergoes a conformational change when it binds to RNA which facilitates interactions in *cis* with the R2 domain. The proteins containing duplicated RRMs would differ from R12 in that the *cis* interactions between homologous domains would either not maximize specific contacts with poly U (G-R2R2) or would generate a configuration of RRM domains which is overly permissive (G-R1R1).

A number of observations are consistent with the idea that *cis* interactions between R1 and R2 might enhance recognition and binding to canonical target sequences. First, the R1 and R2 domains are capable of heterologous (as well as homologous) protein:protein interactions. Second, changing the linker length between R1 and R2 in the R1R2 protein did not have a major effect on either affinity or specificity, while it altered the RNA binding properties of the duplicate fusion proteins. Although our sample size is only two in each case, this would suggest that the surfaces of R1 and R2 that normally face one another in the Sxl protein may be designed to pack together. Third, when the order of the RRM domains is reversed, the resulting protein, G-R2-R1 binds poorly to RNAs containing a single Sxl binding site, but binds almost as well as G-R12 to RNAs containing two Sxl protein binding sites. The most plausible interpretation of this unusual behavior is that when there is only a single Sxl binding site, the normal *cis* interactions between R1 and R2 that stabilize the RNA:protein complex cannot occur. In contrast, when there are two Sxl binding sites, *trans* interactions between the R1 domain of one fusion protein and the R2 domain of the other recapitulate at least partially the *cis* interactions which normally ensure the tight binding of R12 to appropriate target RNAs.

Specific interactions between two contiguous RRM domains of the sort suggested here for Sxl have been observed in crystallographic studies on the hnRNP A1 protein (47,60). Each of the RRM domains in the A1 protein are folded in structures that conform closely to the canonical βαββαβ architecture exhibited by this protein motif. The folded domains are packed together in an anti-parallel arrangement, interfacing with each other along the 2nd α-helix. The two RRM domains are aligned with each other to form a single large RNA binding surface by two inter-domain Arg–Asp ion bridges. Moreover, it is has been proposed that these specific inter-domain contacts may play a critical role in RNA recognition by the A1 protein (47,60). Although we do not know how R1 and R2 pack in the Sxl protein, an anti parallel configuration like that found in A1 seems unlikely. In order to orient the two RRM domains of A1 in an anti parallel configuration, there must be a large linker separating the two domains. The linker between the the A1 RRMs is 17 amino acids, while it is only 9 amino acids in Sxl. It is difficult to see how the shorter Sxl linker would be compatible with the antiparallel packing observed in the hnRNP A1 protein. An additional problem is binding to RNA. In order to maintain the same 5′→3′ orientation of the RNA with respect to the binding surfaces of each domain, Shamoo *et al.* (47) suggested that the hnRNP A1 protein generates a loop when it binds to RNA. A loop would not be consistent with the RNA recognition properties of the Sxl protein. Our results as well as those of others suggest that the two Sxl RRMs bind to a contiguous stretch of uridine residues, presumably maintaining the same 5′→3′ orientation as the RNA leaves one domain and enters the domain. For these reasons, we suspect that the two Sxl RRM domains are most likely to pack together in a parallel (or head-to-tail) fashion.

A number of studies have suggested that the Sxl protein binds cooperatively to RNAs containing two closely spaced target sites. It has been reported that the N-terminal domain of the Sxl protein is responsible for these cooperative interactions (34,62). However, in our assays for Sxl protein:protein interactions, the Sxl N-terminus had no detectable activity. Instead, our experiments indicate that interactions between different Sxl proteins are most likely mediated by the RRM domains. The simplest model is that the R1 domain of one Sxl protein contacts the R2 domain of the adjacent protein. Since the Sxl protein appears to form dimers, but not multimeric arrays, this would imply that the configuration of the R1 and R2 domains in *cis* normally precludes *trans* interactions in both directions along the RNA. Given the propensity of the R1 duplicates to form dimers or multimers, it also possible that *trans* interactions are normally mediated by homologous R1:R1 contacts. (If both proteins bind in the same 5′→3′ orientation, homologous contacts would likely require the formation of a loop.)

Although recognition of appropriate target sequences is certainly essential, there are good reasons to believe that RNA binding is not in itself sufficient to account for the regulatory functions of the Sxl protein (11,31,34,61). In order to carry out these functions, the Sxl protein must be able to interact with components of the splicing and translational machinery (9,13,15,28). One protein that interacts directly with Sxl is Snf, a component of the fly U1 and U2 snRNPs. Our results indicate that this specialized function is mediated by the Sxl R1 RRM domain. It will be interesting to determine if specialization is the rule in RRM mediated interactions between Sxl and other cellular proteins. This would provide another reason why Sxl has two RRM domains that are similar but not precisely identical in amino acid sequence.

ACKNOWLEDGEMENTS

The authors would like to thank members of the laboratory for helpful discussions and for comments on the manuscript. This work was supported by a grant from NIH. We would also like to thank the anonymous reviewers for their helpful suggestions in revising the manuscript.

REFERENCES

1 Baker, B. S. (1989) *Nature*, **340**, 521–524.
2 Bopp, D., Bell, L. R., Cline, T. W. and Schedl, P. (1991) *Genes Dev.*, **5**, 403–415.
3 Cline, T. W. (1988) *Genetics*, **119**, 829–862.
4 Gergen, J. P. (1987) *Genetics*, **117**, 477–485.
5 Hodgkin, J. (1989) *Cell*, **56**, 905–906.
6 Keyes, L. N., Cline, T. W. and Schedl, P. (1992) *Cell*, **68**, 933–943.
7 Parkhurst, S. M. and Meneely, P. M. (1994) *Science*, **264**, 924–932.
8 Estes, P. A., Keyes, L. N. and Schedl, P. (1995) *Mol. Cell. Biol.*, **15**, 904–917.
9 Albrecht, E. B. and Salz, H. K. (1993) *Genetics*, **134**, 801–807.
10 Bell, L. R., Horabin, J. I., Schedl, P. and Cline, T. W. (1991) *Cell*, **65**, 229–239.
11 Bell, L. R., Maine, E. M., Schedl, P. and Cline, T. W. (1988) *Cell*, **55**, 1037–1046.
12 Burtis, K. C. and Baker, B. S. (1989) *Cell*, **56**, 997–1010.

13 Bashaw, G. J. and Baker, B. S. (1997) *Cell*, **89**, 789–798.
14 Inoue, K., Kazuyuki, H., Sakamoto, H. and Shimura, Y. (1990) *Nature*, **344**, 461–463.
15 Kelley, R. L., Wang, J., Bell, L. and Kuroda, M. (1997) *Nature*, **387**, 195–198.
16 Sosnowski, B. A., Belote, J. M. and McKeown, M. (1989) *Cell*, **58**, 449–459.
17 Sosnowski, B. A., David, D. D., Boggs, R. T., Madigan, S. J. and McKeown, M. (1994) *Dev. Biol*., **161**, 302–312.
18 Bernstein, M., Lersch, R. A., Subrahmanyan, L. and Cline, T. W. (1995) *Genetics*, **139**, 631–648.
19 Cline, T. W. (1984) *Genetics*, **107**, 231–277.
20 Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. and McKeown, M. (1987) *Cell*, **50**, 739–747.
21 Butler, B., Pirrotta, V., Irminger-Finger, I. and Nöthiger, R. (1986) *EMBO J*., **5**, 3607–3613.
22 Bashaw, G. J. and Baker, B. S. (1995) *Development*, **121**, 3245–3258.
23 Kelley, R. L., Solovyeva, I., Lyman, L., Richman, R., Solovyev, V. and Kuroda, M. (1995) *Cell*, **81**, 867–877.
24 Zhau, S., Yang, Y., Scott, M., Pannuti, A., Fehr, K., Eisen, A., Koonin, E., Fouts, D., Wrightsman, R., Manning, J. and Lucchesi, J. (1995) *EMBO J*., **14**, 2884–2895.
25 Valcárcel, J., Singh, R., Zamore, P. D. and Green, M. R. (1993) *Nature* (London), **362**, 171–175.
26 Horabin, J. I. and Schedl, P. (1993) *Mol. Cell. Biol*., **13**, 7734–7746.
27 Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y. and Shimura, Y. (1992) *Nucleic Acids Res*., **20**, 5533–5540.
28 Deshpande, G., Samuels, M. and Schedl, P. (1996) *Mol. Cell. Biol*., **16**, 5036–5047.
29 Salz, H. K., Maine, E. M., Keyes, L. N., Samuels, M. E., Cline, T. W. and Schedl, P. (1989) *Genes Dev*., **3**, 708–719.
30 Samuels, M. E., Schedl, P. and Cline, T. W. (1991) *Mol. Cell. Biol*., **11**, 3584–3602.
31 Samuels, M. E., Bopp, D., Colvin, R. A., Roscigno, R. F., Garcia-Blanco, M. A. and Schedl, P. (1994) *Mol. Cell. Biol*., **14**, 4975–4990.
32 Kanaar, R., Lee, A. L., Rudner, D. Z., Wemmer, D. E. and Rio, D. (1995) *EMBO J*., **14**, 4530–4539.
33 Sakashita, E. and Sakamoto, H. (1994) *Nucleic Acids Res*., **22**, 4082–4086.
34 Wang, J. and Bell, L. R. (1994) *Genes Dev*., **8**, 2072–2085.
35 Singh, R., Valcarcel, J. and Green, M. R. (1995) *Science*, **268**, 1173–1176.
36 Flickinger, T. W. and Salz, H. K. (1994) *Genes Dev*., **8**, 914–925.
37 Steinmann-Zwicky, M. (1988) *EMBO J*., **7**, 3889–3898.
38 Granadino, B., Campuzano, S. and Sanchez, L. (1992) *Genetics*, **130**, 597–612.
39 Hilfiker A., Amrein, H., Dubendorfer, A., Schneiter, R. and Nothiger, R. (1995) *Development*, **121**, 4017–4026.
40 Harper, D. S., Fresco, L. D. and Keene, J. D. (1992) *Nucleic Acids Res*., **20**, 3645–3650.
41 Salz, H. K. (1992) *Genetics*, **130**, 547–554.
42 Lee, A., Kannar, R., Rio, D. and Wemmer, D. E. (1994) *Biochemistry*, **33**, 13775–13786.
43 Govind, S., Whalen, A. and Steward, R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7861–7865.
44 Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Press, Cold Spring Harbor, NY.
45 Sakashita, E. and Sakamoto, H. (1996) *J. Biochem*., **120**, 1028–1033.
46 Shamoo, Y., Abdul-Manan, N. and Williams, K. R. (1995) *Nucleic Acids Res*., **23**, 725–728.
47 Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. and Steitz, T. A. (1997) *Nature, Struct. Biol*., **4**, 215–222.
48 Cartegni, L., Maconi, M., Morandi, E., Cobianchi F., Riva, S. and Biamonti, G. (1996) *J. Mol. Biol*., **259**, 337–348.
49 Bird, C. G. and Dryfess, G. (1994) *Science*, **265**, 615–621.
50 Jessen, T. H., Oubridge, C., Teo, C.-H., Pritchard, C. and Nagai, K. (1991) *EMBO J*., **10**, 3447–3456.
51 Nagai, K., Oubridge, C., Jessen, T., Li, J. and Evans, P. R. (1990) *Nature* (London), **348**, 515–520.
52 Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. and Nagai, K. (1994) *Nature*, **372**, 432–438.
53 Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. and Mattaj, I. W. (1990) *Nature*, **345**, 502–506.
54 Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J. and Mattaj, I. W. (1989) *EMBO J*., **8**, 4163–4170.
55 Lutz, C. S. and Alwine, J. C. (1994) *Genes Dev*., **8**, 576–586.
56 Zamore, P. D. and Green, M. R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9243–9247.
57 Zamore, P. D., Patton, J. G. and Green, M. R. (1992) *Nature* (London), **355**, 609–614.
58 Kuhn, U. and Pieler, T. (1996) *J. Mol. Biol*., **256**, 20–30.
59 Nietfield, W., Mentzel, H. and Pieler, T. (1990) *EMBO J*., **9**, 3699–3705.
60 Xu, R., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A. R. (1997) *Structure*, **5**, 559–570.
61 Bopp, D., Calhoun, G., Horabin, J., Samuels, M. and Schedl, P. (1996) *Development*, **122**, 971–982.
62 Wang, J., Dong, Z. and Bell, L. (l997) *J. Biol. Chem*., **272**, 22227–22235.