RNA (2000), *6*:901–911. Cambridge University Press. Printed in the USA.

Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF^{65} and the splicing repressor sex-lethal

RAVINDER SINGH,[1] HIREN BANERJEE,[1] and MICHAEL R. GREEN[2]

[1]Department of Molecular, Cellular, and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado 80309, USA
[2]Howard Hughes Medical Institute, Program in Molecular Medicine, University of Massachusetts Medical Center, Worcester, Massachusetts 01605, USA

ABSTRACT

The polypyrimidine-tract (Py-tract) adjacent to 3′ splice sites is an essential splicing signal and is recognized by several proteins, including the general splicing factor U2AF^{65} and the highly specific splicing repressor Sex-lethal (SXL). They both contain ribonucleoprotein-consensus RNA-binding motifs. However, U2AF^{65} recognizes a wide variety of Py-tracts, whereas SXL recognizes specific Py-tracts such as the nonsex-specific Py-tract of the *transformer* pre-mRNA. It is not understood how these seemingly similar proteins differentially recognize the Py-tract. To define these interactions, we used chemical interference and protection assays, saturation mutagenesis, and RNAs containing modified nucleotides. We find that these proteins recognize distinct features of the RNA. First, although uracils within the Py-tract are protected from chemical modification by both of these proteins, modification of any one of seven uracils by hydrazine, or any of eight phosphates by ethylnitrosourea strongly interfered with the binding of SXL only. Second, the 2′ hydroxyl groups or backbone conformation appeared important for the binding of SXL, but not U2AF^{65}. Third, although any of the bases (cytosine ≫ adenine > guanine) could substitute for uracils for U2AF^{65} binding, only guanine partially substituted for certain uracils for SXL binding. The different dependence on individual contacts and nucleotide preference may provide a basis for the different RNA-binding specificities and thus functions of U2AF^{65} and SXL in 3′ splice site choice.

Keywords: pre-mRNA splicing; ribonucleoprotein-consensus motif; RNA-binding proteins; RNA recognition motif; sex determination

INTRODUCTION

RNA-binding proteins regulate many important biological processes (Mattaj, 1993; Burd & Dreyfuss, 1994). RNA structure—hairpin stem-loop, internal bulge, or distorted RNA helices—plays a significant role in specific recognition by RNA-binding proteins (Nagai, 1996; Puglisi & Williamson, 1999; Steitz, 1999).

In higher eukaryotes, typical introns have a uracil-rich sequence or polypyrimidine-tract (Py-tract), which is an example of unstructured sequence, adjacent to the 3′ splice site where it serves as an important signal for both constitutive and regulated pre-mRNA splicing (Green, 1991; Nadal-Ginard et al., 1991; McKeown, 1992; Adams et al., 1996; Burge et al., 1999). Several proteins preferentially interact with the Py-tract, and thus have been classified as Py-tract-binding proteins. These include the U2 snRNP auxiliary factor (U2AF) (Ruskin et al., 1988; Zamore & Green, 1989), and the *Drosophila* protein Sex-lethal (SXL) (Bell et al., 1988).

Human U2AF^{65} (the 65-kDa, large subunit of U2AF) is an essential splicing factor that recognizes a diverse array of Py-tracts (Green, 1991; Burge et al., 1999). Early during spliceosome assembly, U2AF^{65} binds to Py-tract, interacts with other factors (Fleckner et al., 1997; Gozani et al., 1998), and facilitates the recruitment of U2 snRNP to pre-mRNA branchpoint sequence (Ruskin et al., 1988), at least in part by promoting an otherwise weak RNA–RNA base-pairing interaction (Valcarcel et al., 1996). As expected, the *Drosophila melanogaster* (Kanaar et al., 1993), *Schizosaccharomyces pombe* (Potashkin et al., 1993), and *Caenorhabditis*

Reprint requests to: Dr. Ravinder Singh, Department of Molecular, Cellular, and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado 80309, USA; e-mail: rsingh@colorado.edu.

elegans (Zorio & Blumenthal, 1999) homologs of U2AF^{65} are essential for viability.

Drosophila sexual differentiation involves a hierarchy of alternative splicing decisions (Cline & Meyer, 1996). The *Drosophila* protein SXL is the key binary switch between the male versus the female modes of somatic sexual differentiation; SXL is on in females and off in males. SXL also plays important roles in dosage compensation and germline sex determination (Cline & Meyer, 1996). In female flies, SXL protein regulates the splicing of the *transformer* (*tra*), *Sex-lethal* (*Sxl*), and *male-specific-lethal-2* (*msl2*) (Cline & Meyer, 1996) pre-mRNAs. SXL specifically binds to the nonsex-specific (NSS) Py-tract/3′ splice site of *tra* (Sosnowski et al., 1989; Inoue et al., 1990; Valcarcel et al., 1993). It competes with U2AF^{65} for the NSS Py-tract/3′ splice site and diverts U2AF^{65} to the lower affinity female-specific Py-tract/3′ splice site, thereby mediating 3′ splice site switching (Valcarcel et al., 1993). SXL controls the expression of *msl-2* pre-mRNA at the levels of splicing and translation (Bashaw & Baker, 1997; Kelley et al., 1997; Gebauer et al., 1998). Deletion of the N-terminal domain of SXL has been shown to uncouple the splicing and translational regulatory activities of SXL (Yanowitz et al., 1999).

Using iterative selection amplification or SELEX (Systematic Evolution of Ligands by EXponential enrichment) from a random RNA library, we found that the consensus SXL-binding site resembles the NSS Py-tract of *tra*, UUUUUGUUGUUUUUUUU (Singh et al., 1995), which is conserved among five *Drosophila* species (O'Neil & Belote, 1992). The A/G U_8C or U_8C (where U_8 is uracil-octamer) sequence, which is a portion of the above sequence, has also been proposed to be the SXL-binding site (Sakashita & Sakamoto, 1994; Samuels et al., 1994; Wang & Bell, 1994; Kanaar et al., 1995; Kelley et al., 1995).

Both U2AF^{65} and SXL contain ribonucleoprotein-consensus (RNP-CS) RNA binding motifs. The RNP-CS motif contains two short conserved sequences (RNP1 and RNP2) that form the two central strands of a four-stranded antiparallel β sheet platform packed against two α helices. Members of this largest family of RNA-binding proteins are involved in several aspects of RNA metabolism (Mattaj, 1993; Burd & Dreyfuss, 1994). U2AF^{65} and SXL, whose functions are well understood, serve as excellent models to define RNA recognition by the RNP-CS motif. Specifically, it remains to be understood how proteins with very similar RNA-binding motif show such a large difference in sequence recognition.

In this report, we have used several approaches to probe the interactions of these proteins with bases, sugars, and phosphates within Py-tracts. Our studies reveal that U2AF^{65} and SXL recognize different features of RNA. These findings may provide a basis for their distinct RNA-binding specificities that are relevant to the functions of these proteins in 3′ splice site selection.

RESULTS

To facilitate comparison between recombinant proteins GST-Δ1-63 U2AF^{65} and GST-SXL, referred hereafter as U2AF^{65} and SXL, we used the NSS Py-tract/3′ splice site of *tra* for all of the studies described here. Both proteins bind to this RNA sequence with high affinity. The binding conditions and affinities for the recombinant proteins have been previously described (Valcarcel et al., 1993; Singh et al., 1995). Over 90% of the RNA was in the protein-bound form for chemical protection, and approximately 50% of the RNA was in the protein-bound form for chemical interference and saturation mutagenesis.

Chemical protection

To define the binding site of SXL, we used chemical protection/footprinting assay (Moazed & Noller 1986), which identifies the functional groups of RNA that become inaccessible to chemical modification upon protein binding. The in vitro-transcribed RNA containing the NSS Py-tract of *tra* was incubated with SXL. The samples were then treated with 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT), which modifies the N3 position of uracils, and the sites of modification were mapped by primer extension. Figure 1 shows that SXL strongly protects the entire NSS Py-tract of *tra*. These results are consistent with our SELEX and mutagenesis experiments on SXL (Singh et al., 1995). We previously showed that U2AF^{65} protected the adenovirus major late Py-tract (Singh et al., 1995).

Chemical interference

The chemical-protection experiment identifies the nucleotides that are rendered inaccessible upon protein binding. However, it does not address the importance of individual contacts. Therefore, we performed a chemical-interference assay (Rymond & Rosbash, 1988; Conway & Wickens, 1989) that monitors the requirement of individual residues and thus complements the protection analysis in defining an RNA target. For this analysis, 5′-end-labeled transcripts corresponding to the NSS Py-tract/3′ splice site of *tra* were modified with diethylpyrocarbonate (DEPC), hydrazine (Rymond & Rosbash, 1988; Conway & Wickens, 1989), or ethylnitrosourea (ENU) (Vlassov et al., 1980), which modify purines, pyrimidines, and phosphates, respectively. The modified RNAs were incubated with U2AF^{65} or SXL to obtain approximately 50% binding, and the RNA:protein complexes were separated on a native polyacrylamide gel. The protein-bound and unbound RNA

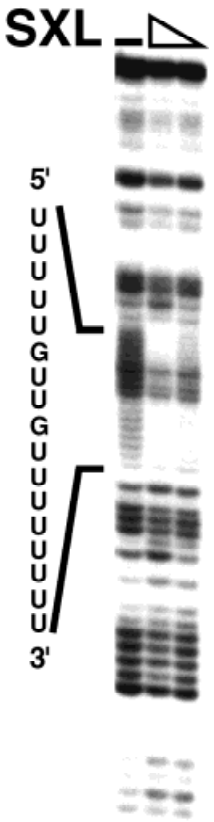

FIGURE 1. Chemical protection by SXL. In vitro-transcribed RNAs containing the NSS Py-tract/3′ splice site of *tra* pre-mRNA (*Bcl*1/*Fok*1 fragment) were incubated with SXL, such that more than 90% of the RNA was bound. The samples were treated with CMCT, which modifies the N3 positions of uracils, and the sites of modification were mapped by primer extension as described (Moazed & Noller, 1986). The NSS Py-tract is shown.

fractions were cleaved at the sites of modification and analyzed on a sequencing gel. A comparison of the cleavage pattern of the modified RNAs from protein-bound and unbound fractions reveals residues that are critical for protein binding; the bands corresponding to the nucleotides that are important for protein binding are excluded from the protein-bound fraction. To analyze the two guanines, the RNA was modified with DEPC, which carboxyethylates the N7 position of guanines. No detectable interference was observed for U2AF[65] or SXL with the RNAs modified with DEPC (data not shown). However, modification of any one of seven consecutive uracils of the Py-tract of *tra* with hydrazine, that opens the pyrimidine ring, strongly interfered with the binding of SXL (Fig. 2A); upstream uracils showed weaker effect. Some of these individual modifications showed barely detectable, if any, interference for U2AF[65] binding.

The interactions with backbone phosphates were similarly analyzed using ENU-modified RNAs. Modification of any one of eight consecutive phosphates within the Py-tract showed significant interference for the binding of SXL (Fig. 2B). In contrast, U2AF[65] showed barely detectable interference in the middle of the Py-tract. We conclude that SXL and U2AF[65] differ with respect to their dependence on individual contacts.

Discrimination between RNA and DNA

To address the role of 2′ hydroxyl groups in the binding of U2AF[65] and SXL, we compared the affinities of these proteins for the NSS Py-tract of *tra* containing either ribose (RNA) or 2′-deoxyribose sugars (DNA). U2AF[65] showed approximately three- to fivefold lower affinity for the DNA probe compared to the RNA of identical sequence (Fig. 3). In contrast, SXL showed no detectable binding to the DNA probe (Fig. 3).

Distinct nucleotide preference for U2AF[65] and SXL binding

Analysis of U2AF[65] and SXL-selected sequences

Differences in the RNA-binding properties of U2AF[65] and SXL (Figs. 2 and 3) prompted us to carefully re-examine their preference for various nucleotides within the uracil-rich sequences selected from a random pool of RNA (Singh et al., 1995). Figure 4 shows that U2AF[65] and SXL have distinct nucleotide preferences. For example, two to three cytosines (positions 7–9) interrupted two blocks of 5–8 uracil-rich sequences in the U2AF[65] binding site, whereas a GUUG/U sequence (positions 6–9) occupied a similar position in the SXL-binding site. Furthermore, any of the bases ($C \gg A > G$) substituted for uracils for U2AF[65] binding; guanine was the least preferred base. In contrast, only guanine partially substituted for certain uracils (positions 9, 11, 13, and 15) for SXL binding; cytosine and adenine were preferentially excluded from the SXL-binding site.

Saturation mutagenesis

There are limitations of the SELEX experiment (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990). For example, iterative amplification may preferentially affect enrichment of certain sequences depending on off-rate, RNA sequence, and/or structure. To exclude the possibility that these factors contributed to the observed nucleotide preference in the U2AF[65]- and SXL-selected sequences (Fig. 4), we performed the following saturation mutagenesis experiment, which avoided iterative amplification.

We doped, at an approximately 10% level, each of the positions of the NSS Py-tract of *tra* with three nonwild-type nucleotides. For example, adenine, cytosine, or guanine replaced uracils (odd and even series), and adenine, cytosine, or uracils replaced the two guanines (2G) in the binding site. Because the binding site is rather long, we chose every other position—odd

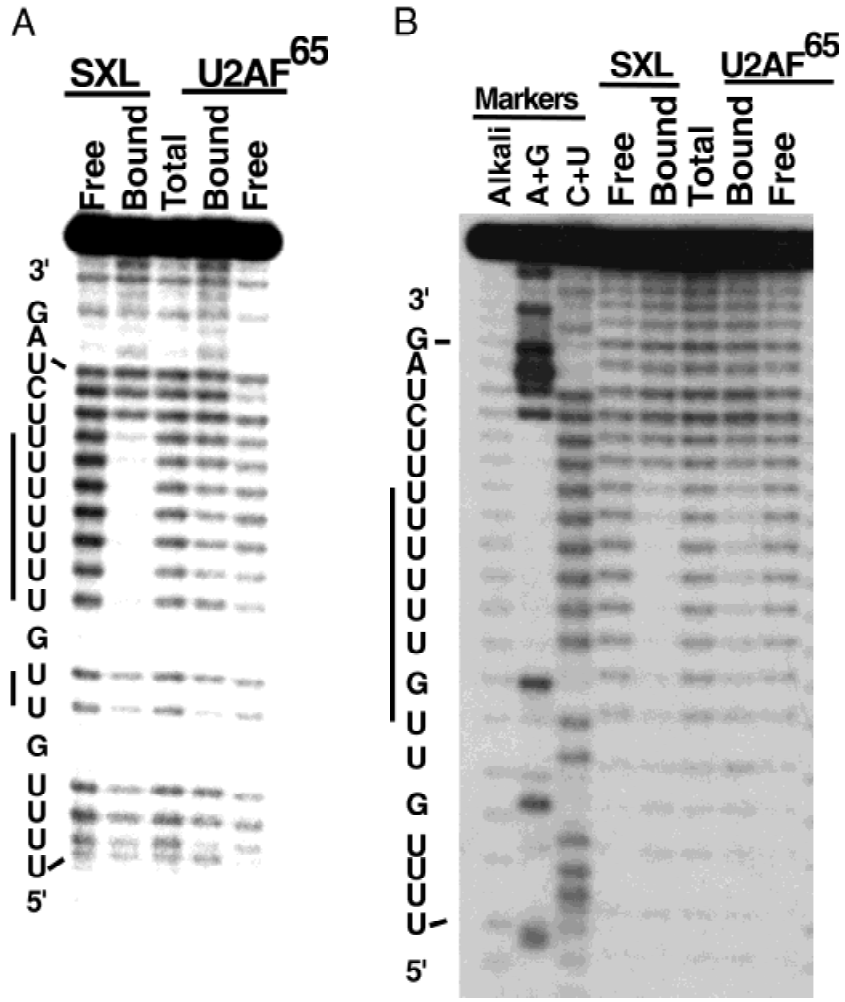

FIGURE 2. Chemical-interference analysis for U2AF^{65} and SXL. **A**: Hydrazine-modified RNA. **B**: ENU-modified RNA. In vitro-transcribed RNAs containing the NSS Py-tract/3′ splice site of the *tra* pre-mRNA were 5′-end labeled and modified with Hydrazine or ENU. The modified RNAs were incubated with U2AF^{65} or SXL, such that approximately 50% of the RNA was bound as described (Singh et al., 1995). The protein-bound and -unbound RNAs were separated on a native polyacrylamide gel. The sites of modification were identified by cleavage and analysis on a denaturing polyacrylamide gel. The sites of significant interference are shown by a vertical line. The RNA sequence is 5′-GGGUUUUGUUGUUUUUUUUCUAGUGUC-3′.

and even series—to span the entire binding site. This important consideration allowed detectable signal without multiple substitutions per molecule. These templates were transcribed in the presence of low levels of appropriate nucleoside 5′ α-thiotriphosphate (α S ATP for the A series, α S CTP for the C series, α S GTP for the G series, or α S UTP for the U series) to obtain single phosphorothioate substitutions per molecule (where sulfur replaces oxygen at the pro-Rp position of RNA backbone). The 5′-end-labeled RNAs were incubated with U2AF^{65} or SXL, such that approximately 50% of the RNA was bound, which is in the linear (thus sensitive) part of the binding curve. The bound RNA fraction was recovered in a filter-binding assay as described (Singh et al., 1995). The total RNA and the bound RNA fractions were cleaved with iodine at the sites of phosphorothioate linkages (Gish & Eckstein, 1988), and approximately equal amounts of radioactivity for each fraction were analyzed on a denaturing gel. It is expected that the nucleotide positions that are important for binding are excluded from the bound fraction, relative to those that are not important.

Given that detection required phosphorothioate backbone, it was necessary to distinguish whether the observed interference was the result of base or phosphorothioate substitution. Phosphorothioate substitution alone showed a detectable interference for SXL at several positions, especially in the 5′ half of the binding site (sU). As shown in Figure 5A, substitution of many of the uracils with either adenines or cytosines interfered with the binding of SXL (sA and sC). The positions that showed strong interference extended beyond those (positions 10–16) revealed by hydrazine modification (Fig. 2A). It is possible that for these upstream uracils, base substitution is more deleterious than pyrimidine ring opening by hydrazine. On the other hand,

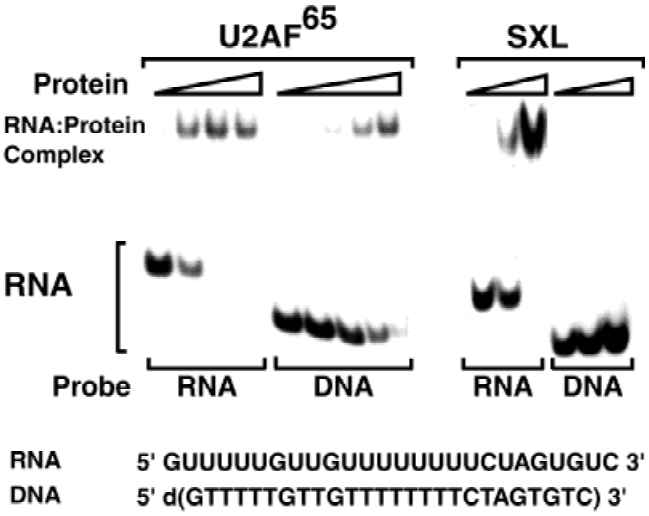

FIGURE 3. SXL and U2AF^{65} differentially recognize the RNA and DNA sequence. The RNA probe was transcribed by T7 RNA polymerase (WT-T7) (Milligan et al., 1987), and the DNA probe was chemically synthesized. The labeled probes were analyzed for U2AF^{65} and SXL binding in a gel mobility shift assay as described (Singh et al., 1995). The protein concentrations were as follows: GST-U2AF^{65}: 0.15 ng/μL, 0.45 ng/μL, 1.35 ng/μL, and 4.0 ng/μL for the RNA probe, and 0.15 ng/μL, 0.45 ng/μL, 1.35 ng/μL, 4.0 ng/μL, and 14.0 ng/μL for the DNA probe; GST-SXL: 0.03 ng/μL, 0.1 ng/μL, and 0.3 ng/μL for the RNA and the DNA probes.

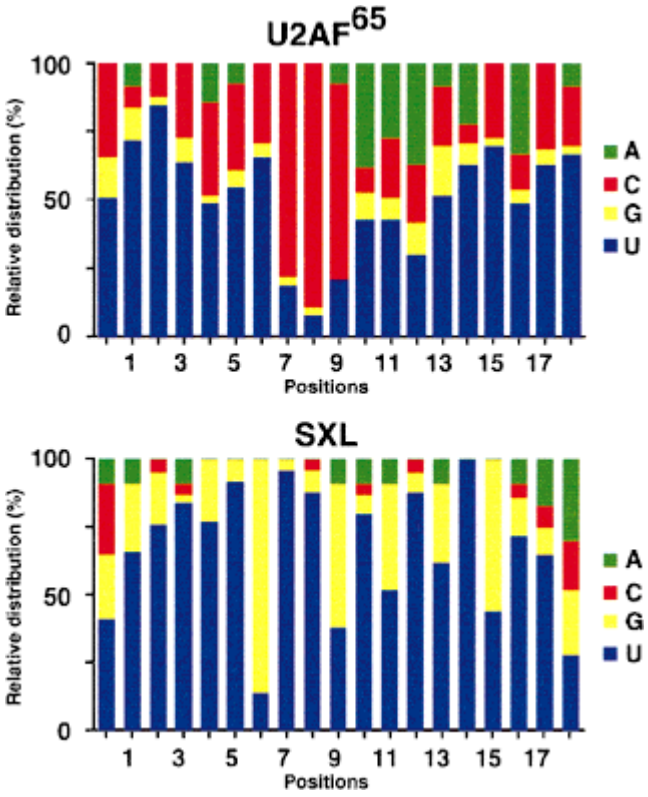

FIGURE 4. Distinct nucleotide preference for U2AF^{65} and SXL binding. Relative distribution of nucleotides at each of the positions within the binding site is shown for U2AF^{65} and SXL (modified from Singh et al., 1995).

only four of the guanines (sG) showed a small but detectable interference. We note that phosphorothioate substitutions (AUC) upstream of the UUUUUGUUGU UUUUUUU sequence also showed interference, including the positions (A and C) that were not doped. These interactions may contribute to binding affinity without affecting specificity. It should be pointed out that the effect of base substitution gets masked at those positions where phosphorothioate substitution is deleterious. Future studies should distinguish between base versus backbone contacts at these positions. In parallel experiments, there was no significant effect of these substitutions on U2AF^{65} binding (Fig. 5B). Occasionally, we observed barely detectable interference for U2AF^{65} from base substitution, but not phosphorothioate substitution.

These results indicate that U2AF^{65} and SXL show distinct nucleotide preferences, and that the SXL-binding site corresponds to the entire NSS Py-tract of *tra*.

Modified nucleotides

U2AF^{65} and SXL have high affinity for uracil-rich sequences. Uracils and cytosines differ only at the N3 and the O4 positions (Fig. 6A). This led us to investigate the interactions of these proteins with the functional groups at the N3 and the O4 positions of uracils in the Py-tract. For these studies we were guided by our observations that deoxythymidines could substitute for uracils in the Py-tract for the binding of U2AF^{65} but not SXL (Fig. 3), and that single-nucleotide substitution was unlikely to affect the binding of U2AF^{65} (Fig. 2A). Therefore, we synthesized RNAs containing a deoxythymidine (dT) analog in eight positions (Fig. 6A) and analyzed them in a mobility shift assay for U2AF^{65} binding. U2AF^{65} showed similar affinities for the uracil (wild-type) and deoxythymidine (dT)-containing RNAs (Fig. 6B). Next, we compared the binding affinities of U2AF^{65} for the RNAs containing deoxythymidines (dT), N3-methyl deoxythymidines (N3-methyl dT), or O4-methyl deoxythymidines (O4-methyl dT). Methyl groups at the N3 or the O4 positions decreased the binding of U2AF^{65} by approximately 80- to 100-fold (Fig. 6C). Availability of only the deoxythymidine analogs precluded our further analysis with SXL. We conclude that the modifications of the N3 and the O4 positions of uracils interfere with U2AF^{65} binding either directly or indirectly through effects on RNA structure.

DISCUSSION

The results presented here show that U2AF^{65} and SXL recognize different features of the Py-tract. First, U2AF^{65} and SXL protected multiple uracils from chemical modification, suggesting that these residues are in close contact. However, they differed with respect to depen-

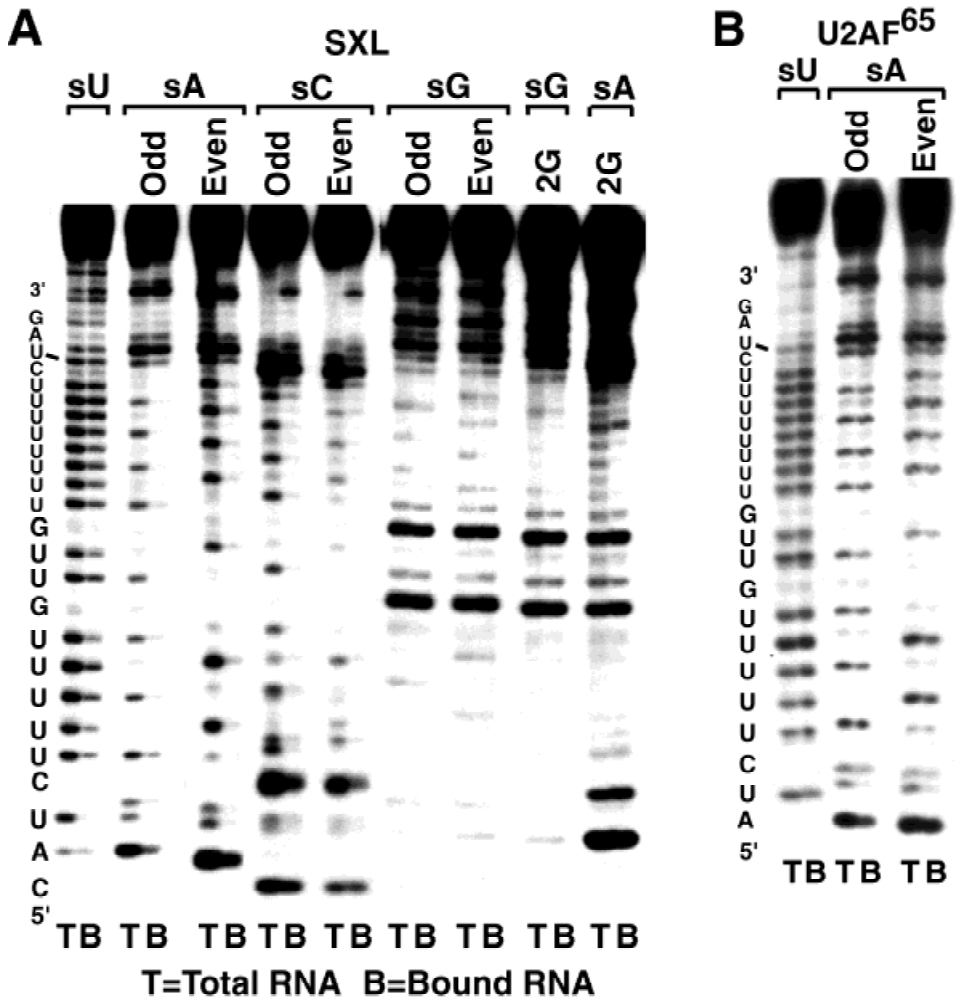

FIGURE 5. Saturation mutagenesis of the Py-tract by phosphorothioate approach. Appropriate phosphorothioate-containing RNAs were 5′-end labeled, incubated with SXL (**A**) or U2AF^{65} (**B**), such that approximately 50% of the RNA was bound, and total- and protein-bound RNA fractions were cleaved with iodine at the sites of phosphorothioate linkages, and analyzed on a denaturing polyacrylamide gel. The RNA sequence is: odd series: 5′-GGGCAXCXUXUXGXUGXUXUXUXUCXAGUGU AGUGAAC-3′; even series: 5′-GGGCAXCUXUXUGUXGUXUXUXUXCXAGUGUAGUGAAC-3′ where X = 90% U, 10% A, C, or G; 2G doping, 5′-GGGCAUCUUUUUXUUXUUUUUUUUUCUAGUGUAGUGAAC-3′, where X = 70% G, 10% each of A, C, and U.

dence on individual contacts: modification of any one of seven uracils by hydrazine or any of eight phosphates interfered with the binding of SXL but not U2AF^{65}. Second, they showed distinct requirements for the 2′ hydroxyl groups or the nucleic acid backbone conformation. Third, they showed distinct nucleotide preference. Finally, contacts with the N3 and/or the O4 positions could help explain recognition of uracil-rich sequences.

SELEX (Singh et al., 1995), chemical protection (Fig. 1; Samuels et al., 1994), and saturation mutagenesis (Fig. 5) experiments suggest that the high affinity SXL-binding site is the UUUUUGUUGUUUUUUUU sequence. Interactions with the N3 and/or the O4 positions offer a basis for the specificity of recognition of uracil-rich sequences by both proteins. We showed that the N3 and the O4 positions of the uracils are recognized by U2AF^{65} (Fig. 6C). Binding of the heterogeneous nuclear ribonucleoprotein C (hnRNP C), that also prefers uracil-rich sequences, was also strongly inhibited by the N3 or O4 methyl groups (data not shown). Although we could not directly analyze SXL binding to the RNAs containing the deoxythymidine derivatives, several considerations suggest that SXL contacts the N3 and/or the O4 positions. First, like U2AF^{65}, SXL protected the N3 position of uracils from CMCT modification (Fig. 1). Second, cytosines that differ from uracils only at the N3 and the O4 positions are excluded from the SXL-binding site in SELEX (Singh et al., 1995) and saturation mutagenesis (Fig. 5A) experiments. Third, guanine and uracil can partially substitute for each other for SXL binding (Fig. 4), presumably because they offer similar functional groups for recognition (Fig. 6A). The imino group at the N3 position and the carbonyl group

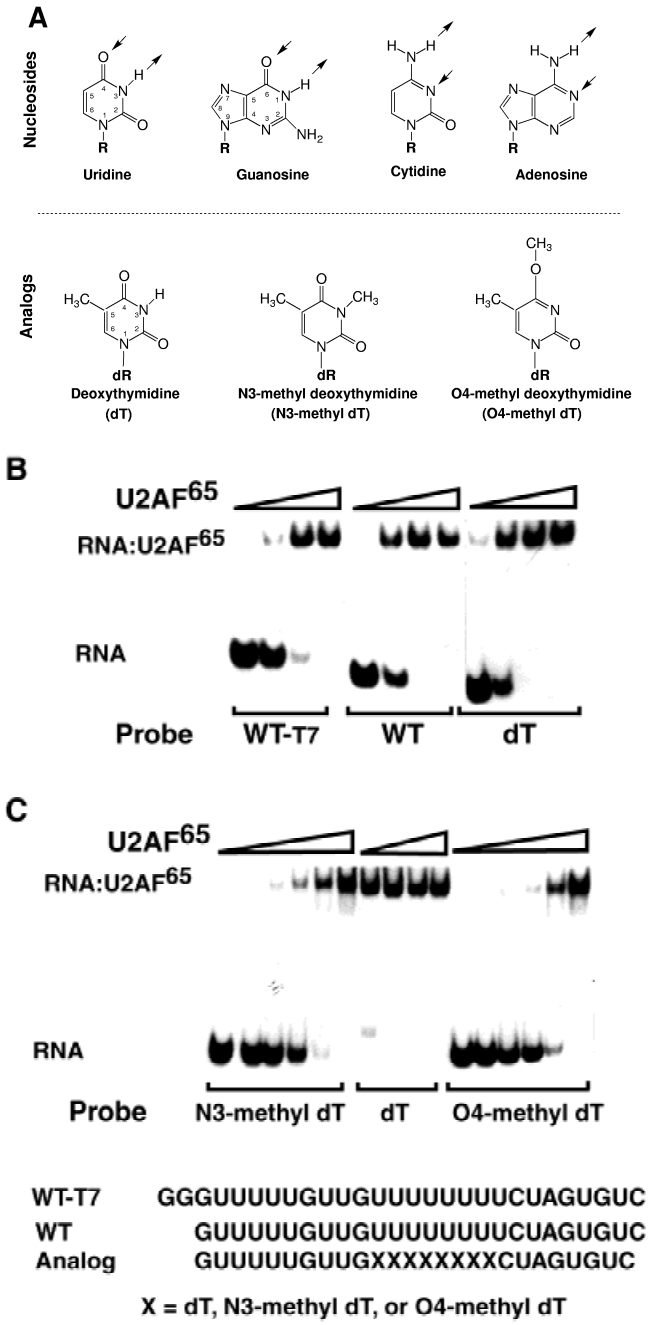

FIGURE 6. N3 and O4 positions of uracils are important for the binding of U2AF^{65}. **A**: Structures of nucleotides and their analogs. Arrows indicate potential H-bonding interactions. **B**: Binding of U2AF^{65} to the enzymatically synthesized (WT-T7) (Milligan et al., 1987), or the chemically synthesized wild type (WT), or deoxythymidine-containing (dT) probes. **C**: Binding of U2AF^{65} to the chemically synthesized probes containing methyl groups at the N3 (N3-methyl dT) or the O4 (O4-methyl dT) positions. Protein concentrations were as follows: 0.5 ng/μL, 1.5 ng/μL, 4.5 ng/μL, and 13.5 ng/μL for dT probe, and 0.5 ng/μL, 1.5 ng/μL, 4.5 ng/μL, 13.5 ng/μL, 40.5 ng/μL, and 121.5 ng/μL for the N3-methyl dT and the O4-methyl dT probes.

at the O4 position of uracil are identical to the N1 and the O6 positions of guanine, respectively. This simplest interpretation is consistent with the recently published X-ray structure of SXL bound to the NSS Py-tract of the *tra* mRNA, discussed below (Handa et al., 1999), and may help explain, at least in part, why cytosine and adenine, that offer different functional groups, are excluded from the binding site. Indeed, contacts with the N1 and/or the O6 positions are extensively used for specific recognition of poly(A) by the poly(A)-binding protein (Deo et al., 1999).

Our biochemical data on backbone- and base-specific contacts can be partly explained by the X-ray structure of SXL bound to the NSS Py-tract of *tra* (Handa et al., 1999); the numbering of nucleotides from Handa et al. (1999) is included below the sequence in Figure 7. For example, six of the nine 2′ hydroxyl groups are engaged in intermolecular (U8, U14, and U16) or intramolecular (U10–U12) interactions (Fig. 7), consistent with our finding that a DNA sequence (Fig. 3) or an RNA containing deoxyuridines (Kanaar et al., 1995) is not recognized by SXL. Thus the effect of 2′ hydroxyls on SXL binding could result from either the loss of direct contacts with SXL or indirect effects through structural requirements for the RNA kink around U11, that includes three direct internucleotide hydrogen bonds between the 2′ hydroxyl and phosphate groups of U10, U12, and U13. Finally, the effect on nucleic acid backbone conformation, which is unusually C2′-*endo*-rich for the SXL–RNA complex (Handa et al., 1999), could also indirectly contribute to the loss of SXL binding to DNA; only a few sugars are in the C2′-*endo* form in the U1A complex (Oubridge et al., 1994). It will be interesting to determine how U2AF^{65}, which binds to DNA and RNA sequences equally well, recognizes RNA. In this regard, U2AF^{65} is similar to hnRNP A1 (UP1), which also binds to RNA and DNA sequences relevant to pre-mRNA splicing or present at chromosomal ends (telomeres) (Ishikawa et al., 1993; Ding et al., 1999).

Only four (U11, U13–U15) of the six phosphates (G9, U10, U11, U13–U15) that are involved in intra- and intermolecular interactions in the X-ray structure showed ENU-modification interference (Fig. 7). Again, the effect may result from the loss of direct interactions with protein (U11, U14, and U15) or indirect effect on the formation of the RNA kink (U10 and U13). The U12, U16, and U17 phosphates, that show no inter- or intramolecular interactions in the X-ray structure, showed interference. On the other hand, the G9 and U10 phosphates, which interact with SXL in the X-ray structure, showed weak or no interference. Replacement of Arg252, which interacts with the G9 phosphate, by alanine is known to abolish SXL binding (Lee et al., 1997).

Our biochemical/molecular data (Singh et al., 1995; Figs. 1 and 5) and NMR analysis (Kanaar et al., 1995) support interactions with the N3 and/or O4 positions and are consistent with the X-ray structure data (Handa et al., 1999). These findings may explain the preference for uracil-rich sequence and thus specificity of SXL. For example, the strong interference from hydrazine modification of uracils (Fig. 2A) or from saturation

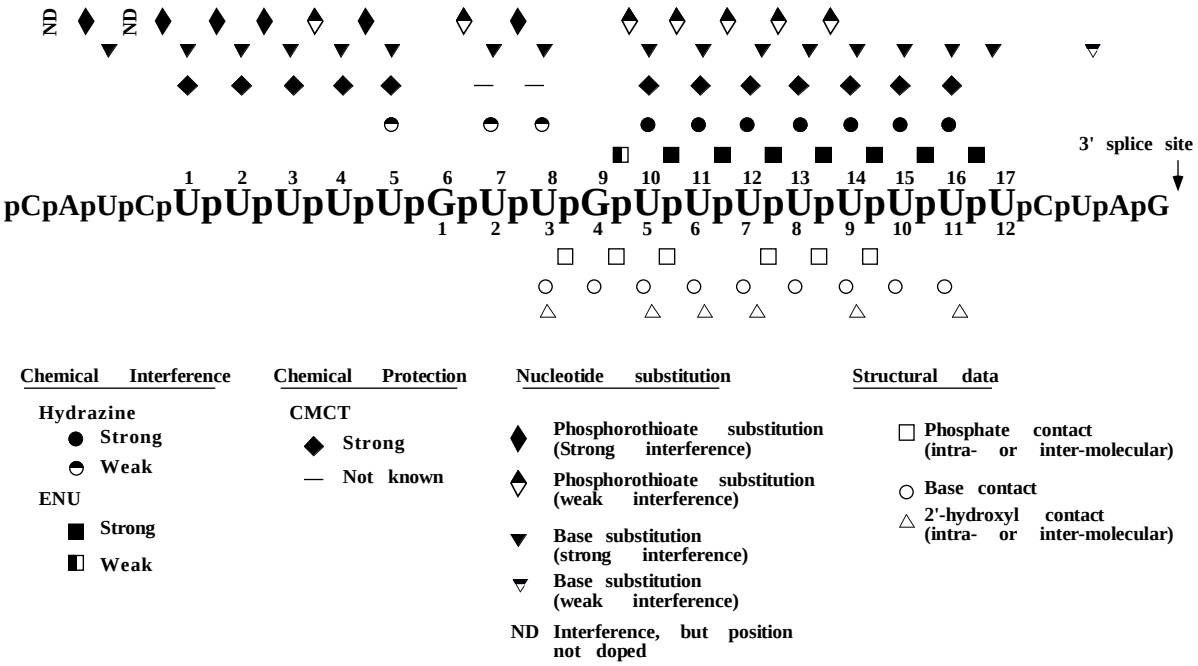

FIGURE 7. Summary of the SXL-Py-tract interactions based on biochemical and structural data. Because the first five residues (uracils) were degraded during X-ray crystallization, the numbering of nucleotides from Handa et al. (1999) is included below the sequence. Because the 3′ termini resulting from aniline and ENU cleavage are different (Ehresmann et al., 1987), our assignment relied on the observation that a fragment (15- as well as 19-mer) containing a 3′ phosphate (ENU or RNase T1 cleavage) migrates faster by the equivalent of ~1 nt relative to that lacking a 3′ phosphate (aniline cleavage) (Been & Cech, 1987; Singh, unpubl.).

mutagenesis (Fig. 5A) can be explained because the N3 and/or the O4 positions of eight uracil moieties (U8 and U10–U16) are recognized by amino acid side chains or α-carbon backbone. The O2 position of five uracils (U5, U6, and U9–U11) is always recognized in combination with the N3, O4, or both; O2 alone cannot distinguish between pyrimidines. Although interaction of SXL with the 2-amino position of the guanine-9 in the X-ray structure (Handa et al., 1999) is consistent with our SELEX data (Singh et al., 1995) and phylogenetic conservation of this residue (O'Neil & Belote, 1992), its substitution to uracil does not significantly compromise SXL regulation in vitro (Singh et al., 1995), and a uracil is present at the corresponding position in the *msl2* pre-mRNA (Zhou et al., 1995). In this context, it is unclear how the register is maintained.

One important aspect of our combined experiments that remains to be explained from the X-ray structure is the contribution of the first five uracils of the UUUUU GUUGUUUUUUUU sequence to SXL binding and splicing regulation. These residues were degraded during crystallization (Handa et al., 1999) and could account for some of the observed differences. We conclude that these residues are important for SXL binding because they are present in the sequences selected by SXL (Singh et al., 1995), protected from chemical modification upon SXL binding (Fig. 1; Samuels et al., 1994), revealed by saturation mutagenesis (Fig. 5A), conserved among five *Drosophila* species (O'Neil & Belote, 1992), and required for SXL regulation in vitro (Singh et al., 1995) and in vivo (Sosnowski et al., 1994). Although uracil-rich sequences are present in all known SXL targets, *tra*, *Sxl*, *msl-2*, and *msl-1* (Sosnowski et al., 1989; Inoue et al., 1990; Sakamoto et al., 1992; Horabin & Schedl, 1993; Bashaw & Baker, 1997; Kelley et al., 1997; Chang & Kuroda, 1998; Gebauer et al., 1998), the binding sites in other RNAs remain to be systematically characterized. Furthermore, the effect of cofactors encoded by genes such as *snf*, *fl(2)d*, and *vir* (Oliver et al., 1988; Steinmann-Zwicky, 1988; Granadino et al., 1990; Hilfiker & Nothiger, 1991; Granadino et al., 1992; Albrecht & Salz, 1993), or the amino-terminal domain of SXL (Wang & Bell, 1994; Granadino et al., 1997) on the RNA-binding properties of SXL remains to be understood. Given that the mechanisms of splicing regulation may not be identical for every target (Sakamoto et al., 1992; Horabin & Schedl, 1993; Granadino et al., 1997; Yanowitz et al., 1999), future studies should clarify sequence requirements for splicing regulation by SXL in other RNAs.

There are notable differences in the arrangement of RNP-CS motif(s) with respect to each other and interactions with the target RNA in known RNA–protein complexes. For example, the single RNP-CS motif of U1A and U2B″/U2A′ (leucine-rich repeat protein) form a distinct network of interactions with target sequences in U1

and U2 snRNAs (Oubridge et al., 1994; Price et al., 1998). On the other hand, the two RNP-CS motifs of SXL form a V-shaped cleft (Handa et al., 1999), PABP forms a continuous RNA-binding trough (Deo et al., 1999), and hnRNP A1 (UP1) forms a dimer such that RNP-CS1 of one monomer and the RNP-CS2 of the other contact one RNA (Ding et al., 1999). RNP-CS motif appears to be a versatile RNA-binding domain. Thus, differences between U2AF^{65} and SXL are less surprising. We propose that the binding of U2AF^{65} to the uracil-rich sequences may involve redundant contacts that obscure the effects of single-nucleotide substitutions or chemical modifications. It is tempting to speculate that interactions with the O2 positions may be relevant to U2AF^{65} because cytosines are tolerated. The O2 position, which alone cannot distinguish between pyrimidines, is always recognized in combination with the N3, O4, or both for SXL and hnRNP A1 (UP1) (Ding et al., 1999; Handa et al., 1999). Steric considerations for purines and interference from the 2-amino group may exclude guanine, which is the least preferred base for U2AF^{65}. Stacking interactions may also partly contribute to the different RNA-binding properties of U2AF^{65}. These features—dispensability of individual contacts and tolerance for other bases—may help explain how the general splicing factor U2AF^{65} is able to recognize a wide variety of natural Py-tracts (Green, 1991).

The N3 and the O4 positions are more likely to be recognized in the context of an unstructured, single-stranded RNA sequence such as the Py-tract (Saenger, 1994). In DNA recognition, the major groove is the key determinant of specificity. However, a typical RNA major groove, that offers a rich ensemble of discriminatory functional groups, including the N3 and the O4 positions, is deep and narrow, and thus inaccessible to proteins (Steitz, 1999). This problem has been circumvented in part by contacting distorted RNA helices, loops, or bulges in the contexts of structured RNAs (Nagai, 1996; Puglisi & Williamson, 1999; Steitz, 1999). On the other hand, our study suggests a basis for the recognition of the Py-tract, which is in largely extended conformation and lacks intramolecular base pairing. In addition, it provides important insights into how U2AF^{65} and SXL may achieve different levels of specificity, which is relevant to the functions of these proteins in 3′ splice site choice, by recognizing different features of the Py-tract, for example varying dependence on individual contacts and 2′ hydroxyl groups and distinct nucleotide preference.

MATERIALS AND METHODS

Materials

GST-U2AF^{65} Δ1-63, GST-SXL proteins and *tra* RNA for interference analysis were previously described (Valcarcel et al., 1993; Singh et al., 1995). *tra* RNA for CMCT chemical-protection assay corresponds to the *Bcl*1/*Fok*1 fragment of the NSS Py-tract/3′ splice site of *transformer* gene. The chemically synthesized RNAs were synthesized on an Applied Biosystems (ABI Model 392) DNA synthesizer. Phosphoramidite for O4 methyl dT was from Glen Research (Virginia), for N3 methyl dT from Chemgen (Massachusetts), and for ribonucleotides from BioGenex (California). Nucleoside 5′-α thiotriphosphates were from NEN (Boston, Massachusetts).

RNA binding

RNA-binding reactions (10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 50 mM KCl, 0.5 U/μL RNasin, 0.09 μg/μL acetylated bovine serum albumin, 0.15 μg/μL tRNA, and various concentrations of recombinant protein) were essentially as described (Valcarcel et al., 1993; Singh et al., 1995). Protein concentrations were determined by the staining of a sodium dodecyl sulphate polyacrylamide gel with Coomassie Brilliant Blue R-250 with bovine serum albumin as a standard.

RNA sequences

For chemical protection, *Bcl*1/*Nsi*I fragment of *tra* (Valcarcel et al., 1993) was cloned into the *Bam*H1/*Pst*I site of pGEM3 (Promega). The *Fok*1-digested template was used for transcription in vitro by T7 RNA polymerase. Other RNAs have been described in the figure legends.

Chemical protection

Nonradioactive *Bcl*1/*Fok*1 RNA was incubated with CMCT, and chemical protection was performed as described (Moazed & Noller, 1986).

Chemical interference

Chemical-interference analysis was performed essentially as described (Rymond & Rosbash, 1988; Conway & Wickens, 1989). Briefly, *tra* RNA was modified with DEPC, hydrazine, or ENU to achieve approximately one modification per molecule. The modified RNAs were incubated with U2AF^{65} or SXL, such that approximately 50% of the RNA was bound. The bound and the unbound fractions were separated on a native polyacrylamide gel as described (Singh et al., 1995). The sites of modification were identified by aniline cleavage for hydrazine (Rymond & Rosbash, 1988; Conway & Wickens, 1989) or mild alkali hydrolysis for ENU (Vlassov et al., 1980), and analyzed by electrophoresis on a 20% wedge-shaped denaturing polyacrylamide gel.

Saturation mutagenesis

Transcription reaction [1× T7 transcription buffer (Promega), 10 mM dithiothreitol, 1 μM doped oligonucleotide, 1 μM T7 oligonucleotide, 2 mM GTP, 1 mM each ATP, CTP, and UTP, and 1–2 U/μL T7 RNA polymerase] was incubated for 2 h at 37 °C (Milligan et al., 1987). For odd and even series, phosphorothioate concentrations were 0.167 mM α-thio ATP, 0.05 mM α-thio UTP, 0.2 mM α-thio CTP, or 0.2 mM α-thio GTP, and were adjusted accordingly for the two-guanine dop-

ing. Cleavage of the RNA at the sites of phosphorothioate incorporation with iodine was as described (Gish & Eckstein, 1988). T7 Primer: 5′-gTAATACgACTCACTATAg-3′; U-doping (10% A, C, or G) at odd positions: 5′-gTTCACTACACTXgAXAXAXCAXCXAXAXgXTgCCCTATAgTgAgTCgTATTAC-3′; U-doping (10% A, C, or G) at even positions: 5′-gTTCACTACACTXgXAXAXAXACXACAXAXAgXTgCCCTATAgTgAgTCgTATTAC-3′; G doping (30% A, C, or U) at two guanines: 5′-gTTCACTACACTAgAAAAAAAAXAAXAAAAAgATgCCCTATAgTgAgTCgTATTAC-3′. X represents the site of doping.

Nucleotide analogs

RNAs containing nucleotide analogs were chemically synthesized and deprotected in ethanolic ammonia or DBU/anhydrous methanol for 24 h. Subsequently, the samples were vacuum dried and the 2′ hydroxyl groups were deprotected in 1 M TBAF/THF for 24 h. The RNA was purified through a DEAE column, ethanol precipitated, and 5′-end labeled for binding experiments.

ACKNOWLEDGMENTS

We thank Tom Cech, Olke Uhlenbeck, Juan Valcarcel, and Andrew Rahn for critical comments on the manuscript; Lisa Hegg for advice on RNA synthesis; Maria Zapp for oligonucleotide synthesis; and Art Zaug for discussion about the assignment of phosphates. This work was supported by grants from the National Institutes of Health to MRG (GM35490) and RS (GM58576).

Received February 21, 2000; returned for revision March 15, 2000; revised manuscript received March 28, 2000

REFERENCES

Adams MD, Rudner DZ, Rio DC. 1996. Biochemistry and regulation of pre-mRNA splicing. *Curr Opin Cell Biol 8*:331–339.

Albrecht EB, Salz HK. 1993. The *Drosophila* sex determination gene *snf* is utilized for the establishment of the female-specific splicing pattern of Sex-lethal. *Genetics 134*:801–807.

Bashaw GJ, Baker BS. 1997. The regulation of the *Drosophila msl-2* gene reveals a function for Sex-lethal in translational control. *Cell 89*:789–798.

Been MD, Cech TR. 1987. Selection of circularization sites in a group I IVS RNA requires multiple alignments of an internal template-like sequence. *Cell 50*:951–961.

Bell LR, Maine EM, Schedl P, Cline TW. 1988. Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell 55*: 1037–1046.

Burd CG, Dreyfuss G. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science 265*:615–621.

Burge CB, Tuschl T, Sharp PA. 1999. Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 525–560.

Chang KA, Kuroda MI. 1998. Modulation of *MSL1* abundance in female *Drosophila* contributes to the sex specificity of dosage compensation. *Genetics 150*:699–709.

Cline TW, Meyer BJ. 1996. Vive la difference: Males vs. females in flies vs. worms. *Annu Rev Genet 30*:637–702.

Conway L, Wickens M. 1989. Modification interference analysis of reactions using RNA substrates. *Methods Enzymol 180*:369–379.

Deo RC, Bonanno JB, Sonenberg N, Burley SK. 1999. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell 98*: 835–845.

Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM. 1999. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes & Dev 13*:1102–1115.

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel J-P, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res 15*:9109–9127.

Ellington AD, Szostak JW. 1990. In vitro selection of RNA molecules that bind specific ligands. *Nature 346*:818–822.

Fleckner J, Zhang M, Valcarcel J, Green MR. 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes & Dev 11*:1864–1872.

Gebauer F, Merendino L, Hentze MW, Valcarcel J. 1998. The *Drosophila* splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *RNA 4*:142–150.

Gish G, Eckstein F. 1988. DNA and RNA sequence determination based on phosphorothioate chemistry. *Science 240*:1520–1522.

Gozani O, Potashkin J, Reed R. 1998. A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol 18*:4752–4760.

Granadino B, Campuzano S, Sanchez L. 1990. The *Drosophila melanogaster fl(2)d* gene is needed for the female-specific splicing of Sex-lethal RNA. *EMBO J 9*:2597–2602.

Granadino B, Penalva LOF, Green MR, Valcarcel J, Sanchez L. 1997. Distinct mechanisms of splicing regulation in vivo by the *Drosophila* protein Sex-lethal. *Proc Natl Acad Sci USA 94*:7343–7348.

Granadino B, San Juan A, Santamaria P, Sanchez L. 1992. Evidence of a dual function in *fl(2)d*, a gene needed for Sex-lethal expression in *Drosophila melanogaster*. *Genetics 130*:597–612.

Green MR. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu Rev Cell Biol 7*:559–599.

Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S. 1999. Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature 398*:579–585.

Hilfiker A, Nothiger R. 1991. The temperature-sensitive mutation *vir*^{ts} (*virilizer*) identifies a new gene involved in sex determination of *Drosophila*. *Roux Arch Dev Biol 200*:240–248.

Horabin JI, Schedl P. 1993. Sex-lethal autoregulation requires multiple *cis*-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5′ splice site. *Mol Cell Biol 13*:7734–7746.

Inoue K, Hoshijima K, Sakamoto H, Shimura Y. 1990. Binding of the *Drosophila* sex-lethal gene product to the alternative splice site of *transformer* primary transcript. *Nature 344*:461–463.

Ishikawa F, Matunis MJ, Dreyfuss G, Cech TR. 1993. Nuclear proteins that bind the pre-mRNA 3′ splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol Cell Biol 13*:4301–4310.

Kanaar R, Lee AL, Rudner DZ, Wemmer DE, Rio DC. 1995. Interaction of the sex-lethal RNA binding domains with RNA. *EMBO J 14*:4530–4539.

Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: Requirement for viability. *Science 262*:569–573.

Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V, Kuroda MI. 1995. Expression of *msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell 81*:867–877.

Kelley RL, Wang J, Bell L, Kuroda MI. 1997. Sex lethal controls dosage compensation in *Drosophila* by a non-splicing mechanism. *Nature 387*:195–199.

Lee AL, Volkman BF, Robertson SA, Rudner DZ, Barbash DA, Cline TW, Kanaar R, Rio DC, Wemmer DE. 1997. Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein sex-lethal. *Biochemistry 36*:14306–14317.

Mattaj IW. 1993. RNA recognition: A family matter? *Cell 73*:837–840.

McKeown M. 1992. Alternative mRNA splicing. *Annu Rev Cell Biol 8*:133–155.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res 15*:8783–8798.

Moazed D, Noller HF. 1986. Transfer RNA shields specific nucleo-

tides in 16S ribosomal RNA from attack by chemical probes. *Cell 47*:985–994.
Nadal-Ginard B, Smith CW, Patton JG, Breitbart RE. 1991. Alternative splicing is an efficient mechanism for the generation of protein diversity: Contractile protein genes as a model system. *Adv Enzyme Regul 31*:261–286.
Nagai K. 1996. RNA–protein complexes. *Curr Opin Struct Biol 6*: 53–61.
Oliver B, Perrimon N, Mahowald AP. 1988. Genetic evidence that the sans fille locus is involved in *Drosophila* sex determination. *Genetics 120*:159–171.
O'Neil MT, Belote JM. 1992. Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics 131*:113–128.
Oubridge C, Ito N, Evans PR, Teo CH, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.
Potashkin J, Naik K, Wentz-Hunter K. 1993. U2AF homolog required for splicing in vivo. *Science 262*:573–575.
Price SR, Evans PR, Nagai K. 1998. Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature 394*:645–650.
Puglisi JD, Williamson JR. 1999. RNA interactions with small ligands and small peptides. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 403–425.
Robertson DL, Joyce GF. 1990. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature 344*:467–468.
Ruskin B, Zamore PD, Green MR. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell 52*: 207–219.
Rymond BC, Rosbash M. 1988. A chemical modification/interference study of yeast pre-mRNA spliceosome assembly and splicing. *Genes & Dev 2*:428–439.
Saenger W. 1994. Defining terms for nucleic acids. In: Cantor CR, ed. *Principles of nucleic acid structure*. New York: Springer-Verlag Inc. pp 21–23.
Sakamoto H, Inoue K, Higuchi I, Ono Y, Shimura Y. 1992. Control of *Drosophila* Sex-lethal pre-mRNA splicing by its own female-specific product. *Nucleic Acids Res 20*:5533–5540.
Sakashita E, Sakamoto H. 1994. Characterization of RNA binding specificity of the *Drosophila* sex-lethal protein by in vitro ligand selection. *Nucleic Acids Res 22*:4082–4086.
Samuels ME, Bopp D, Colvin RA, Roscigno RF, Garcia-Blanco MA, Schedl P. 1994. RNA binding by Sxl proteins in vitro and in vivo. *Mol Cell Biol 14*:4975–4990.
Singh R, Valcarcel J, Green MR. 1995. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science 268*:1173–1176.
Sosnowski BA, Belote JM, McKeown M. 1989. Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. *Cell 58*:449–459.
Sosnowski BA, Davis DD, Boggs RT, Madigan SJ, McKeown M. 1994. Multiple portions of a small region of the *Drosophila transformer* gene are required for efficient in vivo sex-specific regulated RNA splicing and in vitro sex-lethal binding. *Dev Biol 161*: 302–312.
Steinmann-Zwicky M. 1988. Sex determination in *Drosophila*: The X-chromosomal gene *liz* is required for Sxl activity. *EMBO J 7*:3889–3898.
Steitz TA. 1999. RNA recognition by proteins. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 427–450.
Tuerk C, Gold L. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science 249*:505–510.
Valcarcel J, Gaur RK, Singh R, Green MR. 1996. Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science 273*:1706–1709.
Valcarcel J, Singh R, Zamore PD, Green MR. 1993. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA. *Nature 362*:171–175.
Vlassov VV, Giege R, Ebel JP. 1980. The tertiary structure of yeast tRNAPhe in solution studied by phosphodiester bond modification with ethylnitrosourea. *FEBS Lett 120*:12–16.
Wang J, Bell LR. 1994. The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes & Dev 8*:2072–2085.
Yanowitz JL, Deshpande G, Calhoun G, Schedl PD. 1999. An N-terminal truncation uncouples the sex-transforming and dosage compensation functions of sex-lethal. *Mol Cell Biol 19*:3018–3028.
Zamore PD, Green MR. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA 86*:9243–9247.
Zhou SB, Yang YF, Scott MJ, Pannuti A, Fehr KC, Eisen A, Koonin EV, Fouts DL, Wrightsman R, Manning JE, Lucchesi JC. 1995. Male-specific lethal 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO J 14*:2884–2895.
Zorio DA, Blumenthal T. 1999. U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *Caenorhabditis elegans*. *RNA 5*:487–494.