Chemical Shift Mapping of the RNA-Binding Interface of the Multiple-RBD Protein Sex-Lethal[†]

Andrew L. Lee,[‡,§] Brian F. Volkman,[‖] Stephanie A. Robertson,[‡] David Z. Rudner,[⊥] Daniel A. Barbash,[⊥] Thomas W. Cline,[⊥] Roland Kanaar,[⊥,#] Donald C. Rio,[⊥] and David E. Wemmer*,[‡]

Departments of Chemistry and of Molecular and Cell Biology, University of California at Berkeley,
Berkeley, California 94720-3204, Department of Biochemistry and National Magnetic Resonance Facility at Madison,
University of Wisconsin−Madison, Madison, Wisconsin 53706, and Structural Biology Division, Lawrence Berkeley National
Laboratory, Berkeley, California 94720-5230

ABSTRACT: The *Drosophila* protein Sex-lethal (Sxl) contains two RNP consensus-type RNA-binding domains (RBDs) separated by a short linker sequence. Both domains are essential for high-affinity binding to the single-stranded polypyrimidine tract (PPT) within the regulated 3′ splice site of the *transformer* (*tra*) pre-mRNA. In this paper, the effect of RNA binding to a protein fragment containing both RBDs from Sxl (Sxl-RBD1+2) has been characterized by heteronuclear NMR. Nearly complete (85−90%) backbone resonance assignments have been obtained for unbound and RNA-bound states of Sxl-RBD1+2. A comparison of amide ^{1}H and ^{15}N chemical shifts between free and bound states has highlighted residues which respond to RNA binding. The β-sheets in both RBDs (RBD1 and RBD2) form an RNA interaction surface, as has been observed in other RBDs. A significant number of residues display different behavior when comparing RBD1 and RBD2. This argues for a model in which RBD1 and RBD2 of Sxl have different or nonanalogous points of interaction with the *tra* PPT. R^{142} (in RBD2) exhibits the largest chemical shift change upon RNA binding. The role of R^{142} in RNA binding was tested by measuring the K_d of a mutant of Sxl-RBD1+2 in which R^{142} was replaced by alanine. This mutant lost the ability to bind RNA, showing a correlation with the chemical shift difference data. The RNA-binding affinities of two other mutants, F^{146}A and T^{138}I, were also shown to correlate with the NMR observations.

Sex determination and X chromosome dosage compensation in the fruit fly, *Drosophila melanogaster*, are controlled by Sex-lethal (Sxl),[1] an RNA-binding protein belonging to the RNP consensus family [reviewed in Cline and Meyer (1996)]. Functional Sxl is found only in females. The sex-specific expression of Sxl is established through a mechanism involving transcriptional activation (Keyes *et al.*, 1992) and is maintained through an autoregulatory feedback loop in which Sxl protein regulates the sex-specific alternative splicing of its own pre-mRNA (Bell *et al.*, 1991). Sxl also controls the alternative splicing of the feminizing switch gene, *transformer* (*tra*), in somatic sexual differentiation (Bashaw & Baker, 1995; Kelley *et al.*, 1995; Zhou *et al.*, 1995) and the switch gene male-specific-lethal #2 in somatic dosage compensation (Kelley *et al.*, 1995; Zhou *et al.*, 1995; Bashaw & Baker, 1995). Regulation of *tra* by Sxl has been studied extensively both genetically and biochemically (Inoue *et al.*, 1990; Valcárcel *et al.*, 1993; Sosnowski *et al.*, 1994). Sxl binds to the highly conserved (O'Neil & Belote, 1992) intron pyrimidine tract in the regulated intron of the *tra* pre-mRNA, blocking the use of this strong constitutive 3′ splice site, and a weaker downstream 3′ splice site is used instead (Figure 1A). This female-specific alternative splicing event bypasses a premature stop codon, and the resulting mRNA encodes the full-length transformer protein.

The deduced amino acid sequence of Sxl reveals the presence of two segments with strong similarity to the ribonucleoprotein (RNP) consensus sequence domain (Figure 1B) (Bell *et al.*, 1988). This 80−90 amino acid motif is found in over 150 proteins involved in RNA metabolism and is generally found to possess RNA-binding activity (Birney *et al.*, 1993; Burd & Dreyfuss, 1994a). We will refer to the

* To whom correspondence should be addressed.

[‡] Department of Chemistry, University of California at Berkeley, and Structural Biology Division, Lawrence Berkeley National Laboratory.

[§] Present address: Department of Chemistry, Natural Sciences and Mathematics Complex, SUNY at Buffalo, Buffalo, NY 14260-3000.

[‖] Department of Biochemistry and National Magnetic Resonance Facility at Madison, University of Wisconsin−Madison.

[⊥] Department of Molecular and Cell Biology, University of California at Berkeley.

[#] Present address: Department of Cell Biology and Genetics, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1997.

[1] Abbreviations: Sxl, Sex-lethal; RBD, RNA-binding domain; Sxl-RBD1, N-terminal RBD construct from Sxl; Sxl-RBD2, C-terminal RBD construct from Sxl; Sxl-RBD1+2, construct from Sxl containing both RBDs; RBD1, N-terminal RBD within Sxl-RBD1+2; RBD2, C-terminal RBD within Sxl-RBD1+2; α3$_I$, third α-helix in RBD1; β4$_{II}$, fourth strand of β-sheet in RBD2; R^{142}, arginine 142 in the Sxl-RBD1+2 sequence; NMR, nuclear magnetic resonance; *tra, transformer* gene; PPT, polypyrimidine tract; HSQC, heteronuclear single-quantum coherence; NOE, nuclear Overhauser effect; snRNP, small nuclear ribonucleoprotein; hnRNP, heterogeneous nuclear ribonucleoprotein.
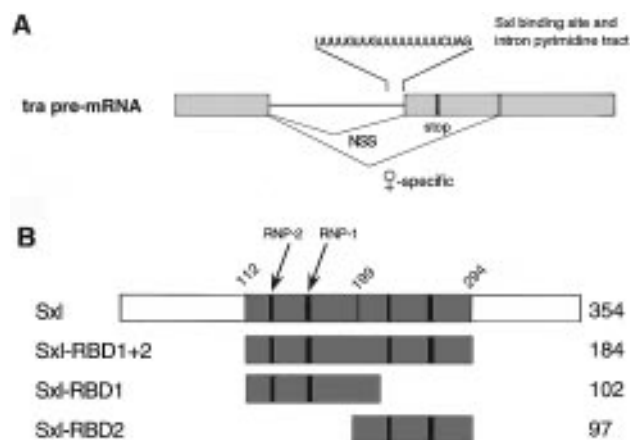
FIGURE 1: (A) Schematic diagram of the sex-specific alternative splicing of the *transformer* (*tra*) pre-mRNA. Shaded boxes represent exons, and lines represent introns. The non-sex-specific (NSS) splice that occurs in both males and females is indicated, and the female-specific splice is also indicated. The premature stop codon that is bypassed in the female-specific splice is indicated. The Sxl-binding site which is also the intron pyrimidine tract is expanded to show the sequence. (B) Sxl protein fragments discussed in the text are shown. The highly conserved RNP-1 octamer and RNP-2 hexamer are shown in black. The number of amino acids in each fragment is shown on the right.

two RNP consensus sequence domains of Sxl as RNA-binding domains (RBDs) since RNA-binding activity has been shown to reside in these domains (Kanaar *et al.*, 1995). The RNA-binding activity of the full-length protein is retained by a protein fragment derived from Sxl that encompasses both RBDs (Sxl-RBD1+2, Figure 1B). It has approximately a 1000-fold higher affinity for the wild-type (wt) Tra-PPT than for a mutant version of this PPT in which three of the uridines in a stretch of eight have been replaced by cytosines (Kanaar *et al.*, 1995). Both RBDs of Sxl are required *in cis* for site-specific RNA binding. The individual RBD protein fragments interact (more weakly) with the RNA but cannot discriminate between wt and mutant Tra-PPTs.

Previously, we determined the solution structure of the second RBD of Sxl (Sxl-RBD2) to intermediate resolution (Lee *et al.*, 1994). The structure of Sxl-RBD2 is similar to the two other RBDs for which three-dimensional structures have been determined, those of U1A snRNP and hnRNP-C (Nagai *et al.*, 1990; Wittekind *et al.*, 1992). These RBDs all have the same global fold: two α-helices packed against a four-stranded antiparallel β-sheet. Two highly conserved amino acid sequences, the RNP-1 octamer and the RNP-2 hexamer (Figure 1B), form the central two strands of the β-sheet. Despite the conserved overall fold, there are a number of differences between the structurally characterized RBDs. The most striking difference is the length of the loop between β-strands 2 and 3 (the β2−β3 loop), which varies from 4 amino acids in hnRNP-C to 10 in Sxl-RBD2. The differences in amino acid sequence and structural subtleties between various RBDs are likely to contribute to the diversity in RNA sites bound by RNP-type RBDs.

A high proportion of RBD-containing proteins have more than one RBD, and in many cases these multiple RBDs are required for proper function. In order to gain a more complete understanding of RBD function, it will be important to carry out biophysical studies on these multiple RBD proteins. Sxl is a simple example of a multiple RBD-containing protein since it contains only two RBDs connected by a short linker sequence (Figure 1B). Structural investiga-

tion of the two RBDs of Sxl (Sxl-RBD1+2) should help to identify important features of multiple-RBD recognition of RNA.

In this paper, we have analyzed the Sxl-RBD1+2/Tra-PPT interaction by chemical shift perturbation mapping of the protein backbone. These data suggest that the two domains of Sxl-RBD1+2 protein (RBD1 and RBD2) respond quite differently to the presence of Tra-PPT RNA. In complementary experiments, three mutants of Sxl-RBD1+2 were constructed and assayed for changes in RNA-binding affinity versus wild-type. One mutant, R[142]A, eliminated RNA-binding activity in Sxl-RBD1+2. Functional aspects of the Sxl/RNA interaction are discussed in terms of these results.

MATERIALS AND METHODS

Sample Preparations. Uniformly ^{15}N- and ^{13}C-, ^{15}N-labeled Sxl-RBD1+2 was prepared as described (Kanaar *et al.*, 1995), with the exception that *E. coli* were grown on M9 minimal media containing ^{15}NH$_4$Cl as the sole nitrogen source. In the case of ^{13}C-, ^{15}N-labeled Sxl-RBD1+2, [U-^{13}C$_6$]-D-glucose was used as the sole carbon source. 5′-GUUUUUUUC RNA was synthesized on an ABI 391 DNA/RNA synthesizer. Deprotection of the 2′-OH-*tert*-butyldimethylsilyl protecting group was carried out with triethylamine hydrofluoride, and the deprotected RNA was purified on a Dionex NucleoPac PA-100 (9 × 250) HPLC anion exchange column. RNA in a stock solution of 5 mM (determined by measurement of optical density) was titrated into 0.8 mM RBD1+2 until the RNA was in slight excess. The final solution composition of the RBD1+2/RNA complex samples used for NMR was 90% H$_2$O/10% D$_2$O solution, 30 mM sodium phosphate buffer, pH 6.3, 30 mM NaCl, and 0.02% NaN$_3$. Due to gradual RNA degradation, fresh RNA was added before each 3D experiment to maintain the 1:1 protein/RNA complex. The complex concentration thus varied from 0.7 to 0.4 mM.

NMR. All NMR experiments were performed at 600.13 or 750.13 MHz on Bruker DMX spectrometers at 20 °C. Both spectrometers were equipped with Bruker ^{1}H, ^{13}C, ^{15}N triple-resonance probeheads with three-axis self-shielded gradient coils. All NMR experiments (Table 1) were acquired with pulse schemes which are based on previously described methods (Grzesiek & Bax, 1993; Kay *et al.*, 1994) and included the following enhancements: magic-angle (instead of *Z*-axis) gradients for coherence selection and sensitivity enhancement, which resulted in improved suppression of the water solvent signal (Zijl *et al.*, 1995); minimal saturation of the water signal in all experiments except HCACO and CBCA(CO)NH; continuous ^{1}H decoupling for the optimization of triple-resonance experiments (Grzesiek & Bax, 1992). All pulse sequences used (Bruker DMX format) may be obtained from the BMRB pulse sequence library (www.bmrb.wisc.edu) or the NMRFAM WWW server (www.nmrfam.wisc.edu/~volkman). Table 1 contains a list of experiments which were performed on the Sxl-RBD1+2/Tra-PPT 10-mer complex, including the spectral parameters. ^{15}N T_2 relaxation measurements were performed with ^{15}N transverse relaxation delays of 0, 8, 16, 24, 32, 40, 48, 64, and 128 ms. The published CPMG sequence (Farrow *et al.*, 1994) was altered such that the minimal relaxation period was reduced by a factor of 2: ^{1}H 180° pulses used for the suppression of cross-correlation
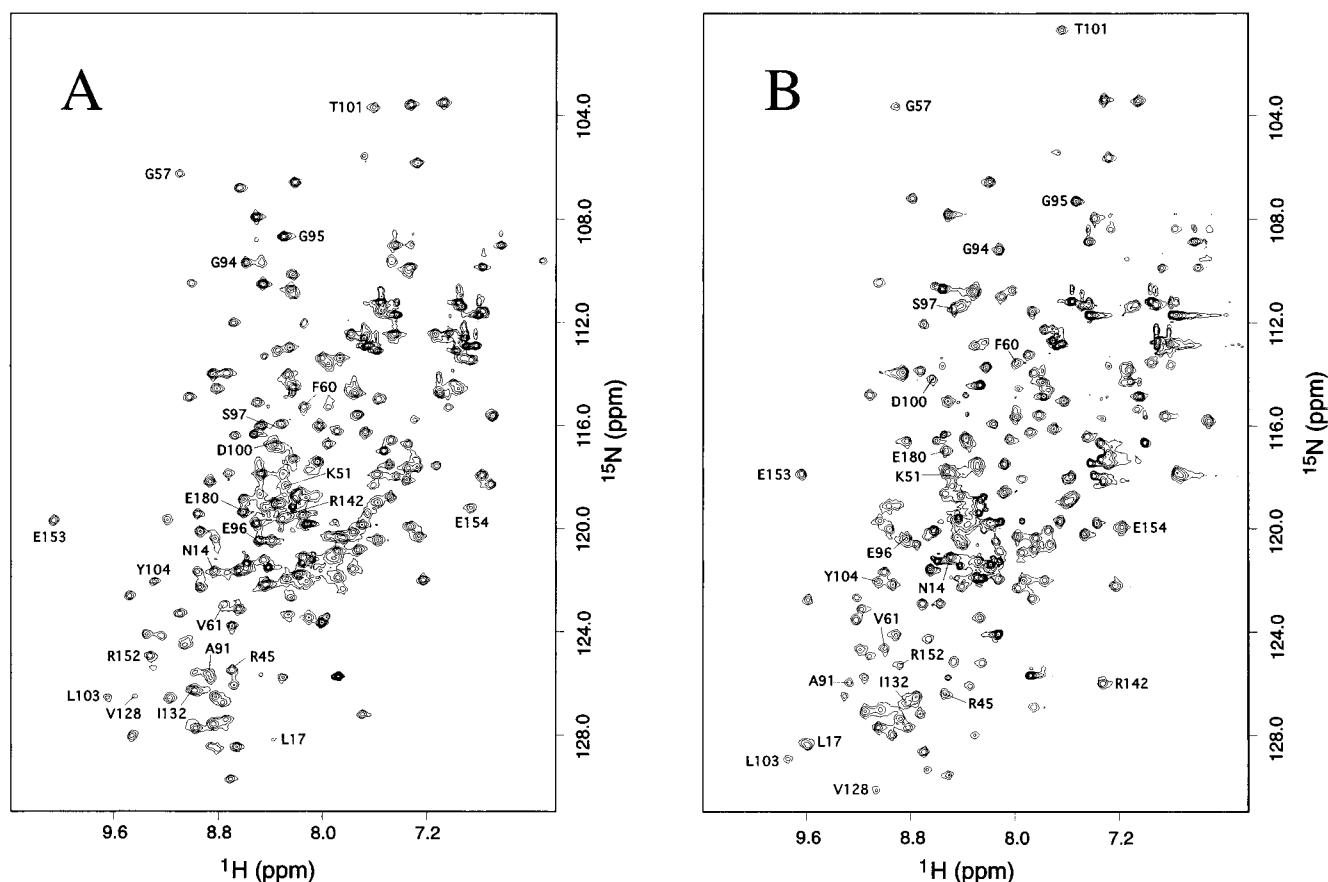
FIGURE 2: Comparison of 750 MHz 2D ^{15}N/^{1}H HSQC spectra of Sxl-RBD1+2 in the presence (A) and absence (B) of the Tra-PPT 10-mer RNA 5′-GUUUUUUUUC from *transformer* pre-mRNA. Spectra were collected at 20 °C for 2 h each. Some peaks which undergo notable resonance shifts upon binding of RNA are indicated.

effects were placed after every four (instead of eight) ^{15}N 180° pulses. ^{15}N T_1 measurements were obtained similarly. All spectra were processed using Felix 2.30 or 95.0β (Biosym Technologies, 1993), and chemical shifts were externally referenced to DSS (Wishart *et al.*, 1995a). Relaxation measurements were analyzed with the suite of macros and programs made available from the laboratory of Prof. Art Palmer, also available on their WWW site (http://cuhhca.h-hmi.columbia.edu/palmer/palmer_group.html).

Construction and Purification of Sxl Mutant Proteins. The cDNA fragment encoding wild-type (wt) Sxl-RBD1+2 (Kanaar *et al.*, 1995) was subcloned into the expression vector pRSETA (Invitrogen) between *Nde*I and *Bam*HI restriction sites. Site-directed mutagenesis was performed in this plasmid as described (Maniatis *et al.*, 1989). The oligos used were T138I, 5′-CACGAGGACGACCGAT-GAGCTTGTCACGC; F146A, 5′-CTTGTTGTACCGAA-CAGCGGCCACACCACG; and R142A, 5′-CAAAGGC-CACACCAGCAGGACGACCTGTG. The wt and three mutant proteins were overexpressed in *E. coli* strain BL21: DE3 pLys-S. Expression and purification were as described (Kanaar *et al.*, 1995) except that the ammonium sulfate precipitates from the pooled blue trisacryl column fractions were dialyzed against buffer H [20 mM HEPES−NaOH (pH 7.6), 1 mM EDTA, 1 mM DTT, 10% glycerol] and loaded on a porosHS column that was equilibrated with buffer H containing 25 mM NaCl. The protein was eluted with a 15 column volume linear gradient from 25 mM to 300 mM NaCl. Peak fractions were aliquoted, frozen in liquid N$_2$, and stored at −85 °C.

Protein−RNA Binding Analysis. Equilibrium dissociation constants (K_d) for interaction of different Sxl-derived proteins

with the *transformer* polypyrimidine tract (Tra-PPT) were determined with the use of native gel electrophoresis as described previously (Kanaar *et al.*, 1995). Binding reactions were performed in a volume of 20 μL and contained the indicated concentrations of proteins (below), 5 pM oligo-nucleotide, 15 mM HEPES−KOH (pH 7.6), 50 mM KCl, 1 mM EDTA, 1 mM DTT, 50 mg/ml BSA, 0.005% Nonidet P40, and 2% glycerol. Protein concentrations of 0.01, 0.03, 0.1, 0.3, 1.0, and 3.0 nM were used for wt Sxl-RBD1+2 and T^{138}I. Concentrations of 0.03, 0.1, 0.3, 1.0, 3.0, and 10.0 nM were used for F^{146}A, and concentrations of 0.1, 1.0, 10.0, 100, and 1000 nM were used for R^{142}A. Incubations were continued for 2 h at 4 °C. One-quarter of the reaction mixture was electrophoresed through a 4% polyacrylamide gel (60:1; 0.5× TBE) at 4 °C for 100 min at 20 V/cm. RNA binding was quantitated with a Fuji Phosphorimager. K_ds were obtained by fitting binding curves to the data obtained from the native gels. A simple two-state binding reaction was assumed, and the data were fit to the following equation: $y = 1/[1 + (K_d/x)]$, in which y represents the fraction of RNA bound by protein and x equals the protein concentration.

RESULTS

^{15}N/^{1}H *HSQC Spectra.* Figure 2A and Figure 2B show ^{15}N/^{1}H HSQC spectra of uniformly ^{15}N-labeled Sxl-RBD1+2 (21 kDa) in the free (unbound) state and in a complex with the Tra-PPT 10-mer RNA (5′-GUUUUUUUUC-3′, from the regulated *transformer* polypyrimidine tract), respectively. Each spectrum was recorded at 750 MHz. The complex has a K_d of ~5 × 10^{-12} M, and exhibits slow-exchange behavior on the NMR time scale (Kanaar *et al.*, 1995). Lines of the

Table 1: Spectral Parameters for NMR Data Sets Collected on Sex-Lethal/Tra-PPT (10-mer) Complex

experiment	^1H[a]		d2			d3			matrix size[b]	completeness[c]
	SF (MHz)	SW (Hz)	nucleus	SW (Hz)	N^*	nucleus	SW (Hz)	N^*	(d1 × d2 × d3)	(%)
HNCA	750.13	11062.00	$^{13}C_\alpha$	5000.00	48	^{15}N	2500.00	40	512 × 128 × 128	85[e]
HN(CO)CA	600.13	8333.34	$^{13}C_\alpha$	4000.00	48	^{15}N	2000.00	44	512 × 128 × 128	80
CBCA(CO)NH	600.13	8333.34	$^{13}C_{\alpha/\beta}$	8333.33	40	^{15}N	2000.00	32	512 × 128 × 128	15−20
HNCO	600.13	8333.34	$^{13}C'$	2000.00	60	^{15}N	2000.00	32	512 × 128 × 128	95
HCACO	600.13	6250.00	$^{13}C'$	2000.00	72	$^{13}C_\alpha$	4000.00	28	512 × 256 × 128	65−70
^{15}N-TOCSY	600.13	8333.34	^1H	5000.00	80	^{15}N	2000.00	36	512 × 256 × 128	45−50[d]
^{15}N-TOCSY	750.13	10000.00	^1H	8333.33	96	^{15}N	2272.73	35	512 × 256 × 128	45−50[d]
^{15}N-NOESY	750.13	8333.00	^1H	8333.33	120	^{15}N	2272.73	35	512 × 256 × 128	100
2D $^{15}N−^1H$ HSQC	750.13	10000.00	^{15}N	2500.00	200				512 × 512	100
$^{15}N\ T_1/T_2$	750.13	10000.00	^{15}N	2500.00	150				512 × 512	

[a] A total of 1024 complex points were collected in the ^1H dimension (d1) of all experiments. In all experiments except HCACO, the right half of the spectrum (upfield of H_2O) in the ^1H dimension contained no signals and was discarded. [b] Data were zero-filled to the final sizes indicated. Linear prediction was used to extend data by 50% in the d3 dimension only. [c] Completeness indicates the percentage of residues for which expected signals in triple-resonance experiments were observed. [d] Indicates completeness for combined total of both 3D ^{15}N-separated TOCSY experiments. [e] Percentage for HNCA includes intraresidue correlations only.

free protein disappeared at a 1:1 stoichiometry as determined from the stock concentrations. The expected number of NH correlation peaks was observed in the HSQC of the complex and is consistent with a single form of the protein in the bound state. Both spectra have regions of overlap due to the number of peaks and the relatively broad lines. The ^{1}-HN line widths of free RBD1+2 are approximately 35 Hz. In the case of RBD1+2 bound to RNA, the ^{1}HN line widths range from 30 to 50 Hz, with an average line width of 35−40 Hz. These broad lines were a major point of concern for obtaining sequence-specific assignments for both free and RNA-bound forms of RBD1+2, although data collection and the subsequent analysis were more problematic for the latter. Therefore, the assignment process for the RNA-bound form of RBD1+2 will be described, even though a similar approach was used for free RBD1+2.

Backbone Resonance Assignments. In order to obtain backbone resonance assignments for the protein, a three-dimensional triple-resonance approach was taken. The initial strategy was to use CBCA(CO)NH and HNCACB experiments, as this combination works quite effectively. Unfortunately, these experiments did not yield sufficiently complete data sets due to the signal loss resulting from broad lines and the relatively high number of magnetization transfer steps (Table 1). Only experiments employing a minimal number of transfer steps and large *J*-coupling constants yielded useful data sets; HNCO, HNCA, HN(CO)CA, and HCACO are experiments which fulfilled these requirements. The final column of Table 1 summarizes the percentage of expected peaks which were actually observed in each data set acquired on the complex. It should be noted that experiments were limited to durations of approximately 48 h on the ^{13}C-, ^{15}N-labeled sample due to RNA degradation. As expected, the HNCO was the most sensitive experiment with 95% of the possible correlations observed, followed by the HNCA (85% of the intraresidue correlations) and HN-(CO)CA (80%) experiments. The HCACO was approximately 65% complete. $^1H_N−^1H_\alpha$ correlations were obtained from 3D ^{15}N-separated TOCSY-HSQC. For free RBD1+2, the corresponding data sets were more complete.

The assignment strategy employed was centered around the sequence-specific identification of backbone amide ^1H and ^{15}N resonances. This was accomplished when enough sequence-adjacent $^1H/^{15}N$ pairs were linked so that $^{13}C_\alpha$ chemical shifts, $^{13}C_\beta$ chemical shifts [when present in CBCA-
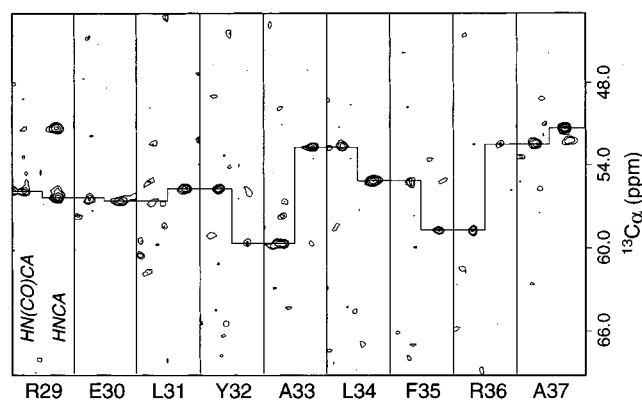

FIGURE 3: Strip plot of sequential links from 3D HN(CO)CA and HNCA spectra of the RBD1+2/Tra-PPT complex. Pairs of $^1H−N$ strips for residues 29−37 are arranged with the interresidue $C_{\alpha(i−1)}$ correlation [HN(CO)CA] on the left and intraresidue $C_{\alpha(i)}$ correlation (HNCA) on the right. Horizontal lines link an intraresidue correlation to the interresidue correlation for the following residue. The noise level illustrates the low sensitivity of the data for the complex, which is the reason for difficulty in obtaining complete sequential assignments from them.

(CO)NH], and NOE patterns would unambiguously place the stretch into the amino acid sequence (Wüthrich, 1986; Wishart *et al.*, 1991). Links between residues were accomplished using three methods:

(1) When signals from all basic triple-resonance experiments [HNCO, HNCA, HN(CO)CA, and HCACO] and a $^1H_N/^1H_\alpha$ correlation from either the ^{15}N-separated TOCSY-HSQC or the ^{15}N-separated NOESY-HSQC were present, the HCACO experiment could be used to link sequential $^1H/^{15}N$ pairs. This was the most desirable method for linking sequence-adjacent $^1H/^{15}N$ pairs because all correlations are extended by the matching of two independent chemical shifts (Ikura *et al.*, 1990).

(2) Candidates for sequential $^1H/^{15}N$ pairs were generated by matching $^{13}C_\alpha$ chemical shifts from HNCA and HN(CO)-CA, and these candidates were evaluated by analysis of NOE patterns from the ^{15}N-separated NOESY-HSQC. This method was used when correlations from the HCACO or ^{15}N-separated TOCSY-HSQC were not identified, and was generally successful since the HNCA and HN(CO)CA experiments were reasonably complete (Table 1). The HNCA/HN(CO)CA approach is illustrated in Figure 3 with data for residues R[29]−A[37] of the complex.
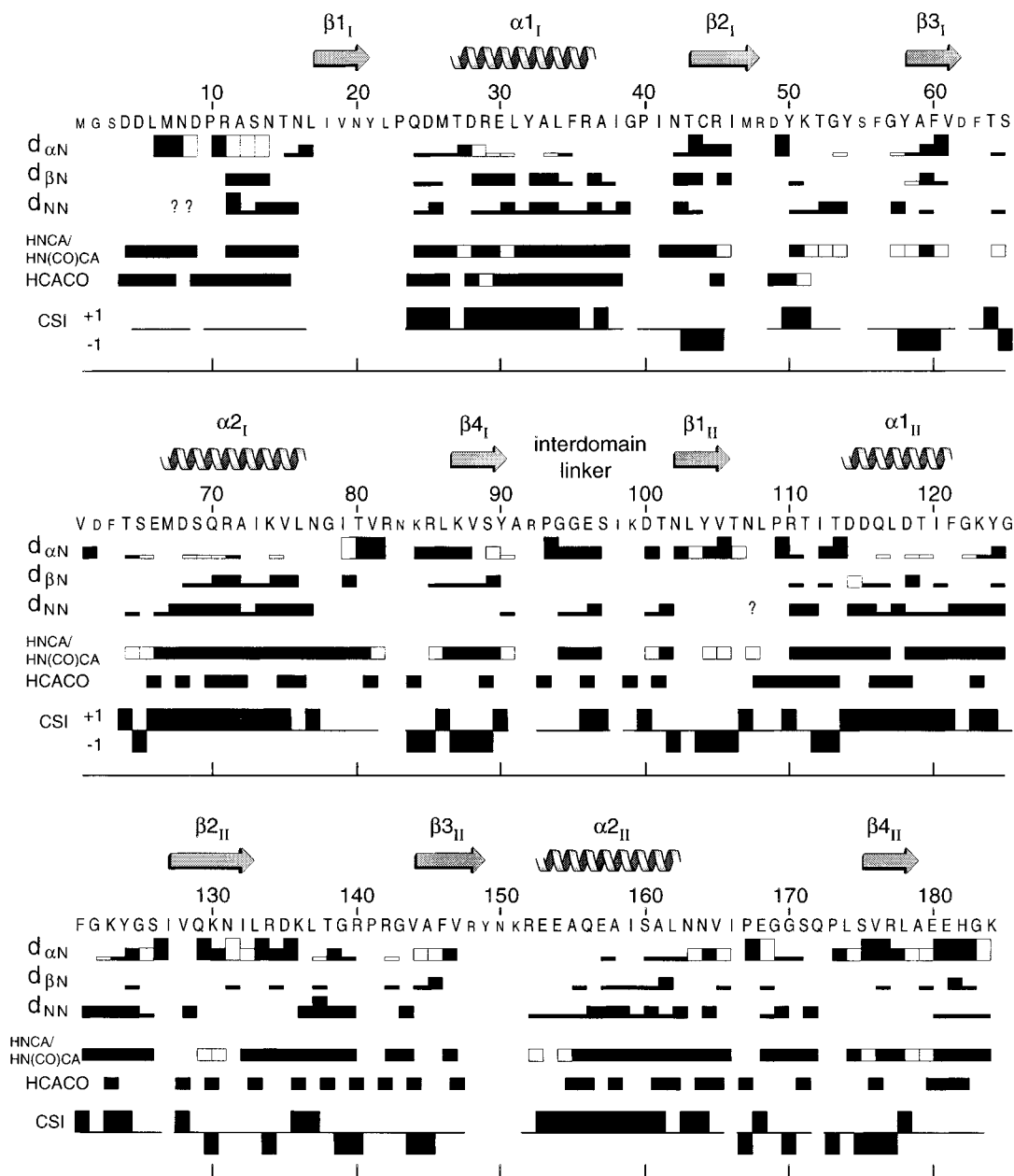
FIGURE 4: Summary of sequential NOEs, chemical shift index (CSI), and through-bond connectivities identified in the process of obtaining sequence-specific backbone resonance assignments of the Sxl-RBD1+2/Tra-PPT complex. The complete sequence is shown in the one-letter amino acid code. Amino acids as smaller letters indicate residues for which no backbone NH resonance was assigned. The location of secondary structure elements is indicated above the sequence. The relative intensity of sequential NOEs is indicated by the height of the connecting box. Unfilled NOE boxes indicate NOEs which are partially overlapped or for which $^1H_\alpha$ or $^1H_\beta$ assignments have been inferred from NOESY data, due to missing HCACO or 3D ^{15}N-TOCSY peaks. Because of the general lack of side chain resonances observed in the 3D ^{15}N-TOCSY, $d_{\beta N}$ NOEs indicate sequential side chain to NH cross-peaks for which assignments are inferred from amino acid type and based upon the presence of both intra- and interresidue NOEs. Question marks denote degeneracy in ^{1}HN chemical shifts. Two types of triple-resonance connectivities are indicated. A box in the HNCA/HN(CO)CA row indicates that a sequential connection was observed by correlating $^{13}C_\alpha$ shifts in HNCA and HN(CO)CA spectra. An unfilled HNCA/HN(CO)CA box indicates a very weak cross-peak in either experiment. The HCACO row indicates HCACO peaks which were assignable via correlations to HNCO, HNCA, HN(CO)CA, and 3D ^{15}N-TOCSY peaks. An unfilled HCACO box indicates that the HCACO peak occurs in a very crowded region of the spectrum. The row labeled CSI contains consensus chemical shift index values for backbone resonances of the Sxl-RBD1+2/Tra-PPT complex. The index is based on the agreement of two out of three shifts from random coil values for $^{13}C_\alpha$, $^{13}C'$, and $^1H_\alpha$ where all three values were assigned, or agreement of both for residues where one of the three assignments was missing.

(3) When the triple-resonance spectra were too incomplete to use, connectivities could in some cases be made through NOESY/TOCSY analysis (Wüthrich, 1986) or from NOEs exclusively. This was most effective when sequential HN−HN NOEs were observed, since connectivities could be confirmed by the presence of a cross-peak on both sides of the diagonal.

Approximately 87% of the backbone ^{1}H, ^{13}C, and ^{15}N resonances in the complex and 90% in free RBD1+2 were

assigned sequence-specifically. Figure 4 summarizes the *J*-correlated connectivities used for the assignments as well as sequential NOEs indicative of secondary structure in the complex. Squares in the row labeled "HNCA/HN(CO)CA" indicate links made via method 2. Squares in the row labeled "HCACO" indicate HCACO peaks which were assigned via correlations to HNCO, HNCA, and HN(CO)CA peaks. An equivalent figure for the free RBD1+2 protein is given in supporting information (see Supporting Information Avail-

able) (Figure S1). The chemical shifts for both forms have been deposited in the BioMagResBank (Accession numbers bmr4029.str for the free protein and bmr4028.str for the complex).

Secondary Structure. Even though prior knowledge of the secondary and tertiary structure for the second domain (Sxl-RBD2) was available (Lee *et al.*, 1994), placement of secondary structure in RBD1+2 was inferred from chemical shifts (Wishart & Sykes, 1994) and NOE patterns. The ability to identify secondary structure from patterns of differences between random coil and observed chemical shifts has been documented (Wishart *et al.*, 1995b). These differences for $^{13}C_\alpha$, $^{13}C'$, and $^1H_\alpha$ were used to calculate a consensus chemical shift index (CSI) which is shown in Figure 4. Since $^{13}C_\beta$ shifts were not generally observed in the CBCA(CO)NH, this consensus index is not strictly the same as that presented originally (Wishart & Sykes, 1994). The index in Figure 4 interprets agreement of two of the three individual indices as a consensus, provided that the third index had a value of 0. Helices and β-strands are indicated by continuous consensus values of $+1$ and -1, respectively. The β-strands are not delineated as clearly as helices, which may be due to the lack of $^{13}C_\beta$ resonance assignments.

The $\beta\alpha\beta-\beta\alpha\beta$ topology found in all structurally characterized RBDs is observed in each of the domains of RBD1+2 in both the free and bound states. There are no apparent changes in secondary structure which accompany RNA binding. The secondary structure of RBD2 is entirely consistent with that previously observed in Sxl-RBD2 (Lee *et al.*, 1994). Chemical shift and NOE patterns indicate that the interdomain linker contains no regular secondary structure in either free or bound RBD1+2. In the case of RBD1, the secondary structure was not previously mapped, and is now found to overlay reasonably well with the secondary structure of other RBDs. However, careful inspection of Figures 4 and 5 reveals a small but significant degree of variability in the spacing between elements of secondary structure in RBD1 relative to RBD2. For example, the $\alpha2-\beta4$ loops from each domain are of slightly different lengths and contain different sequence compositions. These differences will be examined in greater detail below.

Estimation of the Overall Tumbling Correlation Time. In an attempt to obtain more quantitative information on line widths, ^{15}N longitudinal and transverse relaxation times were measured using two-dimensional inverse-detected methods. The uncertainties in these data were too large for the reliable calculation of order parameters, but an estimation of the overall tumbling correlation time for the complex was possible. At 76 MHz, an average T_1 of 0.61 s and a T_2 of 36 ms were calculated from residues involved in secondary structure. The T_1/T_2 ratio was then used to estimate an apparent tumbling correlation time for the complex of approximately 17 ns using TMEST (Dr. Arthur Palmer; see Materials and Methods). A calibration curve was generated using τ_c values determined by NMR for 12 small proteins (supporting information, Figure S2) which predicted a correlation time of 15 ns for this molecular weight. The observed 17 ns τ_c is the same within error, while a dimeric complex, predicted to have a correlation time of 30 ns, can be ruled out. For free RBD1+2, an estimation of the tumbling correlation time was not attempted because the two domains display different relaxation properties, and are therefore likely to tumble with significant but unpredictable

degrees of anisotropy. In free RBD1+2 at 76 MHz, the average ^{15}N T_2 values for residues involved in secondary structure in RBD1 and RBD2 were calculated to be 38 and 45 ms, respectively.

Changes in Chemical Shifts upon RNA Binding. Figure 5 shows the absolute values of chemical shift differences between free and RNA-bound RBD1+2 for amide ^1H and ^{15}N resonances (shown in Hertz from data with a proton frequency of 600 MHz). These differences will be referred to as chemical shift perturbations. For residues which were assigned in one form but not in the other, it was possible to calculate minimum shift perturbations. For example, Y^{21} was assigned in free RBD1+2 but not in the complex. The unassigned HSQC peak in the complex closest to the chemical shift values for Y^{21} in the free state would therefore become the most conservative assignment for Y^{21} in the bound state, yielding a minimum shift perturbation. Since nearly all residues were assigned in both forms and approximately 175 amide NH correlation peaks (183 residues, 8 prolines) were observed in the HSQC spectra of both forms, it was clear that lines were shifting rather than disappearing due to broadening. The HNCO spectrum was checked to show that peaks had not simply moved under another neighboring peak. This made assignment of minimum shifts straightforward. Gray bars with up-arrows in Figure 5 correspond to the sum of these ^1H and ^{15}N minimum shift perturbations. Substantial chemical shift perturbations (>75 Hz) for a given residue can be interpreted in two ways: (1) a change in local protein conformation alters the magnetic shielding; and (2) the residue is involved in direct or protein-mediated contacts with the RNA. Differentiating between these two scenarios at the present stage of the analysis is difficult because chemical shifts are not yet rigorously understood in the context of protein structure. Additionally, RNA−protein interactions are likely to be predominantly mediated through the side chains of the protein residues, and lack of chemical shift changes for the backbone cannot necessarily be interpreted as lack of interaction in that region. These data should therefore be interpreted in a qualitative manner.

As shown in Figure 5, there is generally more extensive shifting of resonances in the vicinity of β-strands than in the helices, although a few helices have substantial perturbations near N-termini and some β-strands have small perturbations. On the whole, this is consistent with previous observations that the antiparallel β-sheet of an RBD serves as an RNA-binding surface (Görlach *et al.*, 1992; Oubridge *et al.*, 1994; Kanaar *et al.*, 1995). This is illustrated in Figure 6 which provides a structural representation of the chemical shift perturbations for RBD2, visualized on the backbone of the three-dimensional solution structure of the single domain Sxl-RBD2 (Lee *et al.*, 1994). For further reference, all elements of secondary structure and loops in RBD1+2 are labeled in Figure 7. The highly conserved RNP-1 and RNP-2 consensus sequences found in the central β-strands of all RBDs are affected in Sxl-RBD1+2 by the binding of RNA. Residues in the interdomain linker also display significant shift perturbations upon RNA binding. This suggests that the entire linker experiences a substantial change in chemical environment, even though it lacks regular secondary structure in either state (Figure 4). $\beta1_{II}$, which contains much of the RNP-2$_{II}$ hexamer (L^{103}YVTNL^{108}), immediately follows the linker, and the shift perturbations for the residues bridging the linker and $\beta1_{II}$ (E^{96}SIKDTN^{102}) are large throughout.
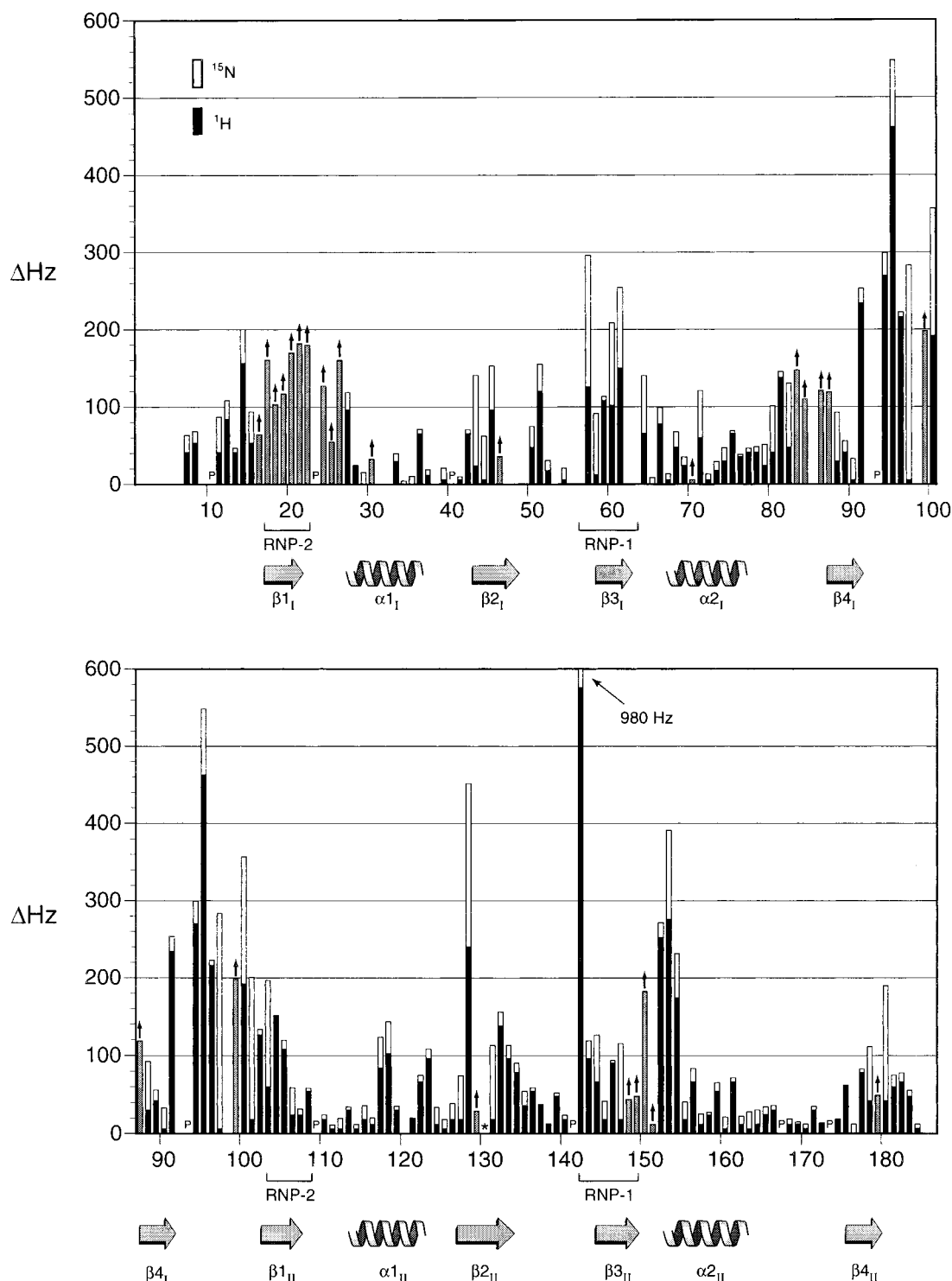
FIGURE 5: Plot of chemical shift perturbations (defined as the absolute value of the chemical shift difference) due to the Tra-PPT 10-mer binding to Sxl-RBD1+2. The sequences of RBD1 and RBD2 are aligned vertically in the two plots to allow comparison of the analogous locations within the two domains (part of the sequence is repeated). Chemical shift perturbations of backbone amide ^{1}H (black) and ^{15}N (white) are stacked to show the total shift observed in hertz, at 600 MHz proton frequency. Gray bars with arrows are calculated minimum shift perturbations (see text for a more detailed description). The location of secondary structural elements is indicated and labeled below each plot as are the positions of the conserved RNP1 and RNP2 sequences. Proline residues are indicated with a "P".

Unlike most regions with large shift perturbations in RBD1+2, there is no interspersion of small shift differences in this stretch. This region will be discussed in more detail below.

Aside from the expected chemical shift perturbations in β-sheet residues, there are a number of other regions in the sequence which are perturbed by RNA binding. R^{142} has the largest chemical shift perturbation in the protein, with a ^{1}HN shift of 1 ppm and a ^{15}N shift of 7 ppm. R^{142} is the

first residue in the RNP-1$_{II}$ octamer consensus sequence, but lies just outside of the β-sheet. The extreme chemical shift perturbations observed for R^{142} combined with its positioning in the sequence and positive charge implicate it as a critical residue in RBD1+2 with respect to RNA binding. Another region in RBD2 which has large chemical shift perturbations is $\beta2_{II}$. I^{127} and V^{128} form a β-bulge in $\beta2_{II}$, which is on one edge of the RBD2 β-sheet (Lee *et al.*, 1994). The perturbations of the following two residues in $\beta2_{II}$ are not
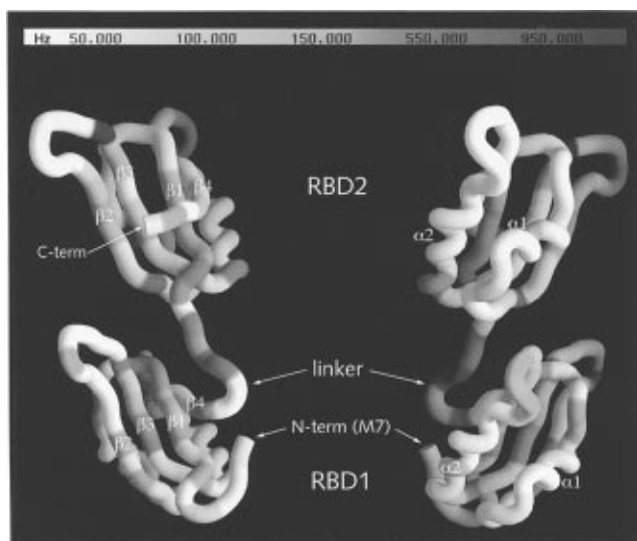
FIGURE 6: Structural representation of RNA-induced chemical shift perturbations. The solution structure of Sxl-RBD2 was used to depict each domain (Lee *et al.*, 1994) and is highlighted with a color gradient to indicate regions of significant chemical shift perturbations in RBD1+2 upon binding of the Tra-PPT 10-mer RNA. The backbone worm representation was created using the program GRASP (Nicholls, 1996). Two views of the same structure are shown, related by a rotation of 180° about the vertical axis. Residues of RBD1+2 for which the combined amide shift perturbation was less than 50 Hz (Figure 5) are colored white. For residues with shift differences greater than 50 Hz, a gradient of color was applied (shown at the top of the figure), with darker purple and blue shading representing increasing chemical shift perturbation. The locations of individual secondary structural elements are labeled. The residues of the linker are modeled in an extended conformation in order to display the significant shift perturbations which are observed in RBD1+2; the conformation of this region and the relative positions of the two domains are NOT actually determined from the NMR data.
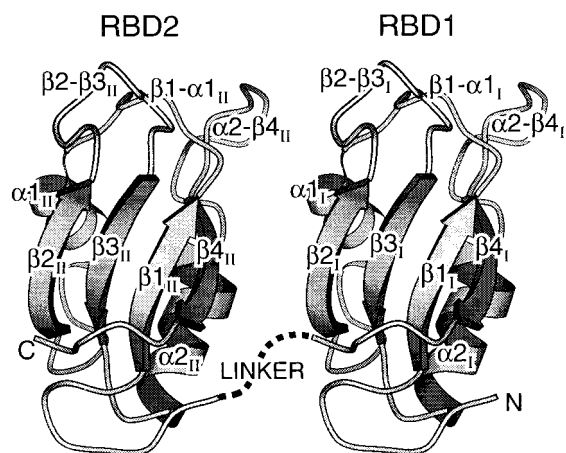

FIGURE 7: Schematic drawing of the Sxl-RBD1+2 protein is shown (based upon the NMR structure of RBD2 alone). The β-strands and loops connecting the helices and strands are labeled for reference to the text. The interdomain linker is shown as a dashed line to emphasize that its structure has not been determined from these data. The relative orientation of the two domains is hypothetical; the figure is intended to clarify the identity and labeling of the loops.

well characterized, although they are likely to have large shift differences as well; using assignments from free Sxl-RBD2 instead of free RBD1+2 (in which Q^{129} and K^{130} were not assigned; RBD2 within free Sxl-RBD1+2 and Sxl-RBD2 have nearly identical assignments), summed ^{1}HN and ^{15}N shift perturbations of 55 and 250 Hz are obtained for Q^{129} and K^{130}. In addition, N^{131} has a summed $^{13}C_\alpha$ and $^{13}C'$
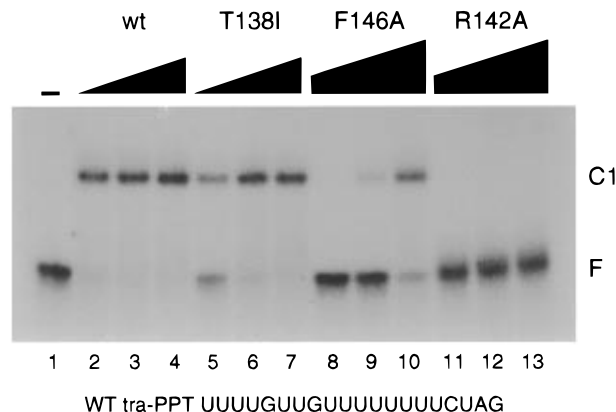

FIGURE 8: Electrophoretic mobility shift analysis of wild type (wt) and mutant Sxl-RBD1+2 with Tra-PPT. wt Sxl-RBD-1+2 and mutants threonine 138 to isoleucine (T138I), phenylalanine 146 to alanine (F146A), and arginine 142 to alanine (R142A) were incubated with 5 pM ^{32}P-labeled *tra* polypyrimidine tract (Tra-PPT) RNA oligonucleotide. The protein concentrations of wt (lanes 2−4) and T138I (lanes 5−7) were 0.01, 0.1, and 1 nM, respectively. The protein concentrations of F146A (lanes 8−10) and R142A (lanes 11−13) were 0.1, 1, and 10 nM, respectively. Protein RNA complex and unbound RNA (F) were separated by electrophoresis through a native polyacrylamide gel and visualized by autoradiography. The dissociation constants (K_d's) of wt, T138I, F145A, and R142A were determined to be ∼5 pM, 10 pM, 1 nM, and >1000 nM, respectively. The sequence of the Tra-PPT used is displayed below the autoradiogram.

chemical shift perturbation of 380 Hz. The last region of RBD2 which displays dramatic chemical shift perturbations is $N^{150}KREE^{154}$. There is a high density of charged hydrogen bond donors and acceptors in this region, which immediately follows the RNP-1$_{II}$ consensus sequence. This area also lies just outside the β-sheet.

RBD1 has fewer dramatic changes than RBD2 (Figure 5). In part this may be due to assignment gaps in the region of M47−F63, but it is clear that RBD1 and RBD2 display qualitative differences in their patterns of chemical shift perturbations. Two regions in RBD1 which show significant perturbations—with no counterparts in RBD2—are the $\beta1-\alpha1_I$ loop and the $\alpha2-\beta4_I$ loop. Interestingly, while the significantly perturbed $\alpha2-\beta4_I$ loop contains four positively charged residues ($G^{78}ITVRNKRLK^{87}$), the corresponding sequence in the loop of RBD2 ($\alpha2-\beta4_{II}$) is not basic at all ($N^{164}VIPEGGSQP^{173}$). Based on existing RBD structures, these positively charged residues would be situated toward the C-terminal end of the $\alpha2-\beta4$ loop, which should be near the RNA-binding surface.

Binding Affinity of Mutants of Sxl-RBD1+2. The most dramatic chemical shift perturbation in Sxl-RBD1+2 is the amide of R^{142}, a highly conserved arginine in the RNP-1$_{II}$ octamer (Birney *et al.*, 1993; Burd & Dreyfuss, 1994a). To test the importance of this residue in RNA binding, we mutated this arginine to alanine ($R^{142}A$). As a comparison, a second solvent-exposed, conserved residue in the RNP-1$_{II}$ octamer, F^{146}, was mutated to alanine ($F^{146}A$). These mutant proteins were analyzed for RNA binding to the *transformer* polypyrimidine tract (5′-UUUUGUUGUUUUUUUUCUAG) by gel retardation analysis, and these results are shown in Figure 8. Protein concentrations varied from 0.01 to 1000 nM in the binding assay (see Materials and Methods). We note that the Sxl-RBD1+2 (wt) K_d found here is approximately 10-fold lower than we have observed previously (Kanaar *et al.*, 1995). This discrepancy is probably due to the difference in the preparation of the labeled RNA and its

concentration determination, as well as the purity of the Sxl protein. This 10-fold difference has not affected interpretation of the data. Consistent with the dramatic chemical shift perturbation in R^{142}, the $R^{142}A$ mutation eliminated RNA binding (Figure 8, lanes 11−13). $R^{142}A$ appears to be properly folded since the HSQC spectrum of this mutant protein was very similar to wt (data not shown). All significant changes in ^1H and ^{15}N chemical shifts were restricted to residues neighboring A^{142}. The $F^{146}A$ mutant protein bound RNA with reduced affinity ($K_d \sim 1$ nM; wt $K_d \sim 5$ pM) (Figure 8, lanes 8−10) as might be expected for a mutation in a conserved residue in the RNP-1 octamer. The K_d of $F^{146}A$ is more similar to that of the Sxl-RBD1 fragment ($K_d \sim 48$ nM) (Kanaar *et al.*, 1995) than to the K_d of wt RBD1+2, consistent with the idea that this mutation reduces the RNA-binding contribution from RBD2.

We previously noted a moderate chemical shift perturbation of the amide of T^{51} in the context of the Sxl-RBD2 fragment bound to a DNA analog of the PPT (Kanaar *et al.*, 1995). This residue has also been found to be important for Sxl function *in vivo*. A mutation of this threonine to isoleucine has been identified in a temperature-sensitive, reduced-function (hypomorphic) allele of Sxl (R. A. Lersh and T.W.C., unpublished). To explore the possible involvement of this residue in RNA binding, we made the analogous mutation in Sxl-RBD1+2 ($T^{138}I$). The RNA-binding activity of this mutant protein was analyzed by native gel electrophoresis. The $T^{138}I$ mutation had a small but significant effect on RNA binding (Figure 8, compare lanes 2−4 with 5−7).

DISCUSSION

Despite widespread interest in understanding protein−RNA interactions, there are presently only a few published structures of protein−RNA complexes. Crystal structures have been determined for three transfer RNA synthetase complexes and for the MS2 bacteriophage capsid protein complexed with an RNA hairpin. While these are important structural examples, these proteins do not belong to any of the RNA-binding motifs typically involved in posttranscriptional gene regulation (Burd & Dreyfuss, 1994a). Only one protein from the RNP consensus-type family has been characterized in structural detail when complexed with RNA: the snRNP-U1A protein bound to U1 snRNA stem−loop II (Oubridge *et al.*, 1994), and in complex with a portion of the 3′UTR of U1A pre-mRNA (Allain *et al.*, 1996). These structures showed details of the specific interactions between an RNP-type RBD and RNA for the first time. However, both of these complexes contain only one RBD and therefore do not address the question of how multiple-RBD proteins interact with RNA. The results reported here on Sex-lethal give an initial glimpse at how more than one RBD can participate in high-affinity, specific binding of target RNA sequences.

Extent of Sxl-RBD1+2 Assignments. Sxl-RBD1+2 appears to retain arrangement of secondary structural elements when bound to a specific, single-stranded target RNA, the Tra-PPT (*transformer* polypyrimidine tract), based on chemical shift data and short- to medium-range NOEs. The assignment of approximately 85% of the backbone resonances of RBD1+2 in the free and RNA-bound states allowed for the mapping of the RNA-interaction surface of RBD1+2 as probed by amide ^1H and ^{15}N chemical shift

perturbations (Figures 5 and 6). Chemical shift perturbations for $^{13}C_\alpha$ and $^{13}C'$ resonances (supporting information, Figure S3) are consistent with the amide perturbation patterns shown in Figure 5.

The assignments of Sxl-RBD1+2 were obtained with some difficulty, particularly for the complex. Many of the experiments suffered from sensitivity problems due to the submillimolar concentrations used and the broad lines which resulted from a tumbling correlation time of about 17 ns for the 24 kDa complex. The extent of the assignments obtained here has reached the limit of what the most sensitive $^1H/^{13}C/^{15}N$ pulse sequences (see Materials and Methods) can achieve on a nondeuterated protein system.

Comparison of Chemical Shift Perturbations between RBDs of Sxl. A primary result from the chemical shift mapping is that the two RBDs of Sxl-RBD1+2 display different patterns of chemical shift perturbations upon RNA binding. Most notable are the $\alpha2-\beta4$ and $\beta1-\alpha1$ loops. These loops in RBD1 have larger chemical shift perturbations relative to those in RBD2, suggesting that they play an active role in RNA binding in RBD1 but not in RBD2. In the solution structure of Sxl-RBD2 (Lee *et al.*, 1994), and represented schematically in Figure 7, the $\alpha2-\beta4$ and $\beta1-\alpha1$ loops are not part of the putative RNA-binding platform provided by the β-sheet.

Significantly broader lines were observed in the triple-resonance data from the complex for loop $\alpha2-\beta4_I$ than the corresponding residues of RBD2 ($\alpha2-\beta4_{II}$). This is consistent with $\alpha2-\beta4_{II}$ having a substantial degree of dynamic disorder and $\alpha2-\beta4_I$ having more rigidity or undergoing motions on a slower time scale than $\alpha2-\beta4_{II}$. The correlation of chemical shift perturbation and structural mobility may reflect upon the relative involvment in RNA recognition for these loops. The high content of basic residues in loop $\alpha2-\beta4_I$ ($G^{78}ITVRNKRLK^{87}$) underscores its potential to form favorable electrostatic interactions with the negatively charged RNA. The $\alpha2-\beta4_{II}$ loop does not contain a corresponding stretch of basic residues, and the $\alpha2-\beta4$ loop has not been directly implicated in RNA binding in other RBD systems. The U1A snRNP $\alpha2-\beta4$ loop, while not basic, has been shown to be important for protein stability (Kranz *et al.*, 1996) and to undergo a structural shift upon RNA binding (Oubridge *et al.*, 1994). Kranz *et al.* also demonstrated that this loop is flexible on a slow time scale.

The $\beta1-\alpha1$ loops in RBD1+2 do not show any apparent differences in either sequence composition or line widths despite the large chemical shift pertubations seen in the $\beta1-\alpha1_I$ loop as compared to the $\beta1-\alpha1_{II}$ loop. In the U1A complex, residues in the $\beta1-\alpha1$ loop make contacts with the RNA near the base of the stem of the hairpin. If the $\beta1-\alpha1_I$ loop in the Sxl/Tra-PPT complex does interact with the RNA, these contacts are likely to be different than those seen in the U1A complex because there is no evidence for a stem−loop structure in the Tra-PPT.

Residues $N^{150}KREE^{154}$ in RBD2 show dramatic chemical shift perturbations, while the corresponding residues ($F^{63}-TSTMD^{71}$) in RBD1 do not. This is an interesting observation since (as in loop $\alpha2-\beta4_I$) this region of an RBD has not been implicated in RNA recognition. This continuous sequence of potentially hydrogen-bonding residues immediately follows the conserved RNP-1_{II} octamer, and it is therefore located at the edge of the RNA-binding platform (Lee *et al.*, 1994), proximal to the RNA. The last two or three residues of this stretch form the N-terminus of $\alpha2_{II}$.

This may also reflect a novel interaction since these helical residues have not been implicated in RNA recognition in other RBD/RNA complexes.

Another region of difference between the RBDs of Sxl is strand $\beta2$ (T[43]CRIMR[48] in RBD1 and I[127]VQKNIL[133] in RBD2). V[128], the β-bulging residue at the N-terminal section of the $\beta2_{II}$ strand, has a chemical shift perturbation at least 3 times as large as any residue in the corresponding section of the β-strand in RBD1. In the structure of free Sxl-RBD2, V[128] (V[41] in Sxl-RBD2 nomenclature) protrudes above the plane of the sheet surface from the corner of the sheet and is presented to the RNA which is predicted to lie across the sheet. As there is thus far no evidence for a similar β-bulge in $\beta2_I$, the presence of the bulge may correlate with chemical shift perturbations. The different behavior of these corresponding sections of β-sheet between the two domains is likely to reflect their unique interactions with the Tra-PPT. This reasoning can be extended to all regions which exhibit substantially different chemical shift perturbation patterns between RBD1 and RBD2. Therefore, the collective body of data presented in Figure 5 argues for a model of RNA recognition by Sxl-RBD1+2 in which the two RBDs take on different or structurally asymmetric roles. In addition to direct protein—RNA contacts, the difference in the chemical shift perturbation patterns may also reflect direct RBD1—RBD2 interactions or intradomain interactions that contribute to RNA-binding affinity. Evidence for homodimeric interactions between two Sxl molecules has recently been reported (Sakashita & Sakamoto, 1996); however, this interaction is RNA-mediated, and we found no evidence for formation of such a complex in previous gel shift studies using the same RNA as in the present work . The observations of one set of resonances and the determined correlation time are inconsistent with dimer formation.

Correlations between Chemical Shift Data and Mutant Binding Data. R[142] undergoes the largest chemical shift perturbation in the protein. The corresponding residue in U1A (R[52]) is involved in numerous critical hydrogen bonds with both RNA and protein (Oubridge *et al.*, 1994). In Sxl and U1A, this arginine occupies the first position in the conserved RNP1 octamer sequence and lies just outside the third strand of β-sheet, which appears to be a critical position for making RNA contacts. The implied importance of R[142] for RNA binding from chemical shift mapping was confirmed by the R[142]A mutant. This single point mutation within RBD2 completely eliminated complex formation. It was unexpected that the binding of the R[142]A mutant was completely lost rather than simply being lowered into the range of the affinity of Sxl-RBD1, which was the effect of the F[146]A mutant. This implies that there is a coupling of the two RBDs in binding the Tra-PPT, rather than simple additivity of their contributions. It is possible that R[142] makes essential RNA contacts in the context of the two RNA-binding domains without which binding is reduced considerably or abolished. Alternatively, RBD2 containing the R[142]A mutation might interfere with RBD1, inhibiting the RNA binding of both domains. This is possible since a comparison of the HSQC spectra from R[142]A and wt proteins revealed small but measurable chemical shift differences (20—30 Hz) for several residues in RBD1. The solution structure of the complex of RBD1+2 and RNA will help to further elucidate the importance of R[142].

The behavior of the two other mutants, T[138]I and F[146]A, is consistent with the chemical shift observations. The F[146]A mutation lowered the binding affinity of Sxl-RBD1+2 into the range of Sxl-RBD1. This reduction in binding affinity (>200-fold) is consistent with this aromatic residue being one of the defining features of the RNP consensus-type RBD. While the chemical shift change of F[146] ($\sim$100 Hz) is not very dramatic, it is significantly larger than the ca. 50 Hz background of changes seen in RBD2. This modest backbone chemical shift change, in spite of the importance of the side chain, could be rationalized by the fact that the backbone is in part of the β-sheet, which probably limits conformational changes which could occur. The inability to quantitatively interpret chemical shift changes in structural terms limits the level of interpretation which can be made.

A substantial chemical shift change at T[138] had previously been noted in Sxl-RBD2 bound to a DNA pyrimidine tract (Kanaar *et al.*, 1995). A mutation of this same residue to isoleucine had also been identified in the temperature-sensitive, hypomorphic Sxl allele, Sxl[f2593]. These observations motivated us to test the T[138]I mutant for RNA binding *in vitro*. This mutant had a very modest reduction in RNA binding ($\sim$2-fold) compared to wt. The limited effect on RNA binding of the T[138]I change encourages the idea that the $\beta2$—$\beta3_{II}$ loop has a relatively passive role in RNA binding, reinforced by the small chemical shift perturbation at this residue in Sxl-RBD1+2. The different response of T[138] in RBD2 alone and RBD1+2 could reflect the difference between RNA and DNA in binding, or an alteration in interaction due to the presence of two domains. Though the effect on RNA binding observed for the T[138]I mutant is small, it could suffice to explain the effect of this partial loss-of-function mutation *in vivo*, since Sxl protein is required for proper splicing of its own transcript. The consequences of a small decrease in RNA binding on splicing regulation could be amplified through this autoregularoty feedback loop. The data presented in Figure 5 should be useful for guiding future mutational analysis of Sxl RNA binding both *in vitro* and *in vivo*.

The $\beta2$-Strand, $\beta2$—$\beta3$ Loop, and C-Terminus. A region of particular interest in RBDs is the so-called specificity loop, the loop connecting $\beta2$ and $\beta3$ strands of an RBD. The idea that the residues in this loop confer RNA-binding specificity is somewhat misleading since the original sequence-swapping experiments carried out on U1A and U2B″ snRNP proteins involved probing residues from the $\beta2$ strand into only the first half of the $\beta2$—$\beta3$ loop (Scherly *et al.*, 1990; Bentley & Keene, 1991); this was a consequence of the sequence identity between U1A and U2B″ in the last few residues of their $\beta2$—$\beta3$ loops. As a result, the specificity conferred by the second half of the loop was not probed. Nevertheless, these studies make clear that residues found in and shortly after the $\beta2$ strand are important for specificity. The substantial chemical shift perturbations in $\beta2_{II}$ and the smaller but significant perturbations observed in $\beta2_I$ of Sxl-RBD1+2 are consistent with the apparent involvement of these strands in sequence-specific recognition.

In the complex of U1A with U1 snRNA stem—loop II (Oubridge *et al.*, 1994), the five-residue $\beta2$—$\beta3$ loop protrudes through the loop of the hairpin RNA, preventing any base-pairing within the loop. This allows the protein to interact with the loop bases as if they were single-stranded RNA. In Sxl-RBD1+2, the two $\beta2$—$\beta3$ loops do not appear to be significantly affected by the RNA. This can be seen in Figure 6, in which the 10-residue $\beta2$—$\beta3_{II}$ loops are primarily colored white. In free Sxl-RBD2, ^{15}N relaxation

measurements showed that the $\beta2-\beta3_{II}$ loop has a significant degree of flexibility (Lee, 1996). This is in agreement with qualitative line widths from free and RNA-bound RBD1+2 which were estimated from triple-resonance data and preliminary ^{15}N T_2 measurements. Therefore, unlike the U1A system, there does not appear to be a rigidification of this loop upon RNA binding. This is not altogether surprising since the RNA substrate for Sxl is not known to be a stem–loop structure. The differences between these systems underscore the wide range of interactions that RBDs can accommodate, presumably by shifted RNA-binding surfaces and structural alterations.

Another region of interest in RBDs is the C-terminus. The C-terminal regions after $\beta4$ in the U1A snRNP and hnRNP-C proteins have been shown to be important for RNA binding (Görlach *et al.*, 1992; Oubridge *et al.*, 1994). Moreover, a short third helix is present at the C-terminus in the free and bound states of U1A (Howe *et al.*, 1994; Oubridge *et al.*, 1994). In Sxl-RBD1+2, E180-K184 comprise the C-terminal region. These residues lack well-defined structure in Sxl-RBD2, but are positioned above the plane of the β-sheet, where interactions with RNA are possible. Indeed, E^{180}, which immediately follows $\beta4_{II}$, experiences a ∼200 Hz chemical shift perturbation. Nevertheless, there are not enough residues following $\beta4_{II}$ to accommodate helix formation, implying that the C-terminal region of Sxl-RBD1+2 (or simply of RBD2) may not be as critical for RNA binding as in hnRNP-C and U1A. In the case of RBD1, if the linker is considered in place of a C-terminus, then its dramatic chemical shift perturbations could make RBD1 more similar to the U1A and hnRNP-C systems.

Potential for Hydrogen Bonding at Peripheral β-Sheet Positions in RBD2. The data in Figure 5 suggest that RBD1 and RBD2 (within Sxl-RBD1+2) are not completely analogous in how they respond to binding of the Tra-PPT 10-mer. In RBD2, the largest shift perturbations are found at the periphery of the β-sheet, rather than in the center of the sheet containing most of the highly conserved residues. The junction of the interdomain linker and $\beta1_{II}$, residues N^{150}–E^{154}, the $\beta2_{II}$-bulge, and the C-terminus collectively form one edge of the putative RNA-binding platform. The linker/$\beta1_{II}$ junction containing the stretch $E^{96}SIKDTN^{102}$ and the neighboring residues $N^{150}KREE^{154}$ could form an array of hydrogen bonds with the RNA or otherwise form intramolecular hydrogen bonds which stabilize protein/RNA contacts. On the opposite side of the sheet, R^{142} is almost certainly involved in critical hydrogen bonds, while the $\beta2_{II}$-bulging residue (V^{128}) and the following residues ($Q^{129}KN^{131}$) form another peripheral ridge with substantial shift perturbations. Most of these peripheral regions have been shown to be important for RNA binding in a variety of RBD proteins. Examples of these include the residues leading into and following the RNP-1 and RNP-2 conserved regions (Jessen *et al.*, 1991; Nagai *et al.*, 1995) as well as the $\beta2$ strand/$\beta2-\beta3$ loop region discussed above. Many of these residues are not conserved among RBDs, suggesting their importance for variability of RNA-binding specificity. The pattern of peripheral perturbations observed in RBD2 is not observed in RBD1. The largest chemical shift perturbations in RBD1 are in the conserved residues in the middle of the β-sheet, and are smaller than the perturbations in RBD2.

Multiple-Domain Proteins and Linkers. The interdomain linker in Sxl-RBD1+2 (R^{92}–T^{101}) experiences a substantial change in chemical environment as evidenced from chemical shift perturbations. No evidence for regular secondary structure was found in the linker in either free or bound states of the protein. Consistent with this, cross-peaks from the central linker residues in all of the triple-resonance data sets were significantly more intense than from structured regions of the protein, suggesting greater flexibility in the linker in both states. Does the linker play a passive role in RNA binding? For a truly passive role, one would not expect the large number of significant chemical shift perturbations seen in the linker of the Sxl-RBD1+2 complex. It is likely that (a) the linker residues are involved in specific contacts and/or that (b) as a result of binding RNA, the two domains become fixed relative to one another and the linker samples a reduced number of environments and conformations, some of which may experience contact shifts from RNA (although reducing flexibility alone could possibly explain the shifts).

The question of linker activity/passivity has been directly addressed in mutational analyses of the first three zinc-finger DNA-binding domains from TFIIIA (Choo & Klug, 1993). These are zinc-fingers of the class I type, the only class in which fingers bind DNA as domain concatemers (Harrison, 1991). The DNA-binding affinity of this protein was demonstrated to be quite sensitive to the amino acid sequences of these two linkers, suggesting that the linkers are somehow *actively* involved in DNA recognition, although exactly how this occurs remains unclear. In the absence of DNA, the two linkers are moderately rigid, and the domains are mobile with respect to each other (Brüschweiler *et al.*, 1995). In the case of the Zif268–DNA complex crystal structure (Pavletich & Pabo, 1991), the relative orientations of the zinc-fingers are fixed, but the linkers apparently do not directly contact the DNA. Unfortunately, there is little structural information on other multiple-domain proteins complexed to nucleic acid sequences.

There may be a fundamental difference between the modular recognition architectures of concatemerized zinc-finger proteins binding DNA and that of Sxl binding the Tra-PPT 10-mer. Zinc-fingers bind to their natural target sequences in a repetitive and predictable manner, each finger binding a three base-pair subsite (Pavletich & Pabo, 1991) via a recognition helix in the major groove of the DNA. In contrast, the asymmetric chemical shift perturbation patterns of RBD1 and RBD2 in Sxl (Figure 5) suggest that each RBD has a unique interaction surface for recognition of the Tra-PPT. Both domains must be interacting with RNA (Kanaar *et al.*, 1995), but the binding appears *not* to be a simple sum of two independent interactions. The $R^{142}A$ mutant data (Figure 8) imply that the RNA-binding activities of RBD1 and RBD2 are coupled, possibly as a result of the two RBDs interacting with each other when bound to overlapping sites on the Tra-PPT. It is also possible that the reduced binding affinity of $R^{142}A$ results only from changes in the free state of RBD1+2 (which prevent RNA binding). Alternatively, Sxl could bind the RNA in two distinct, analogous, subsite interactions. If the RBD/RNA interaction in the Sxl/Tra-PPT complex is similar to the U1A system, at least a seven-base site would be required per RBD (Oubridge *et al.*, 1994; Williams & Hall, 1996). Therefore, at least a 14-base site would be required for high-affinity binding if Sxl binds the Tra-PPT in 2 separate subsite interactions. Because RBD1+2 binds the Tra-PPT 10-mer with the same affinity as the full-length Tra-PPT (Kanaar *et al.*, 1995), the model with overlapping binding sites seems more probable. Another possibility would be that each domain binds distinct five-

base subsites, but this seems less likely in light of the differences of chemical shift perturbation patterns of the two domains. With these considerations, the RNA-binding properties of Sxl have a resemblance to the hnRNP A1 protein, whose two RBDs appear to specifically bind a six-base ssRNA consensus site as a single RNA-binding composite (Burd & Dreyfuss, 1994b).

Unfortunately, with the present data it is not possible to describe the Sxl/Tra-PPT interaction in greater detail. Nevertheless, in one possible scenario, since RBD2 has greater chemical shift perturbations in regions which are suspected to be important for specificity, it might be suggested that RBD2 is more responsible for sequence-specific recognition of the RNA than RBD1. RBD1 is required for specific binding (Kanaar *et al.*, 1995), but a substantial part of its role may be for providing nonspecific stabilizing interactions. This would explain why the largest chemical shift perturbations observed so far in RBD1 are in the conserved RNP-1 and RNP-2 sequences. By definition, these residues should provide a relatively generic contribution to RNA recognition since they are highly conserved across the RNP-type RBD family.

ACKNOWLEDGMENT

SUPPORTING INFORMATION AVAILABLE

Figures showing *J*-correlated connectivities and sequential NOEs for free RBD1+2 protein (Figure S1), calibration curves using τ_c values determined by NMR for 12 small proteins (Figure S2), and chemical shift perturbations for $^{13}C_\alpha$ and $^{13}C'$ resonances due to complexation (Figure S3) (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Allain, F. H. T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D., & Varani, G. (1996) *Nature 380*, 646−650.

Bashaw, G. J., & Baker, B. S. (1995) *Development 121*, 3245−3258.

Bell, L. R., Maine, E. M., Schedl, P., & Cline, T. W. (1988) *Cell 55*, 1037−1046.

Bell, L. R., Horabin, J. I., Schedl, P., & Cline, T. W. (1991) *Cell 65*, 229−239.

Bentley, R. C., & Keene, J. D. (1991) *Mol. Cell. Biol. 11*, 1829−1839.

Biosym Technologies, Inc. (1993) *Felix User Guide Version 2.3*, San Diego, CA.

Birney, E., Kumar, S., & Krainer, A. R. (1993) *Nucleic Acids Res. 21*, 5803−5816.

Brüschweiler, R., Liao, X., & Wright, P. E. (1995) *Science 268*, 886−889.

Burd, C. G., & Dreyfuss, G. (1994a) *Science 265*, 615−621.

Burd, C. G., & Dreyfuss, G. (1994b) *EMBO J. 13*, 1197−1204.

Choo, Y., & Klug, A. (1993) *Nucleic Acids Res. 21*, 3341−3346.

Cline, T. W., & Meyer, J. M. (1996) *Annu. Rev. Genet. 30*, 637−702.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., & Dreyfuss, G. (1992) *EMBO J. 11*, 3289−3295.

Grzesiek, S., & Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.

Grzesiek, S., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593.

Harrison, S. C. (1991) *Nature 353*, 714−719.

Howe, P. W. A., Nagai, K., Neuhaus, D., & Varani, G. (1994) *EMBO J. 13*, 3873−3881.

Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659−4667.

Inoue, K., Hoshijima, K., Sakamoto, H., & Shimura, Y. (1990) *Nature 1990*, 461−463.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.

Kanaar, R., Lee, A. L., Rudner, D. Z., Wemmer, D. E., & Rio, D. C. (1995) *EMBO J. 14*, 4530−4539.

Kay, L. E., Xu, G. Y., & Yamazaki, R. (1994) *J. Magn. Reson. (A) 109*, 129−133.

Kelley, R. L., Solovyeva, I., Lyman, L. M., Richman, R., Solovyev, V., & Kuroda, M. I. (1995) *Cell 81*, 867−877.

Keyes, L. N., Cline, T. W., & Schedl, P. (1992) *Cell 68*, 933−943.

Kranz, J. K., Lu, J., & Hall, K. B. (1996) *Protein Sci. 5*, 1567−1583.

Lee, A. L. (1996) Ph.D. Thesis, University of California, Berkeley.

Lee, A. L., Kanaar, R., Rio, D. C., & Wemmer, D. E. (1994) *Biochemistry 33*, 13775−13786.

Maniatis, T., Sambrook, J., & Fritsch, E. F. (1989) *Molecular Cloning (A Laboratory Manual)* 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Nagai, K., Oubridge, C., Jessen, T., Li, J., & Evans, P. R. (1990) *Nature 348*, 515−520.

Nagai, K., Oubridge, C., Ito, N., Avis, J., & Evans, P. (1995) *Trends Biochem. Sci. 20*, 235−240.

Nicholls, A., Sharp, K., & Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.

O'Neil, M. T., & Belote, J. M. (1992) *Genetics 131*, 113−128.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., & Nagai, K. (1994) *Nature 372*, 432−438.

Pavletich, N. P., & Pabo, C. O. (1991) *Science 252*, 809−816.

Sakashita, E., & Sakamoto, H. (1996) *J. Biochem. 120*, 1028−1033.

Samuels, M. E., Bopp, D., Colvin, R. A., Roscigno, R. F., Garcia-Blanco, M. A., & Schedl, P. (1994) *Mol. Cell. Biol. 14*, 4975−4990.

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., & Mattaj, I. W. (1990) *Nature 345*, 502−506.

Sosnowski, B. A., Davis, D. D., Boggs, R. T., Madigan, S. J., & McKeown, M. (1994) *Dev. Biol. 161*, 302−312.

Valcárcel, J., Singh, R., Zamore, P. D., & Green, M. R. (1993) *Nature 362*, 171−175.

Wang, J., & Bell, L. R. (1994) *Genes Dev. 8*, 2072−2085.

Williams, D. J., & Hall, D. B. (1996) *J. Mol. Biol. 257*, 265−275.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.

Wishart, D. S., & Sykes, B. D. (1994) *J. Biol. NMR 4*, 171−180.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., & Sykes, B. D. (1995a) *J. Biol. NMR 6*, 135−140.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., & Sykes, B. D. (1995b) *J. Biol. NMR 5*, 67−81.

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G., & Mueller, L. (1992) *Biochemistry 31*, 6254−6265.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Zhou, S., Yang, Y., Scott, M. J., Pannuti, A., Fehr, K. C., Eisen, A., Koonin, E. V., Fouts, D. L., Wrightsman, R., Manning, J. E., & Lucchesi, J. C. (1995) *EMBO J. 14*, 2884−2895.

Zijl, P. C. M. V., Johnson, M. O. N., Mori, S., & Hurd, R. E. (1995) *J. Magn. Reson. 113*, 256−270.