Resonance Assignments and Solution Structure of the Second RNA-Binding Domain of Sex-lethal Determined by Multidimensional Heteronuclear Magnetic Resonance[†,‡]

Andrew L. Lee,[§,||] Roland Kanaar,[⊥] Donald C. Rio,[⊥] and David E. Wemmer*[,§,||]

Department of Chemistry, University of California, Berkeley, California 94720-1460, Structural Biology Division of Lawrence Berkeley Laboratory, 1 Cyclotron Road, Berkeley, California 94720, and Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, California 94720

Received July 5, 1994; Revised Manuscript Received September 12, 1994[⊗]

ABSTRACT: The RNA-binding protein Sex-lethal (Sxl) is a critical regulator of sexual differentiation and dosage compensation in *Drosophila*. This regulatory activity is a consequence of the ability of Sxl to bind uridine-rich RNA tracts involved in pre-mRNA splicing. Sxl contains two RNP consensus-type RNA-binding domains (RBDs). A structural study of a portion of Sxl (amino acids 199−294) containing the second RNA-binding domain (RBD-2) using multidimensional heteronuclear NMR is presented here. Nearly complete ^{1}H, ^{13}C, and ^{15}N resonance assignments have been obtained from ^{15}N- and $^{13}C/^{15}N$-uniformly labeled protein. These assignments were used to analyze 3D ^{15}N-separated NOESY and $^{13}C/^{13}C$-separated 4D NOESY spectra which produced 494 total and 169 long-range NOE-derived distance restraints. Along with 41 backbone dihedral restraints, these distance restraints were employed to generate an intermediate-resolution family of calculated structures, which exhibits the $\beta\alpha\beta-\beta\alpha\beta$ tertiary fold found in other RBD-containing proteins. The RMSD to the average structure for the backbone atoms of residues 11−93 is 1.55 ± 0.30 Å, while the RMSD for backbone atoms involved in secondary structure is 0.76 ± 0.14 Å. A capping box [Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605−7609] was identified at the N-terminus of the first helix and has been characterized by short- and medium-range NOEs. Finally, significant structural similarities and differences between Sxl RBD-2 and other RBD-containing proteins are discussed.

The *Drosophila* binary switch gene, *Sex-lethal (Sxl)*,[1] plays a central role in directing events during somatic and germline sexual development. Its expression mode, initiated by upstream events involving the X:A ratio early in development, is dependent upon the sexual state of the organism (Baker, 1989; Cline, 1993; Parkhurst & Meneely, 1994). The functional *Sxl* gene product is the female-specific protein Sxl. The male-specific product is nonfunctional because it is a truncated protein generated by alternative splicing (Bell et al., 1988; Samuels et al., 1993). Sxl protein actively maintains the female-specific expression mode through an autoregulatory feedback loop (Keyes et al., 1992; Cline, 1993). In addition, Sxl is the master regulator of a cascade of downstream events required to maintain female cell fate. This dual regulation is accomplished by a mechanism initially involving transcriptional control (Keyes et al., 1992) and is maintained at the level of pre-mRNA splicing (Bell et al., 1988).

Sxl maintains female cell fate because it acts as a positive regulator of the feminizing gene, *transformer* (*tra*) (Cline, 1979). The *tra* pre-mRNA can be alternatively spliced to yield a non-sex-specific or a female-specific mRNA (Sosnowski et al., 1989). The Sxl protein defines which of these *tra* splicing pathways is taken because the full-length female Sxl protein binds specifically to the pyrimidine-rich tract on the *tra* pre-mRNA just upstream of the 3′ splice site of exon 2 (Inoue et al., 1990). Sxl accomplishes this binding through its two conserved RNA-binding domains (RBDs) (Bell et al., 1988; Valcarcel et al., 1993). Sxl binding prevents splicing to the 3′ splice site of exon 2. Instead, a distal alternative 3′ splice site (exon 3) is activated. It is believed that a mechanism involving Sxl−RNA interactions is also responsible for Sxl autoregulation (Sakamoto et al., 1992). In parallel processes, Sxl is a negative regulator of dosage compensation genes in females and is responsible for

[†] This work was supported by grants from the Director, Office of Energy Research, Office of Biological & Environmental Research, General Life Science Division of the U.S. Department of Energy (under Contract No. DE-AC03-76F00098), the U.S. Department of Energy (DE-FG05-86ER75281), the National Science Foundation (DMB 86-09305 and BBS 87-20134) to D.E.W. for instrumentation grant, the NIH (HD28063 to D.C.R.), and the American Cancer Society, California Division (senior postdoctoral fellowship to R.K.).

[‡] The coordinates of the 17 final structures of Sxl RBD-2 have been deposited in the Brookhaven Protein Data Bank. The file name for these coordinates is 1SXL.

* Corresponding author.

[§] Department of Chemistry.

[||] Structural Biology Division of Lawrence Berkeley Laboratory.

[⊥] Department of Molecular and Cell Biology.

[⊗] Abstract published in *Advance ACS Abstracts,* October 15, 1994.

[1] Abbreviations: Sxl, Sex-lethal; *tra*, transformer gene; RBD, RNA-binding domain; snRNP, small nuclear ribonucleoprotein; hnRNP, heterogeneous nuclear ribonucleoprotein; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; CBCA(CO)NH, C_β to C_α to carbonyl to amide N/HN correlation experiment; 4D NOESY, 4D HMQC-NOESY-HMQC experiment; t_x, time-domain spectrum along the x dimension; f_x, frequency-domain spectrum along the x dimension; TPPI, time-proportional phase incrementation; RMSD, root mean square deviation; N-cap, first residue of a capping box; N1−N3, second through fourth residues of a capping box.

0006-2960/94/0433-13775$04.50/0

increasing the transcriptional activity of the X chromosome in males (Parkhurst & Meneely, 1994). The name Sex-lethal arises from the observation that introduction of a mutant version of the protein in males or females leads quickly to death due to an imbalance of dosage compensation of X chromosomal genes.

It is currently understood that sexual differentiation and maintainance of female cell fate in *Drosophila* is directly linked to the ability of Sxl protein to bind RNA. In fact, the two RBDs are the only well-recognized regions of the 354 residue full-length protein. Studies on other RBD-containing proteins have suggested that specific aromatic side chains are involved in RNA binding (Scherly et al., 1990; Jessen et al., 1991). The amino acid residues carrying these side chains are in the highly conserved octamer (RNP-1) and hexamer (RNP-2) consensus motifs found in all RBDs (Kenan et al., 1991). In this article, we report the resonance assignments and solution structure of the second RBD of Sex-lethal (Sxl RBD-2) as determined by multidimensional heteronuclear NMR techniques.

MATERIALS AND METHODS

Sample Preparation. Expression vector pAR-Sxl-RNP-2 including a T7 promoter (Studier et al., 1990) and cDNA fragment encoding for the C-terminal RNA-binding domain (residues 199−294, hereafter referred to as residues 2−97) of Sex-lethal was transformed into *Escherichia coli* BL21-(DE3)pLysS cells. Cells were adapted and grown on minimal media at 37 °C with $^{15}NH_4Cl$ and [U-$^{13}C_6$]D-glucose as the sole sources of nitrogen and carbon, respectively. Transcription from the T7 promoter was induced by the addition of isopropyl β-D-thiogalactopyranoside to 0.1 mM. Cells were lysed by freeze/thawing, and a crude extract was prepared by centrifugation for 1 h at 40 000 rpm in a 65Ti rotor (Beckman, Fullerton, CA). Nucleic acids were removed from the crude extract by polyethyleneimine precipitation (final concentration 0.5%). $(NH_4)_2SO_4$ was added to the extract to a final concentration of 60%. The resulting precipitate was collected by centrifugation and resuspended in phosphate buffer (50 mM KPO_4, pH 7.0, 50 mM KCl, 1 mM EDTA, 1 mM DTT, and 10% glycerol). This solution was dialyzed against phosphate buffer and run through a Q-Sepharose (Pharmacia, Piscataway, NJ) column. The flow-through was collected and loaded onto a Blue-Trisacryl (IBF Biotechnics, France) column. Sxl RBD-2 protein was eluted with a NaCl concentration gradient of 50 mM to 2 M. Sxl RBD-2 eluted around 0.8 M NaCl. Collected fractions were concentrated, dialyzed against H_2O, and lyophilized. The resultant powder was stored at −20 °C or dissolved in either 0.5 mL of D_2O (99.996%) containing 25 mM d_4-acetic acid, pH 5.0, 25 mM NaCl, and 0.01% NaN_3 or 0.5 mL of an identical mixture with 90%/10% H_2O/D_2O. The final protein concentration was approximately 1.5−2.0 mM. NMR samples were stored at +4 or −20 °C.

Sample purity was assayed with 15% SDS−PAGE and by electrospray-ionization mass spectrometry. Mass spectrometry showed that the first Met residue had been processed off for approximately 70% of the protein molecules. The total mass observed was in agreement with the calculated mass to within 1 mass unit.

NMR Spectroscopy. All NMR experiments were performed at 25 °C on a Bruker AMX 600 spectrometer equipped with an external ENI 50 W linear amplifier and triple-resonance (^1H, ^{13}C, ^{15}N) probe head. Chemical shifts were referenced to 3-(2,2,3,3−$^2H_4$)trimethylsilyl propionate (TSP) (^1H), liquid ammonia (^{15}N) (Live et al., 1984), and TSP (^{13}C) (Bax & Subramanian, 1986). Quadrature detection using TPPI−States (Marion et al., 1989a) was employed for all experiments except for 3D NOESY-HMQC and 3D TOCSY-HMQC, in which case TPPI (Marion & Wüthrich, 1983) was used. All data manipulations as well as assignment analysis were accomplished with the NMR processing software program Felix (2.14β or 2.30β). Linear prediction was carried out using the algorithm supplied with the Felix software package (Barhuijsen et al., 1985; Biosym Technologies, 1993).

^{15}N-Separated 3D NOESY-HMQC (Kay et al., 1989; Marion et al., 1989b) and 3D TOCSY-HMQC (Driscoll et al., 1990) with mixing times of 100 ms each were acquired with spectral widths of 7246 Hz for ^1H dimensions (t_1 and t_3) and 1861 Hz for the ^{15}N (t_2) dimension. In the case of TOCSY-HMQC, a compensation delay during the MLEV-17 spin-lock was used to suppress NOESY-type cross peaks (Griesinger et al., 1988). The number of real data points collected for t_1, t_2, and t_3 were 220, 64, and 2048, respectively. The time domain data were processed with a 70° phase-shifted sine-bell apodization function in each dimension, with linear prediction used in t_2 to double the number of data points prior to apodization. Because the ^1H carrier in t_3 was placed at 4.78 ppm (H_2O resonance), the upfield data points in this dimension were discarded following Fourier transformation. Zero-filling was included for t_1 to yield a 512 × 64 × 512 real data matrix.

The triple-resonance 3D CBCA(CO)NH (Grzesiek & Bax, 1992) experiment was acquired with spectral widths of 8446, 1861, and 7246 Hz for t_1(^{13}C), t_2(^{15}N), and t_3(^1H), respectively. Total acquisition time was 68 h. Using processing similar to 3D NOESY and TOCSY above, a 128 × 64 × 512 real data matrix resulted from 48, 32, and 512 complex data points, respectively. Zero-filling was used in t_1 only.

^{13}C-Separated 3D HCCH-TOCSY was used to assign side chain ^1H and ^{13}C resonances (Bax et al., 1990). Carrier frequencies were centered at 3.00 and 42.94 ppm for proton and carbon, respectively. Spectral widths of 4348 Hz (t_1 and t_3) and 3086 Hz (t_2) resulted in a spectrum aliased in t_2 only. Single aliasing from upfield and downfield is desirable for the ^{13}C dimension because it increases digital resolution up to 3-fold without introducing ambiguities (Clore et al., 1990). A total of 128(^1H), 32(^{13}C), and 256(^1H) complex data points were processed into a 256 × 64 x 512 real data matrix. As usual, linear prediction was used for the dimension with the lowest digital resolution (t_2 in this experiment), and zero-filling was used whenever possible.

The $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC ($^{13}C-^1H-^{13}C-^1H$) experiment (Clore et al., 1991) was acquired over 88 h with a mixing time of 100 ms. Carrier frequencies were centered at 3.81 and 42.94 ppm for proton and carbon, respectively, and aliasing similar to the 3D HCCH-TOCSY was employed for both ^{13}C dimensions. A total of 8, 64, 8, and 256 complex points for each dimension were acquired with corresponding spectral widths of 3086, 4808, 3086, and 5208 Hz. After data acquisition, t_4 was processed with a 70° phase-shifted sine-bell window into a 32 × 128 × 32 × 256 real data matrix followed by similar processing in t_2. Next, t_3 was only transformed (not apodized) in order to spread out the signals to ensure accurate linear prediction in t_1 (Clore et al., 1991). Finally, t_3 was inverse transformed,

linear predicted, and processed to yield the fully processed frequency domain matrix. In both t_1 and t_3, linear prediction was used to extend the data by 50%, and zero-filling was used in t_1, t_2, and t_3.

2D HSQC experiments (Bodenhausen & Ruben, 1980) were used to initially judge the quality of samples, as well as to accurately define resonances due to the high digital resolution available. In the case of ^{15}N-^{1}H HSQC, amide ^{1}H exchange rate information was also obtained. For the ^{13}C−^{1}H correlation spectra, a constant-time evolution period ($2T$ equal to $1/J_{CC}$) was used to eliminate ^{13}C−^{13}C couplings in t_1 (Vuister & Bax, 1992). In the case of aliphatic ^{13}C−^{1}H resonances, the carbon carrier frequency was placed at 46.13 ppm, and spectral widths of 5000 Hz (128 complex points) in t_1 and 7246 Hz (2048 real points) in t_2 were used. In the case of aromatic ^{13}C−^{1}H resonances, a second HSQC was acquired with the carbon carrier placed at 125.13 ppm, using spectral widths of 3968 Hz (56 complex) in t_1 and 7246 Hz (2048 real) in t_2. The aromatic region of the resultant spectrum has a small number of resonances. In addition, the aromatic rings have no carbonyl groups and very few nitrogens. For these reasons, carbonyl and nitrogen decoupling pulses were removed from the pulse sequence. For the ^{15}N−^{1}H HSQC spectra, the ^{15}N carrier was placed at 116.2 ppm, with spectral widths of 1861 Hz (256 complex) in t_1 and 7246 Hz (2048 real) in t_2. For the purpose of estimating amide exchange rates, a rapid-acquisition pulse program (Marion et al., 1989a) was used to collect early time points for a series of ^{15}N−^{1}H HSQC experiments on Sxl RBD-2 immediately following dissolution in 99.9% D_2O, where an entire 2D experiment with 128 complex t_1 points is collected in 6 min.

The 2D HMQC-J experiment (Kay & Bax, 1990) was used to obtain qualitative $^3J_{HN\alpha}$ coupling constants, which in turn could be used to define ϕ dihedral angle ranges via the Karplus relation. For J values >8 Hz, a ϕ dihedral restraint of $-120 \pm 40°$ was implemented in structure calculations. For J values <6 Hz, a ϕ dihedral restraint of $-60 \pm 30°$ was implemented if that residue was known to occur in a helical region. Data aquisition parameters are identical to the ^{15}N−^{1}H HSQC experiments, except that aquisition in t_1 was extended to 200 ms. An exponential line broadening of 7 Hz was applied in the directly detected dimension as described (Kay & Bax, 1990).

Structure Calculations. The program X-PLOR 3.1 (Brünger, 1992) was used for all structure calculations on a Silicon Graphics (Mountain View, CA) workstation. The calculation protocol consisted of distance geometry on substructures comprised of backbone as well as C_β and C_γ atoms, followed by high temperature molecular dynamics, simulated annealing, and Powell energy minimization as described (Nilges et al., 1988; Brünger, 1992). More details of the protocol are presented in Table 1. The calculations were carried out in an iterative fashion, initially using completely unambiguously assigned NOEs. The resultant structures were used to evaluate additional NOEs which were then included to increase the number of restraints. After the global fold had been established, torsional and hydrogen bond restraints were added. NOEs were classified qualitatively as strong (1.8−2.7 Å), medium (1.8−3.5 Å), or weak (1.8−5.0 Å) (Williamson et al., 1985; Clore et al., 1986). As a reference, NOEs between H_δ and H_ϵ in Tyr residues were considered strong, while NOEs between $H_\alpha(i)$ and HN-(i+3, i+4) in α-helices were considered weak. A correction was added to the upper limit for constraints involving methyl protons and nonstereospecifically assigned protons (Wüthrich et al., 1983; Clore et al., 1986). Hydrogen bonds involving slowly exchanging amide protons were included if a unique acceptor could be identified from the structures and the pattern of NOEs. A total of 26 hydrogen bonds, included as 26 HN↔O (1.8−2.3 Å) and 26 N↔O (2.8−3.3 Å) NOE-type distance restraints, were identified.

Table 1: Structural Statistics and RMSDs for 17 Sxl RBD-2 Structures[a]

structural statistics	⟨SA⟩
X-PLOR energies (kcal mol^{-1})	
E_{tot}	293 ± 43
E_{bond}	13 ± 3
E_{angle}	170 ± 14
E_{impr}	24 ± 3
E_{vdw}[b]	46 ± 15
E_{noe}[c]	37 ± 13
E_{cdih}[d]	2.1 ± 1.1
RMSD from idealized covalent geometry	
bonds (Å)	0.0030 ± 0.0003
angles (deg)	0.63 ± 0.02
impropers (deg)	0.45 ± 0.03
rms deviations from experimental distance restraints (Å)[e]	0.036 ± 0.006
rms deviations from experimental dihedral restraints (deg)[f]	0.91 ± 0.26

atomic RMSDs (Å)	N, C_α, C′, O	all non-H
⟨SA⟩ vs $\langle\overline{SA}\rangle_{sxl}$	1.55 ± 0.30	2.20 ± 0.32
⟨SA⟩ vs $\langle\overline{SA}\rangle_{str}$	0.76 ± 0.14	1.35 ± 0.13

[a] Notation is as follows: ⟨SA⟩ is the ensemble of 17 final X-PLOR structures. $\langle\overline{SA}\rangle_{sxl}$ is the mean coordinate set for residues 11−93 obtained from a least-squares superposition to backbone or non-H atoms. $\langle\overline{SA}\rangle_{str}$ is the mean coordinate set for residues involved in secondary structure (15−19, 26−35, 40−46, 57−62, 65−77, 88−91) which was obtained from a least-squares superposition to backbone or non-H atoms. [b] The X-PLOR F_{repel} function was used to simulate the van der Waals potential with atomic radii ranging from 0.9 times their CHARMM (Brooks et al., 1983) values at high temperatures to 0.75 times their CHARMM values at low temperatures (Brünger, 1992). [c] NOE-derived distance restraints were applied with a square-well potential with force constants of 50 kcal·mol^{-1}·Å^{-2}. [d] Dihedral angles were given force constants of 200 kcal·mol^{-1}·rad^{-2} which were applied at the beginning of the annealing/refinement stage. [e] The single NOE violation >0.5 Å in the family of 17 structures is 0.61 Å. [f] There is one dihedral violation >5° in the family of 17 structures, and it is 5.40°.

RESULTS

Resonance Assignments. Figure 1 shows the 2D ^{15}N−^{1}H HSQC of uniformly ^{15}N-labeled Sxl RBD-2. The wide dispersion of peaks and the absence of additional peaks for individual residues indicates that the protein has a well-defined tertiary fold. Labeled peaks were assigned using the methodology described below.

Assignments for N, HN, and H_α as well as unambiguous sidechain proton resonances were made from analysis of ^{15}N-separated 3D TOCSY-HMQC and NOESY-HMQC spectra (Kay et al., 1989; Marion et al., 1989b; Driscoll et al., 1990). After identifying spin systems in the TOCSY-HMQC, these spin systems could be specifically identified by residue number if there were sequential NOEs to the preceding or following amino acid in the known sequence (Wüthrich, 1986). HN(i)−HN(i+1), $H_\alpha(i)$−HN(i+1), and $H_\beta(i)$−HN-(i+1) connectivities were found in helices, while for residues in an antiparallel β-sheet, $H_\alpha(i)$−HN(i+1) and $H_\beta(i)$−HN-(i+1) were generally found. This analysis was carried out

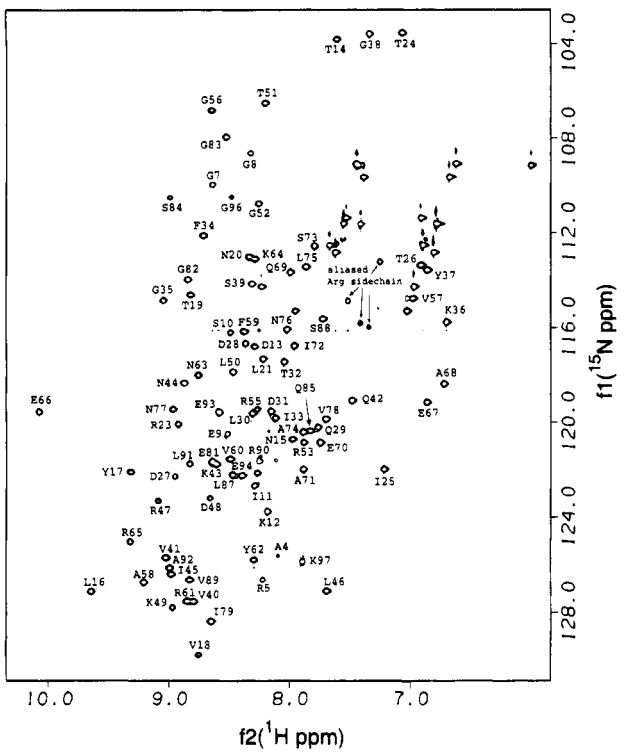

FIGURE 1: 2D HSQC $^{15}N-^{1}H$ correlation spectrum of uniformly ^{15}N-labeled Sxl RBD-2 and assignments. Unlabeled cross peaks between 108 and 116 ppm are NH_2 groups from Asn and Gln side chains.

interactively using locally written macros for the NMR processing and analysis software package Felix (version 2.14β or 2.10). In order to illustrate the quality of the data, Figure 2A shows a series of sequential assignments in the form of plotted 2D strips from the 3D NOESY/TOCSY pair. Figure 2B–D forms a specific example of a sequential assignment from F59 to A58 using Felix macros which are now described for the general case. With the TOCSY-HMQC matrix in one window frame and the NOESY-HMQC matrix in a second frame, $^{1}H(f_2)/^{15}N(f_3)$ 2D slices at the same $^{1}H(f_1)$ frequency were displayed from which a sequential NOESY cross peak or intraresidue TOCSY cross peak could be considered (Figure 2B,C). In the case of considering a $H_\alpha(i-1)$ sequential NOESY cross peak at N(i) and HN(i) frequencies, the corresponding intraresidue cross peak in the TOCSY could be identified by screening all TOCSY spin systems that had peak intensities in the displayed $^{1}H(f_2)/^{15}N(f_3)$ 2D slice. This screen was performed by displaying all candidate $^{1}H(f_1)$ TOCSY 1D vectors in a third window frame (Figure 2D) as a stacked plot. These TOCSY spin system candidates could then be lined up with the NOESY $^{1}H(f_1)$ vector which contained the $H_\alpha(i-1)$ sequential cross peak, and the correct $i-1$ spin system could be confirmed if its chemical shift pattern matched the expected pattern based on the known amino acid sequence. Once this was confirmed, this procedure could be continued by displaying new 2D planes which corresponded to a $H_\alpha(i-2)$ sequential NOESY cross peak at N($i-1$) and HN($i-1$) frequencies. Following this procedure allows one to "walk" backward through the sequence until a proline or amino terminus is encountered. In a similar fashion, a second Felix macro allows moving forward through the sequence by starting from an intraresidue TOCSY cross peak and screening $H_\alpha(i)$/HN($i+1$) sequential NOESY cross peak candidates. The search can also be conducted from H_β or HN resonances instead of H_α cross peaks. This approach to the sequential resonance assignments proved to be convenient since all sequential spin system candidates in three dimensions could be reduced, visualized, and evaluated as a compact spectral subset in the 1D stacked plot display. In addition, the visualization of very weak resonances is enhanced with this method because there is no need to define a contour threshold as required for 2D plots.

^{13}C assignments and all remaining unassigned proton side chain resonances (except aromatic rings) were assigned from the ^{13}C-separated 3D HCCH-TOCSY experiment (Bax et al., 1990; Clore et al., 1990). Figure 3 displays the high efficiency of magnetization transfer observed for various spin systems. The more complete spin systems here were matched with incomplete proton spin-system data from the ^{15}N-separated 3D TOCSY-HMQC, thereby allowing unambiguous sequence-specific assignments to be made for nearly all side chain resonances. Exact placement of correlated $^{1}H/^{13}C$ pairs in the side chains was relatively unambiguous due to the characteristic ^{13}C chemical shifts for amino acid side chain resonances in proteins (Clore et al., 1990; Ikura et al., 1991). Aromatic resonances were assigned by analysis of the $^{13}C/^{13}C$-separated 4D NOESY (Clore et al., 1991) spectrum. NOEs were generally observed from H_β (and/or H_α) to H_δ. In addition, strong NOEs were generally observed between adjacent ring protons, allowing assignments to be made for most aromatic resonances.

In order to verify and complete the resonance assignments, a triple-resonance CBCA(CO)NH spectrum (Grzesiek & Bax, 1992) was acquired. This experiment correlates backbone N(i) and HN(i) with $C_\alpha(i-1)$ and $C_\beta(i-1)$, using one-bond scalar couplings via the backbone carbonyl carbon to bridge sequential resonances. The resultant 3D data proved to be an excellent complement for NOE-based sequential assignments, identifying erroneous assignments derived from proton chemical shift degeneracies, as well as allowing the specific assignment of C_α and C_β resonances of Pro residues.

Secondary Structure. The secondary structure in Sxl RBD-2 consists of two α-helices and one antiparallel β-sheet. The characteristic NOE patterns for an α-helix include the following: HN(i)–HN($i+1$), $H_\beta(i)$–HN($i+1$), $H_\alpha(i)$–HN($i+3$), $H_\alpha(i)-H_\beta(i+3)$, and $H_\alpha(i)$–HN($i+4$). The characteristic NOE pattern for an antiparallel β-sheet is a strong $H_\alpha(i)$–HN($i+1$) along with long-range NOEs to opposite strands, such as strong $H_\alpha-H_\alpha$ and weak H_α–HN, HN–H_α, and HN–HN NOEs (Figure 5) (Wüthrich, 1986). Backbone amide protons involved in secondary structure are often hydrogen bonded in a regular fashion, giving rise to slow exchange of these protons. On the basis of these criteria, the secondary structure in Sxl RBD-2 was mapped and is summarized in Figure 4. Recently, carbon chemical shifts have been observed to correlate with secondary structural elements (Spera & Bax, 1991; Wishart et al., 1991). It has been observed that helices give rise to downfield shifts from random coil values of a few ppm for C_α, as well as a small upfield shift for C_β. An opposite trend has been observed for β-sheets. Figure 4 also reports C_α and C_β deviations from random coil values.

The first α-helix begins with T26 and ends at G35. T26 is the N-cap residue in a capping box, an α-helix initiating signal comprised of four amino acids which recently has been recognized in both peptides and proteins (Harper & Rose, 1993). In a capping box, the side chain of the first (N-cap) of these four residues forms a hydrogen bond with the

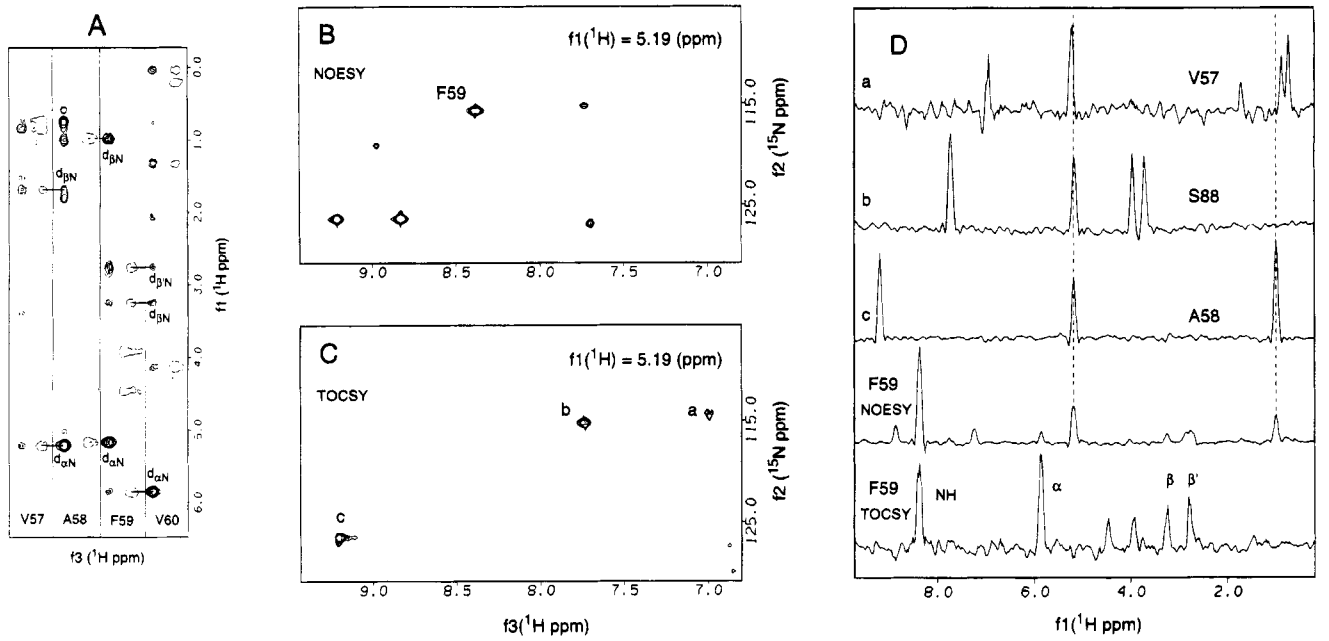

FIGURE 2: Various 2D and 1D slices from ^{15}N-separated 3D NOESY-HMQC and TOCSY-HMQC to illustrate the resonance assignments strategy used. (A) 2D strips are plotted for four different f_3(^{1}H) and f_2(^{15}N) values, which correspond to residues V57−V60. Within each strip, NOESY peaks are on the left with multiple contour levels, and TOCSY peaks are on the right with a single contour level. Horizontal lines represent sequential NOE connectivities. Figure 2B−D displays the three window frames viewed while running the Felix assignment macros (see text). Figure 2 panels B and C show a 2D slice at f_1(^{1}H) = H$_\alpha$(i−1) from NOESY and TOCSY, respectively. (B) 2D slice from 3D NOESY-HMQC. The peak labeled F59 is a sequential cross peak from HN(i) of F59 to H$_\alpha$(i−1) of A58. (C) 2D slice from 3D TOCSY-HMQC. Candidate H$_\alpha$(i−1) peaks for A58 are labeled a−c. (D) 1D TOCSY vectors along f_1(^{1}H) corresponding to a−c are shown on the top. The bottom two 1D vectors correspond to f_2(^{15}N)/f_3(^{1}H) of F59, with H$_\alpha$, H$_\beta$, and H$_{\beta'}$ labeled for the TOCSY spin system. Dashed lines in panels A−D represent searches based on sequential NOESY peaks from F59 HN to A58 H$_\alpha$ and H$_\beta$. Because these dashed lines line up with TOCSY spin system c, c must correspond to A58.

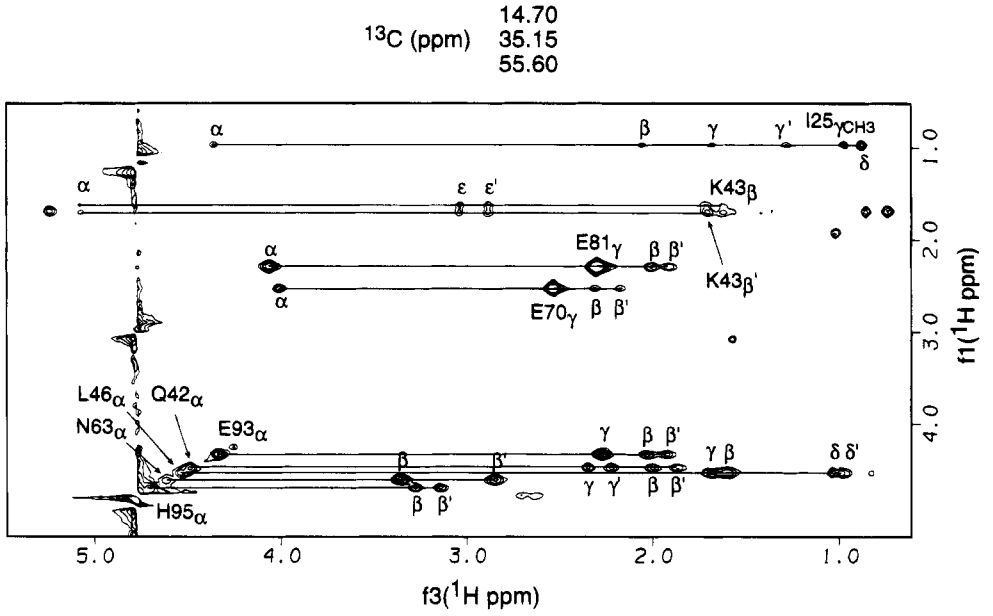

FIGURE 3: 2D slice of ^{13}C-separated 3D HCCH-TOCSY at f_2(^{13}C) = 14.70, 35.15, and 55.60 ppm. Three ^{13}C frequences are represented due to aliasing in f_2. All resonances centered in this plane are labeled with their assignments. Lines represent individual amino acid spin systems.

backbone amide of the fourth (N3) residue, while in turn a reciprocal hydrogen bond from the N3 residue side chain back to the backbone amide of the N-cap residue completes the "box". With such an interaction, two net hydrogen bonds are gained, which has been shown to be energetically favorable. Hence the N-cap and N3 residues are preferably Thr, Ser, Asn, Gln, Asp, or Glu, since the side chain carbonyl or hydroxyl oxygen serves as a suitable hydrogen bond acceptor. A capping box has previously been proposed for hnRNP-A1 on the basis of sequence analysis and ^{13}C chemical shift data (Garrett et al., 1994). In the case of Sxl RBD-2, characteristic NOEs, chemical shifts, and amide protection patterns are observed as expected for the capping box. There is a medium-intensity NOE from T26 HN to Q29 H$_\beta$, a strong NOE from T26 H$_\alpha$ to D27 HN, and NOEs from the side chain protons of T26 to D28 HN. This NOE pattern is consistent with the amide of T26 being oriented nearly antiparallel to all other amides in helix 1, allowing the side chain (Q29 O$_\epsilon$)-to-main chain (T26 NH) hydrogen bond. A similar NOE pattern was observed for a capping

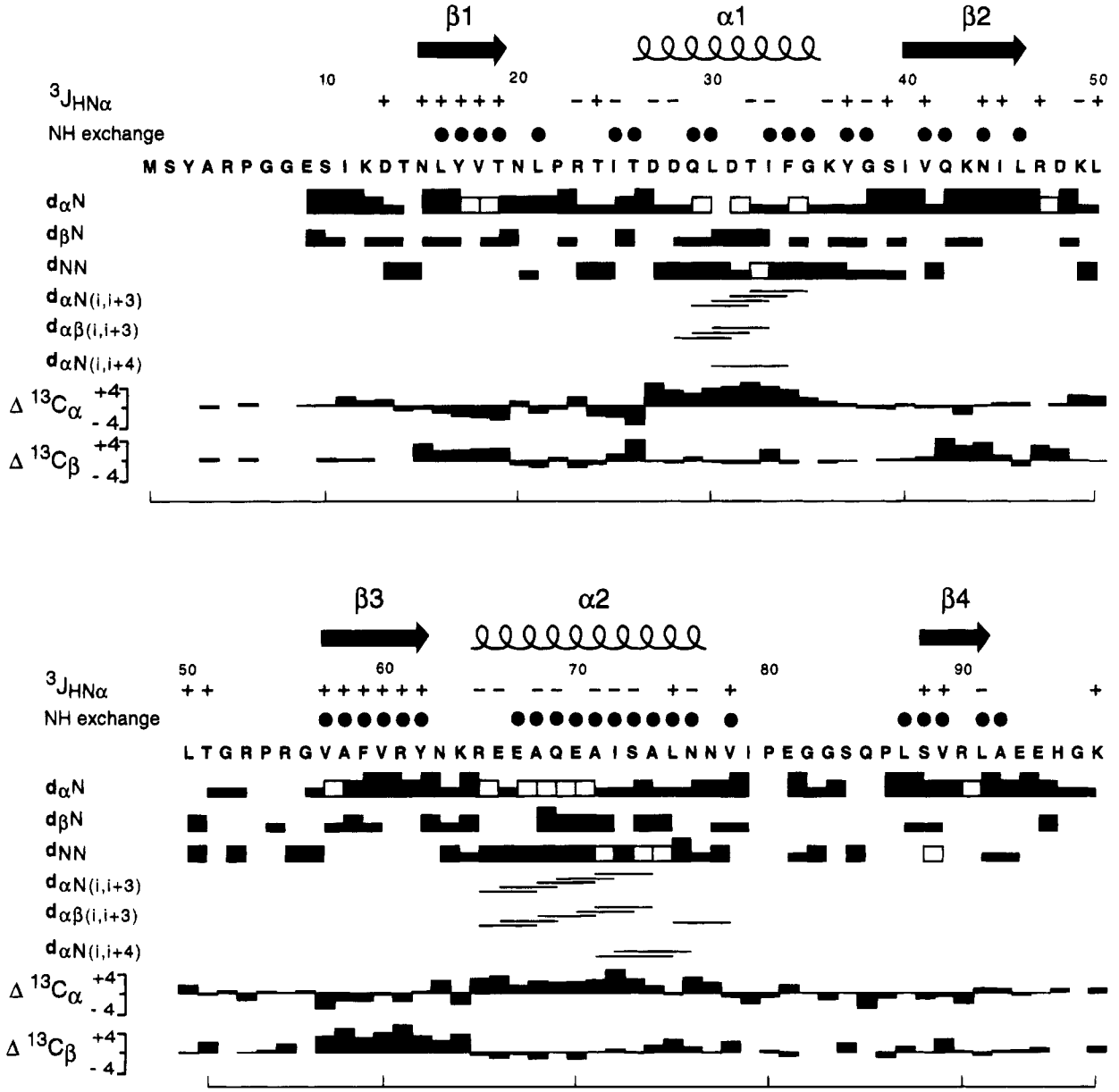

FIGURE 4: Summary of sequential and medium-range NOE connectivities, amide hydrogen exchange rates, $^3J_{HN\alpha}$ coupling constants, and deviations of $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts from random coil values. NOE intensities are denoted by bar height as weak, medium, or strong. Open boxes indicate ambiguity due to chemical shift degeneracy or solvent presaturation. For Pro residues, NOE intensities to their H_δ protons were substituted for NOE intensities to HN protons. Filled circles indicate residues with slowed amide exchange (HN resonances observable 3 h after dissolving lyophilized protein into 99.9% D_2O). Pluses and minuses represent $^3J_{HN\alpha}$ coupling constants greater than 8 Hz or less than 6 Hz, respectively. Secondary structure as determined by NOESY data is indicated above the sequence numbering.

box in a short peptide (Zhou et al., 1994). Chemical shift values for C_α and C_β of T26 have shifts from random coil values of −4 and +3 ppm, respectively. These shifts are characteristic for capping boxes (Gronenborn & Clore, 1994). Finally, the amides of T26 and Q29 are protected from exchange with solvent (Figure 4), while the amides of D27 and D28 exchange rapidly. This pattern of amide protection is consistent with the previous model of a capping box (Zhou et al., 1994). Along with the first α-helical hydrogen bond from L30 NH to T26 CO, T26 is completely fulfilled with three hydrogen-bonding interactions. From the calculated structures, the values for ϕ and ψ of T26 are $-105 \pm 21°$ and $150 \pm 10°$, respectively, which is consistent with other observed capping boxes (Harper & Rose, 1993). The C-terminal end of helix 1 will be discussed below. The second α-helix begins at R65 and ends with N76, running slightly longer than helix 1. It is very well defined due to a continuous stretch of helical-type NOEs (Figure 4) and hydrogen bonds (as detected by slowed amide exchange).

For the previously structurally characterized snRNP and hnRNP RBD proteins (Nagai et al., 1990; Wittekind et al., 1992), it has been suggested that a four-stranded antiparallel β-sheet provides a "platform" for RNA binding. This could be the case for Sxl RBD-2 because it also contains a four-stranded antiparallel β-sheet. The topology as well as NOEs and hydrogen bonds for this sheet are shown in Figure 5. Strands β1 and β3 contain the highly conserved RNP-2 and RNP-1 consensus sequences, respectively. The two aromatic side chains of Y17 and F59 are expected to interact directly with *tra* pre-mRNA (Merrill et al., 1988; Hoffman et al., 1991), and they are situated directly cross-strand from one another in the center of the sheet, which forms a regular pleated pattern. There is a β-bulge involving I40 and V41, which appear similar to residues I40 and L41 in the snRNP-U1A X-ray crystal structure (Nagai et al., 1990). Figure 5 shows the unusual pattern of NOEs resulting from the bulge.

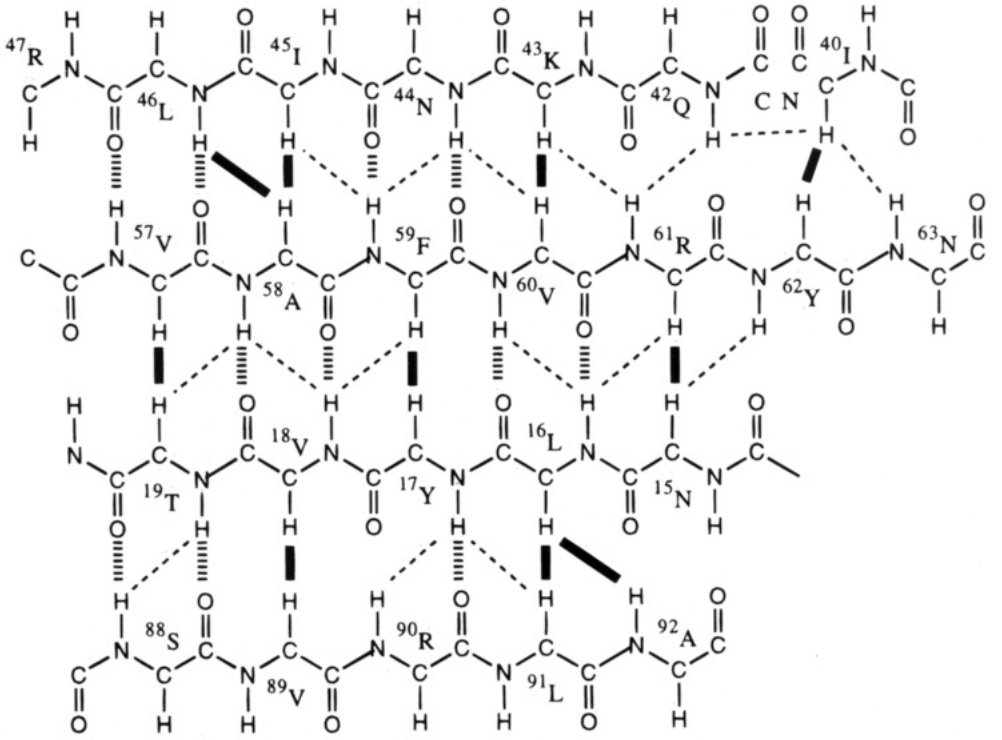

FIGURE 5: Antiparallel β-sheet of Sxl RBD-2. Residues are labeled at their own C_α. Observed weak NOEs are represented by dashed lines. Observed medium or strong NOEs are represented by thick lines. Hydrogen bonds consistent with NOE and exchange data are represented by hatched lines.

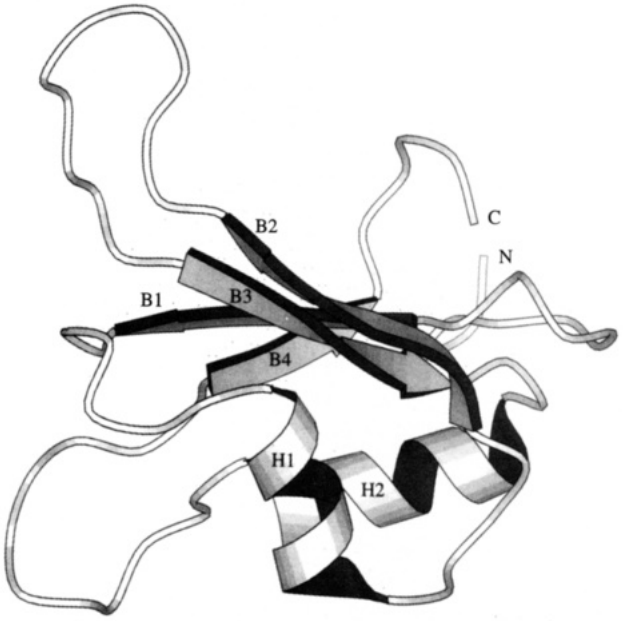

FIGURE 6: Ribbon diagram of Sxl RBD-2. Secondary structural elements are labeled accordingly and shown as ribbons. All other regions are shown as tubes. The figure was created with MOLSCRIPT software (Kraulis, 1991).

The bulge can be visualized in a ribbon diagram of Sxl RBD-2 shown in Figure 6. Aside from this β-bulge, the only other unusual feature of the sheet structure is the fourth strand, S88−L91. Despite observation of typical $H_\alpha-H_\alpha$, $H_\alpha-HN$, and HN−HN cross-strand NOEs, the amide exchange pattern is atypical. For example, the amide proton of R90 exchanges rapidly with solvent, implying the absence of a stable hydrogen bond with Y17 (Figure 5). In addition, V89 HN is oriented down toward helix 2, giving rise to numerous NOEs between it and helix 2. Finally, a $^3J_{HN\alpha}$ value <6 Hz is observed for L91, where one might expect a value >8 Hz for an extended β conformation. In the calculated structures, residues S88−L91 appear to be in an extended conformation, but the strand as a whole is tilted away from the rest of the sheet. This feature is also observed in the crystal structure of snRNP-U1A.

NOE-Derived Restraints. In order to carry out structure calculations, interproton NOEs were assigned from the ^{15}N-separated 3D NOESY-HMQC and $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC. A large number of intraresidue and short-range NOEs had been previously identified during the assignment of the backbone. Medium- and long-range (as well as more intraresidue and short-range) NOEs were assigned by matching NOE cross peak frequencies against existing resonance assignments. Each NOE was assigned an intensity of weak, medium, or strong on the basis of the number of contour levels for that peak. In the end, a total of 241 NOEs were obtained from the ^{15}N-separated 3D NOESY-HMQC and 253 NOEs were obtained from the $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC.

Calculated Structures. A total of 587 experimental restraints were included in the final round of structure calculations. This total consisted of 50 intraresidue, 185 sequential, 90 medium-range ($|i - j| \leq 4$), and 169 long-range NOEs as well as 41 ϕ dihedral and 52 hydrogen bond restraints (from 26 hydrogen bonds). Hydrogen bonds were included for nonexchangeable amides in stretches of regular secondary structure where hydrogen-bond acceptors were unambiguous. In the case of the capping box, two additional hydrogen bonds were included. Justification for these is given in the Discussion. Of the 30 structures calculated, 17 converged to reasonably low energies. This subfamily had a single NOE violation greater than 0.5 Å. Table 1 and Figure 7 contain calculation statistics for the 17 accepted structures. A superposition to the average coordinates for these 17 structures is shown in Figure 8 and exhibits the well defined backbone atoms involved in secondary structure. The RMSD to the average coordinates for backbone atoms

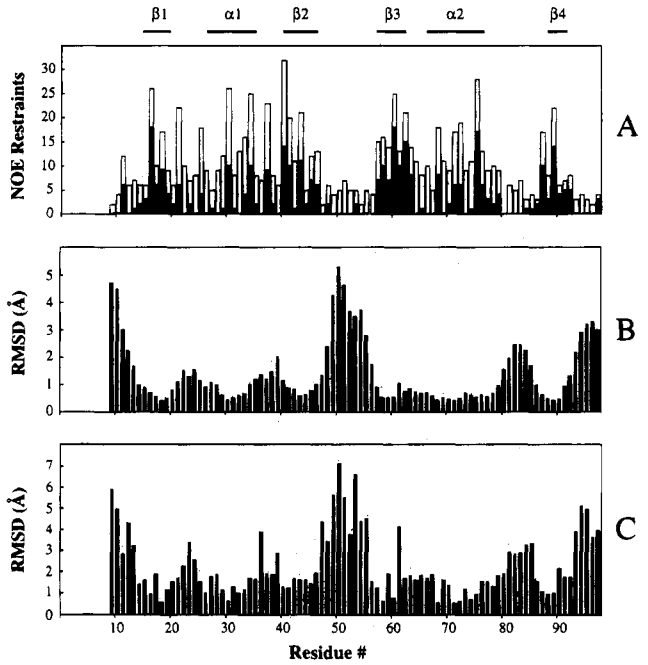

FIGURE 7: (A) Distribution of NOE restraints used in structure calculations. The bar height reflects the total number of NOEs for a given residue. Filled portions represent long-range NOEs. Secondary structure is denoted at the top of the figure. (B) Residue-based root mean square deviation to the mean coordinates for backbone atoms. (C) Residue-based root mean square deviation to the mean coordinates (backbone-superimposed) for non-hydrogen atoms.

(of the accepted family) involved in secondary structure is 0.76 ± 0.14 Å. The RMSD for backbone atoms of residues 11−93 is 1.55 ± 0.30 Å. This difference is primarily due to the relatively unrestrained loop regions (47−56, 79−86) connecting elements of secondary structure. The RMSD for all non-hydrogen atoms for residues 11−93 is higher at 2.20 ± 0.32 Å. However, Figure 7C shows that residues involved in secondary structure frequently have non-hydrogen RMSDs below 1 Å, especially residues in the hydrophobic core (see below). Usually this is a consequence of a relatively large number of long-range NOEs for these particular residues, as shown in Figure 7A.

The folding topology of Sxl RBD-2 is similar to previously described RNA-binding proteins, consisting of a $\beta\alpha\beta-\beta\alpha\beta$ folding pattern (Figure 6). However, Sxl RBD-2 has some significant differences, such as the 10 residues between $\beta 2$ and $\beta 3$ as opposed to five, four, and eight residues in snRNP-U1A, hnRNP-C, and hnRNP-A1, respectively. This turn or loop has been implicated in conferring specificity to binding RNA for the snRNP-U1A protein (Scherly et al., 1990; Mattaj, 1993) and is expected to be important for all proteins of this class. Sxl RBD-2 has one Lys and three Arg residues in this 10 residue stretch alone, and in the absence of RNA, the lack of long-range NOEs suggests that this loop is flexible. A second loop connecting α2 and $\beta 4$ also has few medium- and long-range NOEs, leaving it poorly defined in the calculated structures.

The two α-helices are anchored to the bottom of the sheet by a network of helix−helix and helix−sheet NOEs, which define the relative orientations of these elements (Figures 6 and 8) and reflect the contacts that define the hydrophobic core of the protein. At the present stage of refinement, the overall conformation of the protein resembles hnRNP-C more closely than snRNP-U1A. This is primarily because the helices in both hnRNP-C and Sxl RBD-2 are nearly perpendicular to each other. In snRNP-U1A, helix 1 extends for an extra turn, and the two helices form a more acute interhelical angle (Nagai et al., 1990).

The backbone atoms involved in the RNP-1 and RNP-2 conserved sequences are especially well defined. Consequently, this portion of the ensemble of 17 structures presents an accurate representation of what is believed to be the RNA binding surface. Very few long-range NOEs involving the side chains of F59 and Y17, aromatic groups implicated in RNA binding affinity, were observed, suggesting that these side chains are mobile in the absence of RNA. In the calculated structures, these side chains are found in quite different orientations. Similarly, other side chains (N15, T19, R47, K49, R53, R55, R61, H95, K97) which might be presumed to be involved in RNA recognition sample a wide range of orientations in the structures. These observations are not surprising since all of these side chains are either directed away from the hydrophobic core (side chains of L16, L30, I33, F34, Y37, I40, V60, Y62, A68, A71, and L75) or in flexible regions.

DISCUSSION

Expressed, purified, and uniformly isotopically labeled (^{15}N or ^{13}C/^{15}N) Sxl RBD-2 protein has been 90% assigned using a combination of ^{15}N-separated 3D TOCSY-HMQC and NOESY-HMQC, ^{13}C-separated 3D HCCH-TOCSY, and triple resonance 3D CBCA(CO)NH spectra. A complete list of these assignments is provided in Table 2. Once the assignments were completed, peaks were manually picked from 4D ^{13}C/^{13}C-separated HMQC-NOESY-HMQC and 3D ^{15}N-separated NOESY-HMQC spectra. The chemical shifts of these peaks were then matched against the assignments to generate qualitative distance restraints, which were used to calculate families of structures using X-PLOR 3.1. The overall three-dimensional fold was found to be consistent with folds of previously structurally characterized RBD-containing proteins.

Comparison with Other RBD Structures. There are many proteins which contain single or multiple RBDs. Many of these are known to be involved in RNA processing as well as regulation of gene expression (Kenan et al., 1991; Mattaj, 1993). In addition to an overall sequence similarity, these proteins have highly conserved octamer (RNP-1) and hexamer (RNP-2) sequences, which occur in the central two strands of the β-sheet. Of this class of proteins, a few have been structurally characterized to date. The first is snRNP-U1A, for which a 2.8-Å resolution X-ray crystal structure was solved (Nagai et al., 1990). The second is a solution structure of the hnRNP-C RBD solved by NMR methods (Wittekind et al., 1992; Görlach et al., 1992). Most recently, NMR resonance assignments and secondary structure analysis have been given for hnRNP-A1 (Garrett et al., 1994). It is clear that all four proteins have the same global folding pattern, two helices packed up against a four-stranded antiparallel β-sheet. This fold was originally predicted through secondary structure prediction algorithms of published RBD sequences followed by model building based on acylphosphatase (Ghetti et al., 1990). The primary difference among these proteins is the variability in the length of loop $\beta 2-\beta 3$: five amino acids in snRNP-U1A, four in hnRNP-C, eight in hnRNP-A1, and ten in Sxl RBD-2. In the case of snRNP-U1A, this loop has been implicated in

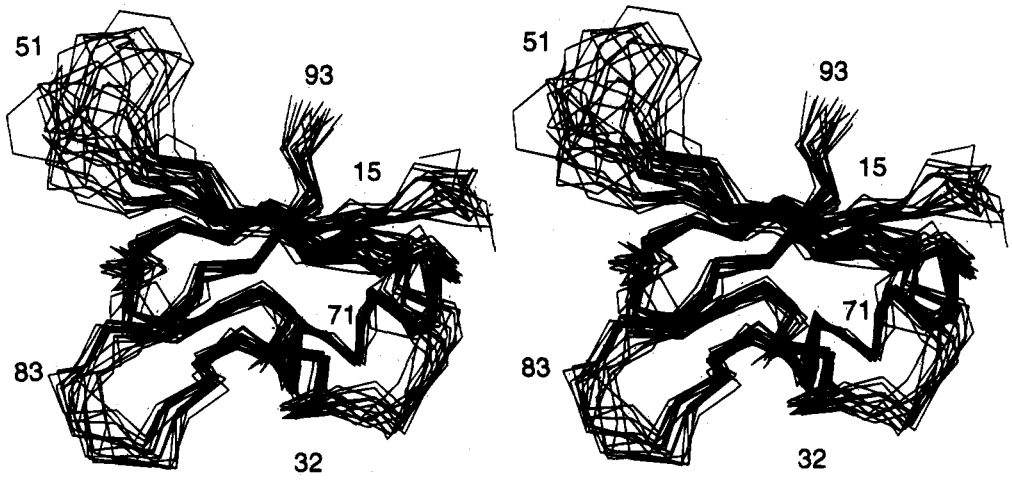

FIGURE 8: Stereoview of C_α traces for a family of 17 structures superimposed to mean backbone atoms involved in secondary structure.

conferring specificity for binding of RNA (Scherly et al., 1990; Mattaj, 1993), and it is quite likely that it plays a major role in Sxl RBD-2 RNA binding. However, the RNAs recognized by snRNP-U1A and Sxl may adopt entirely different conformations. The RNA recognized by snRNP-U1A is known to have a stem−loop structure (Jessen et al., 1991), but Sxl recognizes a site comprised of U-tracts, as do some hnRNP proteins. It is difficult to imagine these U-tracts forming continuous base pairs as is required for stem−loop structures. Therefore, it is not unreasonable to think that different RNAs are recognized by different $\beta 2-\beta 3$ loop structures. In the absence of RNA this loop appears to be mobile in solution, suggesting further structural characterization will be possible only after RNA has been complexed with the protein.

Görlach et al. (1992) successfully complexed the hnRNP-C RBD with a poly-U octamer. They observed that resonances for terminal residues of the protein which originally appeared to be unstructured in solution shifted significantly upon addition of $r(U)_8$. Moreover, deletion of only five amino-terminal residues from the hnRNP-C1 protein abolishes its RNA-binding activity (Görlach et al., 1992). The calculated structures here are consistent with the notion that the termini interact with RNA substrate. The residues entering β-strand 1 and exiting β-strand 4 are above the surface of the face of the sheet (Figures 6 and 8). Thus, it is likely that these residues will contact the RNA upon binding, and as a result they may adopt a more rigid conformation.

The β-sheet in Sxl RBD-2 has a right-handed twist as in the other RBD proteins. However, there is a bulge at the beginning of β-strand 2 (I40−V41), which results in V41 popping above the plane of the sheet. Bulges have been observed in β-strands 1 and 4 in hnRNP-C, β-strand 2 of snRNP-U1A, and not at all in hnRNP-A1. It may well be that these perturbations in local geometry contribute to the specificity of binding RNAs of differing sequences. Another interesting feature of these proteins is the relative orientation of the two helices to each other and the sheet. In this regard, snRNP-U1A appears distinct from hnRNP-C and Sxl RBD-2, where the two helices are nearly perpendicular to each other. It is possible that this is due to the extended length and C-terminal structure of helix 1 in snRNP-U1A, which leaves fewer residues to connect with the beginning of β-strand 2. Helix 1 is reported to terminate near the first conserved aromatic residue (residue preceding position A in Table 3) for both Sxl RBD-2 and hnRNP-C and near the second conserved aromatic residue (position C in Table 3) for snRNP-U1A and hnRNP-A1.

It is intriguing to compare the ^1H chemical shift data near the C-terminal region of helix 1 for the three RBDs with published chemical shifts (hnRNP-C, hnRNP-A1, Sxl RBD-2). In Table 3, A−D correspond to aligned amino acids from these three RBDs near the C-terminus of helix 1. For all three proteins, B amide protons exchange freely, while C and D amide protons are protected from exchange. In addition, all A amide protons are shifted downfield from random coil values (Wishart et al., 1991), and all B, C, and D amide protons are shifted significantly upfield. This evidence argues that there should be a common structural feature shared by these proteins, despite the inconsistencies in the reported termination sites of helix 1. For Sxl RBD-2, if restraints could be included that extend helix 1 beyond G35 and up to G38, there is little doubt that the C-terminus would move toward the sheet, reducing the interhelical angle in the process. However, presently there is not sufficient experimental evidence to suggest helix 1 extends beyond G35. This issue of helix 1 termination can be more accurately discussed when structures of higher resolution are attained.

Helix Capping. As described in the Results, a capping box α-helix initiation signal (Harper & Rose, 1993) has been identified in helix 1 (Figure 9). This structural motif enables two extra intramolecular hydrogen bonds to be formed, further stabilizing the helix. Side chain-to-main chain hydrogen bonds as observed by Harper and Rose were included in the structure calculations with T26 as the N-cap residue and Q29 as the N3 residue. Inclusion of these two hydrogen bonds was further justified from structures calculated without capping box hydrogen bond restraints, where the relative orientations of residues 26−29 were retained. In these structures, Q29 O_ϵ was the closest potential acceptor for HN of T26. Also, the only potential acceptors for HN of Q29 were T26 O_γ and the backbone carbonyl oxygen of T26, as HN of Q29 was nearly equidistant from these two atoms. As a result, the N-terminus of helix 1 is well-defined, shown in Figure 9 as a superposition of all atoms from T26 to Q29. Interestingly, HN of Q29 does not adopt a linear orientation with respect to its hydrogen bond acceptor, O_γ of T26, but instead occupies the space between O_γ of T26 and the backbone carbonyl of T26. Hence it appears from the structure calculations that the amide of Q29 is potentially involved in a bifurcated hydrogen bond. The carbonyl

Table 2: ^{15}N, ^{13}C, and ^{1}H Resonance Assignments (ppm) for Sxl RBD-2 at pH 5.0 and 25 °C[a]

residue	N	C_α	C_β	other
M1	* (*)	* (*)	* (*, *)	C_γ, *(*, *)
S2	* (*)	* (*)	* (*, *)	
Y3	* (*)	* (*)	* (*, *)	C_δ, * (*); C_ϵ, * (*)
A4	125.6 (8.09)	51.8 (4.29)	19.4 (1.30)	
R5	126.6 (8.22)	* (*)	* (*, *)	C_τ, *(*, *); C_δ, *(*, *)
P6	− (−)	* (4.45)	* (*, *)	C_γ, * (*, *); C_δ, 50.7 (3.67, 3.84)
G7	109.9 (8.63)	* (*, *)		
G8	108.6 (8.32)	* (*, *)		
E9	120.4 (8.50)	56.5 (4.34)	* (1.97, 2.11)	C_γ, 35.9 (2.32, 2.37)
S10	116.2 (8.48)	58.3 (4.48)	63.6 (3.90, 3.95)	
I11	122.6 (8.29)	63.0 (4.20)	38.2 (1.94)	$C_{\gamma1}$, 27.6 (1.20, 1.45); $C_{\gamma2}$, 17.4 (0.94); C_δ, 13.2 (0.86)
K12	123.7 (8.18)	57.4 (4.28)	32.8 (1.81, 1.81)	C_γ, 24.8 (1.43, 1.43); C_δ, 29.1 (1.66, 1.66); C_ϵ, 41.9 (3.00, 3.00)
D13	116.8 (8.29)	55.1 (4.75)	* (2.88, 2.88)	
T14	103.7 (7.61)	61.2 (4.40)	* (4.75)	C_γ, * (0.87)
N15	120.7 (7.98)	52.2 (5.36)	41.0 (2.54, 3.12)	N_δ, 109.2 (6.00, 7.43)
L16	127.1 (9.64)	53.7 (5.06)	44.1 (1.17, 2.07)	C_τ, 27.0 (1.96); $C_{\delta1}$, 24.8 (0.84); $C_{\delta2}$, 27.3 (1.11)
Y17	122.1 (9.31)	55.9 (4.79)	40.8 (2.63, 2.71)	C_δ, 132.6 (6.34); C_ϵ, 117.7 (6.37)
V18	129.8 (8.75)	60.0 (5.05)	34.3 (1.81)	$C_{\gamma1}$, 21.2 (0.77); $C_{\gamma2}$, 22.1 (1.06)
T19	114.6 (8.82)	59.3 (5.04)	71.5 (4.21)	C_τ, 22.5 (0.79)
N20	113.0 (8.33)	53.6 (4.32)	37.1 (3.17, 3.17)	
L21	117.3 (8.22)	53.7 (4.14)	41.0 (1.18, 1.23)	C_γ, 26.8 (1.46); $C_{\delta1}$, 24.1 (0.89); $C_{\delta2}$, 27.2 (0.60)
P22	− (−)	62.5 (4.54)	32.2 (1.91, 2.44)	C_τ, * (2.04, 2.04); C_δ, 49.4 (3.40, 3.82)
R23	120.1 (8.92)	57.8 (3.94)	28.9 (1.82, 2.08)	C_γ, 27.9 (1.60, 1.82); C_δ, 42.8 (3.17, 3.17); N_ϵ, 113.2 (7.28)
T24	103.5 (7.06)	60.1 (4.27)	68.5 (4.65)	C_γ, 21.5 (1.18)
I25	121.9 (7.22)	59.1 (4.35)	39.2 (2.06)	$C_{\gamma1}$, 28.5 (1.28, 1.68); $C_{\gamma2}$, 16.6 (0.88); C_δ, 14.8 (0.97)
T26	113.3 (6.91)	58.4 (4.71)	73.1 (4.67)	C_γ, 21.8 (1.35)
D27	122.2 (8.95)	58.5 (4.17)	40.6 (2.64, 2.64)	
D28	116.6 (8.37)	57.0 (4.45)	40.4 (2.60, 2.70)	
Q29	120.2 (7.77)	58.4 (4.18)	29.0 (2.06, 2.23)	C_τ, 33.8 (2.41, 2.52); N_ϵ, 109.1 (6.62, 7.45)
L30	119.6 98.31)	58.5 (4.22)	42.0 (1.57, 1.97)	C_γ, 26.6 (1.76); $C_{\delta1}$, 24.6 (0.74); $C_{\delta2}$, 26.3 (0.75)
D31	119.5 (8.16)	57.7 (4.37)	40.3 (2.75, 2.95)	
T32	117.4 (8.05)	66.6 (3.94)	68.7 (4.38)	C_γ, 21.8 (1.24)
I33	119.8 (8.12)	65.0 (3.74)	40.0 (1.75)	$C_{\gamma1}$, 28.5 (1.02, 1.95); $C_{\gamma2}$, 18.0 (0.40); C_δ, 13.5 (0.83)
F34	112.1 (8.71)	61.0 (4.46)	38.4 (3.01, 3.38)	C_δ, 132.3 (7.80); C_ϵ, 130.4 (7.01); C_ζ, *(*)
G35	114.8 (9.05)	46.6 (4.29, 4.37)		
K36	115.7 (6.70)	57.2 (3.95)	32.1 (1.08, 1.18)	C_γ, 24.1 (0.93, 0.95); C_δ, 28.8 (1.54, 1.54); C_ϵ, 41.6 (2.92, 2.92)
Y37	113.5 (6.86)	58.4 (4.41)	38.7 (2.59, 3.36)	C_δ, 132.4 (7.15); C_ϵ, 118.7 (6.77)
G38	103.5 (7.34)	44.7 (3.97, 4.23)		
S39	114.1 (8.30)	57.6 (4.57)	63.3 (3.94, 3.94)	
I40	127.5 (8.79)	61.5 (3.82)	38.2 (1.68)	$C_{\gamma1}$, * (*, *); $C_{\gamma2}$, 17.9 (0.57); C_δ, 14.2 (0.72)
V41	125.7 (9.02)	61.9 (4.21)	32.7 (1.92)	$C_{\gamma1}$, 19.7 (0.74); $C_{\gamma2}$, 20.9 (0.88)
Q42	119.0 (7.48)	55.8 (4.49)	32.5 (1.87, 1.99)	C_γ, 33.8 (2.22, 2.34); N_ϵ, 111.6 (6.79, 7.55)
K43	121.7 (8.61)	54.8 (5.08)	35.2 (1.62, 1.70)	C_γ, 23.6 (1.34, 1.40); C_δ, 29.6 (1.57, 1.57); C_ϵ, 41.1 (2.88, 3.04)
N44	118.3 (8.87)	52.8 (4.98)	41.2 (2.72, 2.87)	N_δ, 109.7 (6.68, 7.39)
I45	126.4 (8.97)	61.8 (4.03)	38.9 (1.80)	$C_{\gamma1}$, 29.8 (0.63, 1.64); $C_{\gamma2}$, 17.4 (0.83); C_δ, 13.5 (0.75)
L46	127.1 (7.69)	55.6 (4.54)	41.2 (1.59, 1.59)	C_γ, 28.4 (1.69); $C_{\delta1}$, 23.1 (0.98); $C_{\delta2}$, 25.0 (1.03)
R47	123.3 (9.09)	* (*)	33.4 (*, *)	C_γ, *(*, *); C_δ, *(*, *)
D48	123.2 (8.66)	54.3 (4.54)	42.9 (2.61, 2.97)	
K49	127.8 (8.97)	58.5 (4.04)	32.5 (1.90, *)	C_γ, 24.9 (1.51, 1.51); C_δ, 29.1 (1.73, 1.73); C_ϵ, 41.9 (*, 3.04)
L50	117.8 (8.47)	56.9 (4.39)	42.1 (1.69, 1.93)	C_γ, 27.3 (1.64); $C_{\delta1}$, 23.4 (0.91); $C_{\delta2}$, 24.4 (0.98)
T51	106.4 (8.20)	61.7 (4.45)	71.3 (4.38)	C_γ, 21.0 (1.25)
G52	110.7 (8.26)	45.6 (3.86, 4.25)		
R53	120.8 (7.88)	54.5 (4.66)	30.2 (1.71, 1.91)	C_γ, 27.2 (1.73, 1.73); C_δ, 43.2 (3.24, 3.24)
P54	−(−)	63.5 (4.45)	32.1 (1.96, 2.33)	C_γ, 27.4 (2.04, 2.08); C_δ, * (3.67, 3.85)
R55	119.4 (8.27)	56.2 (4.42)	32.0 (1.53, 2.05)	C_γ, 27.5 (1.53, 1.71); C_δ, 43.6 (3.01, 3.21)
G56	106.8 (8.64)	45.6 (3.40, 4.37)		
V57	114.7 (6.98)	58.8 (5.24)	35.4 (1.69)	$C_{\gamma1}$, 20.8 (0.85); $C_{\gamma2}$, 21.9 (0.74)
A58	126.7 (9.21)	50.7 (5.19)	23.8 (0.98)	
F59	116.1 (8.38)	56.3 (5.86)	41.9 (2.78, 3.25)	C_δ, 132.2 (7.23); C_ϵ, 131.1 (7.28); C_ζ, *(*)
V60	121.5 (8.49)	61.9 (4.13)	36.0 (1.33)	$C_{\gamma1}$, 20.3 (0.05); $C_{\gamma2}$, 20.8 (0.22)
R61	127.5 (8.85)	54.3 (5.59)	35.9 (*, *)	C_γ, * (*, *); C_δ, 44.2 (3.03, 3.03)
Y62	125.7 (8.29)	58.4 (5.28)	42.1 (2.58, 4.10)	C_δ, 132.7 (6.94); C_ϵ, 117.2 (6.64)
N63	118.0 (8.75)	55.6 (4.61)	40.3 (2.85, 3.36)	N_δ, 114.3 (6.98, 8.24)
K64	113.0 (8.27)	53.9 (4.88)	36.1 (1.60, 2.26)	C_γ, 24.9 (1.48, 1.48); C_δ, 29.1 (1.74, 1.74); C_ϵ, 41.8 (3.03, 3.03)
R65	125.0 (9.31)	59.3 (3.98)	29.8 (*, *)	C_γ, * (*, *); C_δ, 43.4 (*, 3.27); N_ϵ, 114.9 (7.51)
E66	119.5 (10.05)	60.1 (4.04)	28.5 (*, *)	C_γ, 36.7 (2.35, 2.53)
E67	119.1 (6.85)	58.1 (3.92)	29.3 (*, *)	C_γ, * (*, *)
A68	118.3 (6.72)	54.9 (3.85)	17.7 (1.59)	
Q69	113.6 (8.00)	58.4 (3.92)	28.4 (2.10, 2.10)	C_γ, 33.7 (2.45, 2.45)
E70	120.8 (7.76)	58.9 (4.01)	28.3 (2.18, 2.31)	C_γ, 35.0 (2.54, 2.54)
A71	121.9 (7.89)	55.2 (3.09)	19.0 (1.57)	
I72	116.7 (7.96)	66.4 (3.29)	38.6 (*)	$C_{\gamma1}$, 28.2 (1.67, *); $C_{\gamma2}$, 17.3 (0.92); C_δ, 14.0 (0.81)
S73	112.5 (7.80)	61.2 (4.11)	63.0 (3.90, *)	
A74	120.4 (7.88)	53.8 (4.19)	20.4 (1.10)	

Table 2 (Continued)

residue	N	C_α	C_β	other
L75	113.4 (7.87)	54.7 (4.47)	44.9 (*, *)	C_γ, * (*); $C_{\delta 1}$, 22.4 (0.66); $C_{\delta 2}$, 27.3 (0.21)
N76	116.0 (8.02)	56.2 (4.34)	38.9 (2.92, 3.03)	N_δ, 115.3 (7.03, 7.96)
N77	119.4 (8.97)	55.1 (4.39)	37.7 (2.95, 3.17)	N_δ, 112.8 (6.81, 7.62)
V78	119.8 (7.70)	61.4 (4.22)	34.3 (2.03)	$C_{\gamma 1}$, 21.1 (0.85); $C_{\gamma 2}$, 21.1 (0.85)
I79	128.4 (8.64)	58.9 (4.41)	* (1.69)	$C_{\gamma 1}$, 27.9 (0.30, 1.57); $C_{\gamma 2}$, 16.6 (0.67); C_δ, 13.2 (0.78)
P80	−(−)	62.2 (4.45)	32.1 (*)	C_γ, *(*, *); C_δ, *(*, *)
E81	121.6 (8.64)	58.2 (4.06)	28.9 (1.91, 2.01)	C_γ, 35.5 (2.30, 2.30)
G82	113.9 (8.84)	45.1 (3.73, 4.21)		
G83	107.9 (8.52)	43.6 (3.56, 4.56)		
S84	110.4 (8.99)	58.1 (4.72)	65.0 (3.79, 3.93)	
Q85	120.3 (7.83)	52.6 (4.92)	* (1.90, 2.11)	C_γ, 33.2 (2.35, 2.35); N_ϵ, 111.6 (6.79, 7.41)
P86	−(−)	62.0 (4.14)	31.3 (1.78, 1.78)	C_γ, *(*, *); C_δ, *(*, *)
L87	122.2 (8.47)	55.1 (4.39)	43.4 (1.29, 2.23)	C_γ, 21.9 (0.86); $C\delta_1$, 24.2 (0.83); $C_{\delta 2}$, 26.1 (0.85)
S88	115.6 (7.73)	56.9 (5.17)	63.1 (3.71, 3.96)	
V89	126.6 (8.82)	61.5 (4.66)	34.9 (1.93)	$C_{\gamma 1}$, 21.8 (1.02); $C_{\gamma 2}$, 24.3 (1.02)
R90	121.6 (8.25)	53.7 (4.55)	30.5 (1.89, 1.97)	C_γ, 26.8 (1.67, 1.67); C_δ, 43.3 (*, 3.23)
L91	126.1 (8.99)	56.1 (4.54)	41.7 (1.68, 1.68)	C_γ, * (*, *); $C_{\delta 1}$, 23.1 (0.82); $C_{\delta 2}$, 25.4 (0.95)
A92	121.7 (8.82)	53.2 (4.24)	19.4 (1.48)	
E93	119.5 (8.58)	55.9 (4.33)	30.2 (1.92, 2.03)	C_γ, 35.7 (2.27, 2.27)
E94	122.2 (8.39)	56.3 (4.35)	30.5 (1.88, 1.98)	C_γ, 35.8 (2.19, 2.23)
H95	119.1 (8.61)	55.2 (4.70)	29.1 (3.14, 3.28)	C_δ, 119.9 (7.27); C_ϵ, 136.4 (8.53)
G96	110.4 (8.49)			
K97	125.8 (7.90)	57.5 (4.17)	33.6 (1.70, 1.82)	C_γ, 24.8 (1.37, 1.37); C_δ, 29.1 (1.64, 1.64); C_ϵ, 41.8 (2.98, 2.98)

[a] In each column, ^{15}N or ^{13}C shift precedes the corresponding ^1H shift given in parentheses. An asterisk indicates that the chemical shift was not determined. Methyl group resonances for Val and Leu residues have not been stereospecifically assigned.

Table 3: Chemical Shift Values Near the C-Terminus of Helix 1 for Three RBD-Containing Proteins[a]

	position[b]/NH exchange[c]			
	A/mixed	B/fast	C/slow	D/slow
hnRNP-C	^{38}S (8.95)	^{39}K (7.10)	^{40}Y (7.01)	^{41}G (7.47)
hnRNP-A1	^{34}E (9.30)	^{35}Q (6.98)	^{36}W (7.05)	^{37}G (7.34)
Sxl RBD-2	^{35}G (9.05)	^{36}K (6.70)	^{37}Y (6.86)	^{38}G (7.34)

[a] ^1H chemical shifts in ppm are given in parentheses. [b] Positions A−D are a result from aligning residues near the C-terminus of helix 1 for the three proteins. [c] Fast and slow refer to rapidly and slowly exchanging amide protons, respectively, observed for the corresponding position in all three RBDs. Mixed refers to both fast and slow exchange rates observed. Although all authors have categorized amide exchange rates into these two regimes, their respective criteria and measurements differ slightly.

oxygen of T26 was included as a hydrogen bond acceptor for the amide proton of L30 (a position in the middle of the α-helix), and this donor/acceptor pair adopts a linear orientation. It should be noted that NOE and torsional restraints alone do not exclude the possibility of a 3_{10}-helix for the first turn of helix 1. However, these data are sufficient such that the accompanying shift in backbone geometry would not change the side chain positions of the capping box.

As discussed above, there is some uncertainty as to where helix 1 is terminated. In the light of a recent survey (Aurora et al., 1994), it is tempting to speculate that G35 and/or G38 might be involved in a C-terminal glycine helix-capping interaction. Unfortunately, the sequence LDTIFGKYGSI (residues 30−40) is not consistent with the rules provided, although the amide protection data is suggestive of the Schellman motif (Schellman, 1980). It is probable that there is a distinct pattern of hydrogen bonds since residues 37 and 38 (C and D in Table 3) have slowly exchanging amide protons. Unfortunately, no NOEs could be identified to support this type of structural motif. Again, higher resolution structures should be able to resolve this matter.

Protein−RNA Recognition. Ultimately, we are interested in understanding how RBD proteins specifically interact with their target RNAs. In this paper, the solution structure of Sxl RBD-2 is reported. Preliminary band-shift experiments (our unpublished results) indicate that both RBDs of Sxl are required for high affinity, site-specific RNA binding. It will

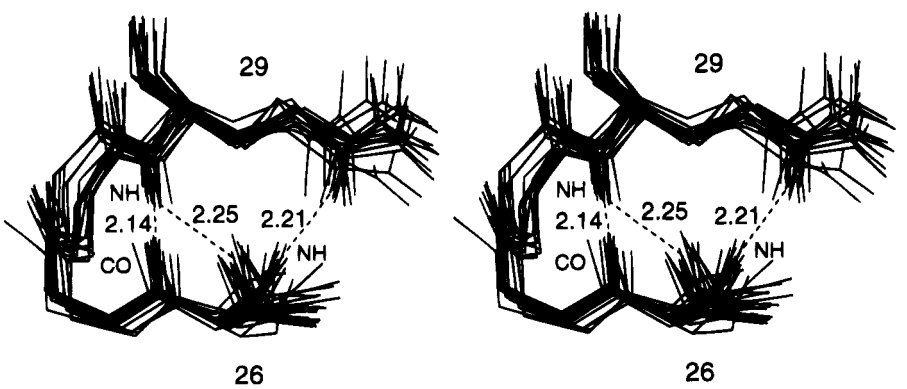

FIGURE 9: Stereoview of the "capping box". Backbone heavy atoms for T26−Q29, as well as side chains for T26 and Q29, are shown in a superposition of 17 calculated structures. All atoms for T26−Q29 were used in the superposition to mean coordinates. Labels for residues 26 and 29 occur near their respective side chains. Backbone amide protons are denoted by NH for T26 and Q29, and the backbone carbonyl oxygen of T26 is denoted by CO. Average distances in (Å) are shown (left to right) as dashed lines for $Q29_{HN}$−$T26_{O'}$, $Q29_{HN}$−$T26_{O\gamma}$, and $T26_{HN}$−$Q29_{O\epsilon}$. Standard deviations for these distances (left to right) are 0.09, 0.08, and 0.10 Å.

be interesting to further characterize the nature of this protein−RNA interaction. Because RBDs are frequently found together in groups of two or three (Keene & Query, 1991), it is hoped that a more general understanding of protein−RNA interactions will be attained through structural studies on these complexes. At the present it is unclear whether RBD-2 alone will bind RNA tightly enough to allow characterization of a complex in the slow-exchange regime. This will be the focus of future studies.

ACKNOWLEDGMENT

We thank Dr. Jeff Pelton and Fred Damberger for many interesting discussions about NMR experiments and structure calculations. We also thank Dr. Valerie Copiè for initial work on the project, Dr. David King for performing electrospray-ionization mass spectrometry, Brian Volkman for a critical reading of the manuscript, and Biosym for β-test versions for the program Felix.

REFERENCES

Aurora, R., Srinivasan, R., & Rose, G. D. (1994) *Science 264*, 1126−1130.
Baker, B. S. (1989) *Nature 340*, 521−524.
Barhuijsen, H., de Beer, R., Bouee, W. M. M. J., & van Ormondt, D. (1985) *J. Magn. Reson. 61*, 465−480.
Bax, A., & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.
Bax, A., Clore, G. M., & Gronenborn, A. M. (1990) *J. Magn. Reson. 88*, 425−431.
Bell, L. R., Maine, E. M., Schedl, P., & Cline, T. W. (1988) *Cell 55*, 1037−1046.
Biosym Technologies (1993) in *Felix User Guide Version 2.3*, San Diego, CA.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.
Brünger, A. T. (1992) in *X-PLOR Manual Version 3.1*, Yale University Press, New Haven, CT.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.
Cline, T. W. (1979) *Dev. Biol. 72*, 266−275.
Cline, T. W. (1993) *Trends Genet. 9*, 385−390.
Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *EMBO J. 5*, 2729−2735.
Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 8172−8184.
Clore, G. M., Kay, L. E., Bax, A., & Gronenborn, A. M. (1991) *Biochemistry 30*, 12−18.
Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542−3556.
Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M., & Gronenborn, A. M. (1994) *Biochemistry 33*, 2852−2858.
Ghetti, A., Bolognesi, M., Cobianchi, F., & Morandi, C. (1990) *FEBS Lett. 277*, 272−276.
Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., & Dreyfuss, G. (1992) *EMBO J. 11*, 3289−3295.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Magn. Reson. 110*, 7870−7820.
Gronenborn, A. M., & Clore, G. M. (1994) *J. Biomol. NMR 4*, 455−458.
Grzesiek, S., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291−6293.
Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605−7609.
Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., & Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495−2499.
Ikura, M., Spera, S., Barbato, G., Kay, L. E., Krinks, M., & Bax, A. (1991) *Biochemistry 30*, 9216−9228.
Inoue, K., Hoshijima, K., Sakamoto, H., & Shimura, Y. (1990) *Nature* 344, 461−463.
Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.
Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.
Keene, J. D., & Query, C. C. (1991) *Prog. Nucleic Acid Res. 41*, 179−202.
Kenan, D. J., Query, C. C., & Keene, J. D. (1991) *Trends Biochem. Sci. 16*, 214−220.
Keyes, L. N., Cline, T. W., & Schedl, P. (1992) *Cell 68*, 933−943.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1934−1941.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515−1517.
Mattaj, I. W. (1993) *Cell 73*, 837−840.
Merrill, B. M., Stone, K. L., Cobianch, F., Wilson, S. H., & Williams, K. R. (1988) *J. Biol. Chem. 263*, 3307−3313.
Nagai, K., Oubridge, C., Jessen, T., Li, J., & Evans, P. R. (1990) *Nature 348*, 515−520.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317−324.
Parkhurst, S. M., & Meneely, P. M. (1994) *Science 264*, 924−932.
Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y., & Shimura, Y. (1992) *Nucleic Acids Res. 20*, 5533−5540.
Samuels, M. L., Weber, M. J., Bishop, J. M., & McMahon, M. (1993) *Mol. Cell. Biol. 13*, 6241−6252.
Schellman, C. (1980) in *Protein Folding* (Jaenicke, R., Ed.) pp 53−61, Elsevier/North-Holland, New York.
Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., & Mattaj, I. W. (1990) *Nature 345*, 502−506.
Sosnowski, B. A., Belote, J. M., & McKeown, M. (1989) *Cell 58*, 449−459.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490−5492.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60−89.
Valcarcel, J., Singh, R., Zamore, P. D., & Green, M. R. (1993) *Nature 362*, 171−175.
Vuister, G. W., & Bax, A. (1992) *J. Magn. Reson. 98*, 428−435.
Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295−315.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G., & Mueller, L. (1992) *Biochemistry 31*, 6254−6265.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
Zhou, H. X., Lyu, P. C., Wemmer, D. E., & Kallenbach, N. R. (1994) *Proteins 18*, 1−7.