^1H, ^{13}C, and ^{15}N NMR Assignments and Global Folding Pattern of the RNA-Binding Domain of the Human hnRNP C Proteins†

Michael Wittekind,*,‡ Matthias Görlach,§ Mark Friedrichs,‡ Gideon Dreyfuss,§ and Luciano Mueller‡

Macromolecular NMR Department, The Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, New Jersey 08543-4000, and the Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148

Received January 30, 1992; Revised Manuscript Received April 16, 1992

ABSTRACT: The hnRNP C1 and C2 proteins are abundant nuclear proteins that bind avidly to heterogeneous nuclear RNAs (hnRNAs) and appear to be involved with pre-mRNA processing. The RNA-binding activity of the hnRNP C proteins is contained in the amino-terminal 94 amino acid RNA-binding domain (RBD) that is identical for these two proteins. We have obtained the ^1H, ^{13}C, and ^{15}N NMR assignments for the RBD of the human hnRNP C proteins. The assignment process was facilitated by extensive utilization of three- and four-dimensional heteronuclear-edited spectra. Sequential assignments of the backbone resonances were made using a combination of ^{15}N-edited 3D NOESY-HMQC, 3D TOCSY-HMQC, and 3D TOCSY-NOESY-HSQC as well as 3D HNCA, HNCO, and HCACO spectra. Side-chain resonances were assigned using 3D HCCH-COSY and 3D HCH-TOCSY spectra. Four-dimensional $^{13}C/^{13}C$-edited NOESY and $^{13}C/^{15}N$-edited NOESY experiments were used to unambigously resolve NOEs. The overall global folding pattern was established by calculating a set of preliminary structures using constraints derived from the sequential NOEs and a small number of long-range NOEs. The $\beta\alpha\beta$-$\beta\alpha\beta$ domain structure exhibits an antiparallel β-sheet with the conserved RNP 1 and RNP 2 sequences [Dreyfuss et al. (1988) *Trends Biochem. Sci. 13*, 86-91] located adjacent to one another as the two inner strands of the β-sheet.

Heterogeneous nuclear RNAs (hnRNAs) form complexes (hnRNP complexes) with a specific set of abundant nuclear proteins termed hnRNP proteins [for reviews see Dreyfuss (1986), Chung and Wooley (1986), Dreyfuss et al. (1988), and Bandziulis et al. (1989)]. There are approximately 20 major hnRNP proteins, designated A through U, that range in molecular weight from 34 000 to 120 000 (Piñol-Roma et al., 1988). Although the specific functions of hnRNP proteins have not been elucidated in detail, several of their properties indicate that they have a role in pre-mRNA metabolism. Most, if not all, have direct RNA-binding activity (Cobianchi et al., 1988; Dreyfuss et al., 1988; Merril et al., 1988; Piñol-Roma et al., 1988; Swanson & Dreyfuss, 1988a; Piñol-Roma et al., 1989; Matunis et al., 1992a), and they associate with nascent hnRNA transcripts (Economidis & Pederson, 1983; Fakan et al., 1986; Piñol-Roma et al., 1989; Matunis et al., 1992; Matunis et al., 1992b), indicating a function in early posttranscriptional events.

The hnRNP C1 and C2 proteins are among the most avid pre-mRNA-binding proteins (Piñol-Roma et al., 1988; Swanson & Dreyfuss, 1988a,b). Antibody inhibition, immunodepletion, and immunoprecipitation experiments suggested that they are involved in pre-mRNA splicing (Choi et al., 1986; Swanson & Dreyfuss, 1988b). Molecular cloning of the hnRNP C1 and C2 proteins (Nakagawa et al., 1986; Swanson et al., 1987; Burd et al., 1989) revealed that they belong to a family of RNA-binding proteins which share a highly homologous domain of 90-100 amino acids (Dreyfuss et al., 1988; Bandziulis et al., 1989; Mattaj, 1989; Query et al., 1989; Kenan et al., 1991), initially described in the yeast poly(A)-binding protein (Adam et al., 1986; Sachs & Kornberg, 1986) and in the mammalian hnRNP A1 protein (Kumar et al., 1986; Cobianchi et al., 1986). This RNA-binding domain (RBD)[1] was proposed (Dreyfuss et al., 1988; Bandziulis, 1989) to confer RNA-binding activity to this class RNA-binding proteins. Studies on various proteins of this class supported this prediction (Merrill et al., 1988; Query et al., 1989; Scherly et al., 1989, 1990; Lutz-Freyermuth et al., 1990; Nietfeld et al., 1990; Burd et al., 1991; Jessen et al., 1991). Some of the RNA-binding proteins of this class contain multiple (up to four) RBDs having different RNA-binding activities (Nietfeld et al., 1990; Burd et al., 1991). The canonical RBD contains two short stretches of amino acids which are conserved to a higher degree than the remainder of the RBD; the most highly conserved is the eight amino acid RNP 1 or RNP consensus sequence (RNP-CS; Bandziulis et al., 1989), and the somewhat less conserved is the six amino acid RNP 2 sequence located approximately 30 amino acids amino-terminal to RNP1 (Dreyfuss, 1988). In contrast to these highly conserved RNP1 and RNP2 sequences within a typical RBD, RNP 1 is preceded by a stretch of amino acids of highly variable length and sequence which are the least conserved residues among the RBDs and which play a role in determining the RNA-binding specificity of this domain (Scherly et al., 1990). In the RBD of the hnRNP C1 and C2 proteins, which extends from amino acid 1 through 94, designated K94 (M.G., C. Burd, and G.D., manuscript in preparation), this variable

† This work was supported by the Howard Hughes Medical Institute and by grants from the NIH.

* To whom correspondence should be addressed.
‡ Bristol-Myers Squibb Pharmaceutical Research Institute.
§ Howard Hughes Medical Institute.

[1] Abbreviations: RBD, RNA-binding domain; FID, free induction decay; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; TOCSY, total correlation spectroscopy; τ_m, mixing time; DSS, 3-(trimethylsilyl)-1-propanesulfonic acid; TSP, (trimethylsilyl)[2,2,3,3-$^2H_4$]propionate; 2D, two dimensional; 3D, three dimensional; 4D, four dimensional.

0006-2960/92/0431-6254$03.00/0 © 1992 American Chemical Society

segment is reduced to only a few residues (Swanson et al., 1987; Burd et al., 1989; Bandziulis et al., 1989; Kenan et al., 1991). Their status as major hnRNP proteins and their well-defined, single and small RBD makes the hnRNP C proteins a prototype of this class of RNA-binding proteins. To better understand the structural basis of the RNA-binding activity and specificity of the hnRNP C proteins (Swanson et al., 1988a,b), we have produced K94 in *Escherichia coli* and purified it to homogeneity in active, RNA-binding form. Nuclear magnetic resonance (NMR) techniques allow for the determination of the structure of polypeptides and of protein–ligand complexes in solution (Wüthrich, 1986; Clore & Gronenborn, 1991; Otting et al., 1990). Here we describe the ^1H, ^{15}N, and ^{13}C NMR assignments of K94 and the global folding pattern of this RNA-binding domain of the hnRNP C proteins. The overall folding pattern is the same as that recently described for the U1A protein (Nagai et al., 1990; Hoffman et al., 1991), but several differences that are important for understanding the significance of this structure have been noted and are discussed.

Experimental Procedures

Sample Preparation. The amino-terminal 94 amino acids of the human hnRNP C proteins, designated K94, encompassing the RNA-binding domain of the hnRNP C proteins (M.G., C. Burd, and G.D., manuscript in preparation), were overexpressed in *E. coli* BL21(DE3) from a cDNA fragment of the hnRNP C1 protein (Swanson et al., 1987) cloned into pET3d (Novagen Inc., Madison, WI; Studier et al., 1990). Labeling of the hnRNP C RBD was achieved by growing the cells in minimal media (Sambrook et al., 1989) with 1 g/L $[^{15}N]NH_4Cl$ (Isotech Inc., Miamisburg, OH) as the sole nitrogen source and, for double labeling, with 2.5 g/L uniformly labeled [^{13}C]glucose (Cambridge Isotope Laboratories, Woburn, MA) as the sole carbon source. The K94 was purified by chromatography over DEAE Sephacell (Pharmacia LKB, Piscataway, NJ), ssDNA cellulose (USB, Cleveland, OH), and S-Sepharose (Pharmacia LKB, Piscataway, NJ) and reconcentrated by $(NH_4)_2SO_4$ (Sigma Chemical Company, St. Louis, MO) precipitation as will be described elsewhere.

The purified K94 is homogeneous as determined by 2D polyacrylamide electrophoresis (O'Farrel, 1977). The purified K94 was dissolved in 50 mM NH_4OAc, pH 6.8, at a concentration of 5.4 mg/mL, and electrospray mass spectra (Covey et al., 1988) were recorded on a Sclex API-III triple quadrupole mass spectrometer fitted with a standard pneumatically assisted nebulization probe and an atmospheric pressure ionization source (Sclex, Ontario, Canada). The determined average molecular mass of K94 is 10053.5 Da; the calculated mass for K94 lacking its amino-terminal methionine residue is 10053.6. The lack of the N-terminal methionine residue was shown by amino-terminal sequencing performed by automated Edman degradation on an Applied Biosystems 473A Sequencer with an online PTH amino acid analyzer as recommended by the manufacturer (Applied Biosystems, Forster City, CA). For NMR spectroscopy, samples were exchanged into 50 mM deuterated sodium acetate, 100 μM EDTA, 0.02% NaN_3, and 9% ^2H_2O/91% ^1H_2O or 100% ^2H_2O, at a pH/pD of 5.5 at 20 °C. The protein concentration of the samples was approximately 3 mM.

NMR Spectroscopy. All spectra were collected at 20 °C and were obtained from the sample prepared in ^1H_2O unless otherwise noted. All proton-detected spectra were obtained on a Varian UNITY-600 spectrometer equipped with triple-channel capabilities and a ^1H–^{13}C–^{15}N triple-tuned probe. Multidimensional spectra were obtained using spectral widths of 6250, 3000, 2200, and 2000 Hz in the ^1H, $^{13}C\alpha$, ^{13}C carbonyl, and ^{15}N dimensions, respectively. The ^1H carrier was set at the water resonance (4.89 ppm), and presaturation was used for solvent suppression followed, in some cases, by the SCUBA sequence (Brown et al., 1988) for samples in ^1H_2O solution. Carrier frequencies for the heteronuclei were set at 121 ppm for ^{15}N, 41.2 ppm for ^{13}C (aliphatic), 58 ppm for ^{13}C (α), and 177 ppm for ^{13}C (carbonyl). Chemical shifts for ^1H and ^{13}C are referenced to TSP and DSS, respectively. Quadrature detection for nonacquisition dimensions in all multidimensional spectra were obtained using the States-TPPI type acquisition scheme (Marion et al., 1989a) unless otherwise noted. Heteronuclear decoupling during proton acquisition was achieved using WALTZ16 (Shaka et al., 1983) for ^{15}N or GARP1 (Shaka et al., 1985) for ^{13}C.

Spectra were processed using a version of FELIX (Hare Research, Bothell, WA) modified to incorporate complex linear prediction (M. Friedrichs, unpublished; L. Kay, personal communication). In some cases, "mirror image" linear prediction was performed after extending the signal into the negative time domain (Zhu & Bax, 1990). For spectra recorded with samples in ^1H_2O, low-freqeuncy deconvolution (Marion et al., 1989b) was applied to the acquisition FIDs prior to transformation in order to reduce the size of the residual ^1H_2O signal. For experiments where only amide protons were detected in the acquisition dimension, the right half of the spectrum (upfield of the ^1H_2O signal) was discarded.

2D HSQC ^{15}N–^1H correlation spectra (Bodenhausen & Ruben, 1980) were collected with an acquisition time of 40 ms using eight scans per t_1 increment and hypercomplex acquisition (Mueller & Ernst, 1978; States et al., 1982) in the F_1 (^{15}N) dimension. To measure ^1H–^2H exchange of amide protons, 2D ^{15}N–^1H correlated HSQC spectra were recorded 6 h after initiating exchange of the sample from buffered ^1H_2O solution into buffered ^2H_2O solution using Centricon-3 (Amicon, Danvers, MA).

3D ^{15}N-edited TOCSY-HMQC, NOESY-HMQC (Marion et al., 1989c; Zuiderweg & Fesik, 1989), and TOCSY-NOESY-HSQC (Mueller et al., 1992) spectra were obtained using TPPI-type acquisitions (Marion & Wüthrich, 1983) for the F_1 (^1H) and F_2 (^{15}N) dimensions. Two 3D TOCSY-HMQC spectra were obtained with τ_m equal to 32 and 50 ms using MLEV17 (Bax & Davis, 1985) during the TOCSY transfers. The 3D NOESY-HMQC spectrum was acquired with a 140-ms τ_m. The 3D TOCSY-NOESY-HSQC spectrum was acquired using 32 and 140 ms for the TOCSY τ_m and NOESY τ_m, respectively. The data were comprised of 256 real × 56 real × 512 complex data points, and four scans were used to acquire each FID. The data were processed into 512 × 64 × 512 real point matrices using a 45°-shifted skewed sine-bell window function in F_3 and 90°-shifted sine-bells in F_1 and F_2 followed by zero-filling in all dimensions.

The 3D HNCA spectrum (Kay et al., 1990a) was collected using a modified pulse sequence (Farmer et al., 1992). Carbonyls were decoupled during t_1 with a shifted and truncated hermite pulse. Each F_1 (^{13}C)–F_2 (^{15}N) point was collected with 32 scans. Acquisition times of 8.7 ms (26 complex points) and 14 ms (28 complex points) were collected for F_1 and F_2, respectively, with the acquisition dimension collected into 832 complex points. Complex linear prediction was used to extend the F_1 and F_2 FIDs to 56 and 40 points, respectively. The data were processed into a 128 × 64 × 1024 real point matrix using matched cosine bell window functions in each dimension.

The 3D HNCO spectrum (Kay et al., 1990a) was acquired using the same pulse sequence as for the HNCA experiment. The carbon carrier frequencies were switched and the delay time for creating ^{13}C–^{15}N double-quantum coherence set to $1/(2J_{N\text{-}CO})$. The spectrum was acquired with acquisition times of 13.3 ms (40 complex points) and 8 ms (16 complex points) for F_1 and F_2, respectively, with the F_3 dimension collected into 1024 complex points. Mirror imaged linear prediction was used to extend the F_2 dimension by 32 points. The final data matrix was 256 × 64 × 256 real points.

The HCACO experiment was collected using the sample in ^2H_2O utilizing a pulse sequence featuring constant time evolution in the F_1 dimension (Powers et al., 1991). The data were collected with 16 scans with acquisition times of 6.7 ms (20 complex points), 24.5 ms (54 complex points), and 164 ms (2048 complex points) for F_1 ($^{13}C\alpha$), F_2 (^{13}C carbonyl), and F_3 (^1H), respectively. Mirror image linear prediction was used to extend the F_1 dimension to 40 complex points, and the final data matrix was 64 × 128 × 256 real points.

The HCCH-COSY and HCCH-TOCSY spectra were obtained from the doubly labeled sample in ^2H_2O using the published pulse sequences (Bax et al., 1990) modified to allow eight-step phase cycling per F_1 (^1H)–F_2 (^{13}C) point (Mueller, unpublished). Carbonyl decoupling during t_2 was performed with a frequency shifted hermite pulse. The DIPSI-3 isotropic mixing sequence (Shaka et al., 1988) with a mixing time of 24 ms was used for the HCCH-TOCSY experiment. Acquisition times of 16 ms (100 complex points) and 6.7 ms (16 complex points) were collected for F_1 (^1H) and F_2 (^{13}C), respectively, with the acquisition dimension collected into 1024 complex points. The data were processed into a 256 × 64 × 256 real data matrix. Linear prediction to 128 and 32 complex points for the F_1 and F_2 dimensions, respectively, was performed prior to transformation.

The 4D $^{13}C/^{15}N$-edited NOESY spectrum was obtained using the published pulse sequence (Kay et al., 1990b) with a mixing time of 100 ms. Acquisition times of 2.7 ms (8 complex points), 16 ms (100 complex points), and 4.0 ms (8 complex points) were collected for F_1 (^{13}C), F_2 (^1H), and F_3 (^{15}N), respectively, with the acquisition dimension collected into 512 complex points. The data were processed into a 32 × 256 × 32 × 128 real data matrix. Prediction of an additional 12 complex points for the F_1 and F_3 dimensions was performed prior to transformation by using linear prediction after extending the signal into the negative time domain (Zhu & Bax, 1990). For this experiment, the data collection in F_1 was delayed by half of the dwell time.

The 4D $^{13}C/^{13}C$-edited NOESY spectrum was obtained using the published pulse sequence (Clore et al., 1991) with several modifications (Mueller, unpublished). The mixing time was 120 ms. Data were collected with acquisition times of 2.7 ms (8 complex points), 10.2 ms (64 complex points), and 2.7 ms (8 complex points) for F_1 (^{13}C), F_2 (^1H), and F_3 (^{13}C), respectively, and the acquisition dimension collected into 512 complex points. The data were processed into a 32 × 128 × 32 × 256 real data matrix. Linear prediction (Zhu & Bax, 1990) of an additional 12 complex points for the F_1 and F_3 dimensions was performed prior to transformation.

Structure Calculations. NOE intensities were evaluated by counting contour levels of peaks in printed contour diagrams of the NOESY spectra. The NOE intensities were classified as being either strong, medium, or weak and converted to distance constraints by assigning upper bounds of 2.7, 3.5, and 5.0 $Å_3$, respectively. The lower bounds for all cases was 1.8 Å. These bounds were augmented by explicit hydrogen-bond constraints in cases where amide protons were protected from solvent exchange, and the identity of the hydrogen-bond acceptor for a given protected amide proton was clear on the basis of NOE data. Appropriate cross-strand NOEs or $d\alpha N_{(i+4)}$ NOEs had to be present in order for hydrogen bonds to be entered for residues involved in antiparallel β-sheets or α-helices, respectively. Each hydrogen bond was represented by two constraints: the amide proton to carbonyl oxygen (1.8–2.20 Å) and the amide nitrogen to carbonyl oxygen (2.8–3.2 Å). The preliminary bounds file consisted of constraints derived from 313 sequential, 91 medium-range, and 86 long-range NOEs and 30 hydrogen bonds. Only 11 constraints derived from intraresidue NOEs were included.

The variable target function minimization program DIANA (Güntert, 1990; Güntert et al., 1991) was used to generate 200 structures. Contributions to the target function from distance, dihedral angle, and van der Waals violations were weighted in a ratio of 1:10:0.2 for 4150 iterations, while the residue window size was gradually increased until all residues were included. The van der Waals contribution was then increased first to 0.6 (500 iterations) and then to 1.0 (500 iterations). The 11 best structures, as evaluated by residual NOE constraint violations, all exhibited the same overall folding topology and were further refined using DIANA for another 6250 iterations. They were then subjected to a dynamical simulated annealing protocol (Nilges et al., 1988) using X-PLOR (Brünger, 1990). The refinement protocol used was essentially the same as described for refinement of BDS-I (Driscoll et al., 1989). All of the 11 structures maintained the same global folding pattern, although each had a number of residual constraint violations. In most cases, these violations could be attributed to the incorrect interpretation of spin diffusion NOE peaks as primary NOEs as a result of the qualitative treatment of cross-peak volumes obtained from NOESY spectra collected with relatively long mixing times. A more detailed analysis including NOESY data collected with shorter mixing times is in progress. However, as the goal at this stage was to establish the global folding pattern, no attempt was made to refine these structures further.

RESULTS

^1H–^{15}N Inverse Experiments. The 2D HSQC ^{15}N–^1H correlation spectrum of uniformly ^{15}N-labeled K94 and the cross-peak assignments that were ultimately arrived at are shown in Figure 1. Both dimensions exhibit adequate chemical shift dispersion providing for good overall resolution. A number of peaks with narrower ^1H line widths were observed and were determined to correspond to residues at the N (residues 3–13) and C (residues 90–94) termini, indicating that these regions are more mobile than the rest of the isolated domain.

A series of ^1H–^{15}N correlated 3D experiments including 3D NOESY-HMQC, 3D TOCSY-HMQC, and 3D TOCSY-NOESY-HSQC were obtained and analyzed in order to identify spin systems and to determine sequential connectivities. An example of the 3D NOESY-HMQC spectra are shown in Figure 2, where the tracing of the *d*NN connectivities for an α-helical segment is shown.

To establish connectivities between residues in extended secondary structures, the 3D TOCSY-NOESY HSQC experiment proved particularly useful. This is illustrated in Figure 3, where connectivities between Ile_{75} and Ala_{76} are shown. The $d\alpha_{i-1}N_i$ NOESY connectivity is strong for this type of connectivity (Figure 3G). However, it is often difficult to identify the spin system corresponding to the $i-1$ residue on the basis of the $d\alpha_{i-1}N_i$ connectivity alone as a number of

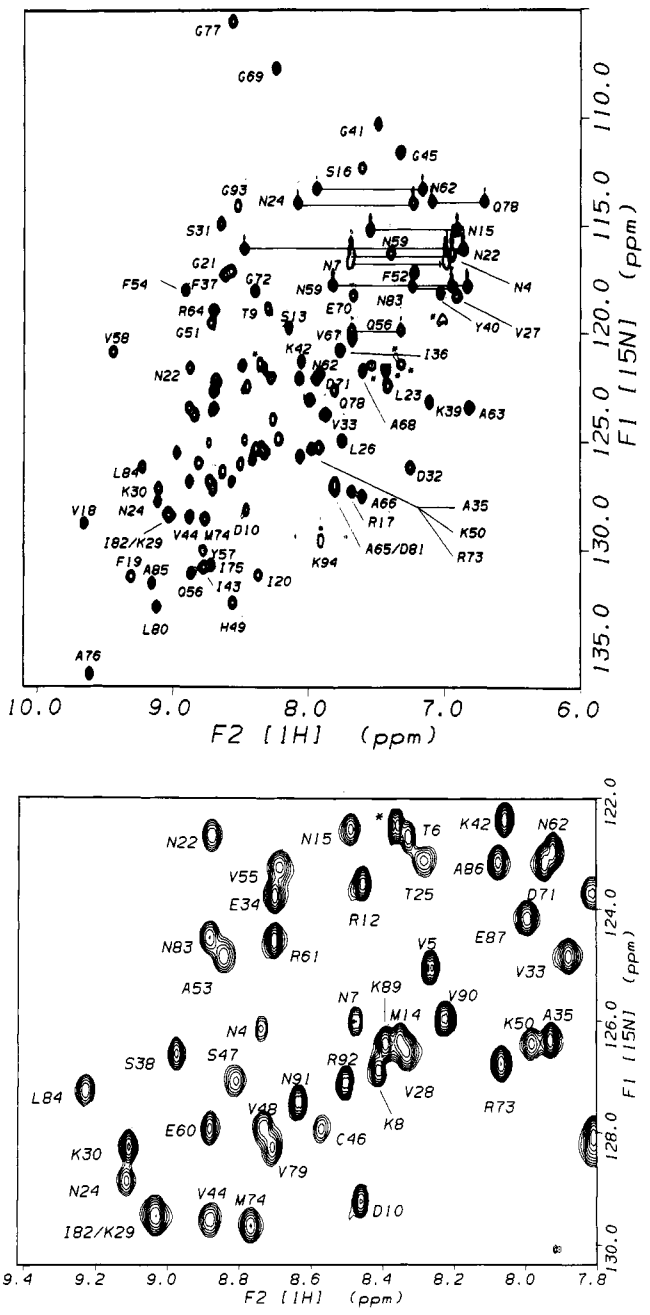

FIGURE 1: 2D HSQC ^{15}N–^{1}H correlation spectrum of uniformly ^{15}N-labeled hnRNP C RDB and assignments. The cross-peaks are labeled with the one-letter amino acid code and residue number to indicate the assignments. Horizontal lines connect cross-peaks corresponding to the NH_2 groups of asparagine and glutamine side chains. The asterisks indicate folded peaks from arginine side chains. (A, top) Full spectrum. (B, bottom) Expanded region.

spin systems may have an α-proton at the same resonance position. The 3D TOCSY-NOESY spectrum (Mueller et al., 1992) provides additional information by including $d\mathrm{N}_{i-1}\mathrm{N}_i$ cross-peaks (Figure 3H). In most cases, the combination of the α and amide proton resonance positions for residue $i-1$ is sufficiently unique to identify this residue. Additional $i-1$ to i connectivities are often observed, such as the $d\beta_{i-1}\mathrm{N}_i$ connectivity from Ile_{75} to Ala_{76} seen in Figure 3 (Figure 3D).

Spin systems were identified and then linked using $d\mathrm{N}_{i-1}\mathrm{N}_i$, $d\alpha_{i-1}\mathrm{N}_i$, and $d\beta_{i-1}\mathrm{N}_i$ NOESY cross-peaks for segments of α-helical secondary structure. The $d\alpha_{i-1}\mathrm{N}_i$ and $d\beta_{i-1}\mathrm{N}_i$ NOESY and $d\mathrm{N}_{i-1}\mathrm{N}_i$ TOCSY-NOESY cross-peaks were used to link residues making up segments of extended secondary structure. In this manner, sequential assignments were derived

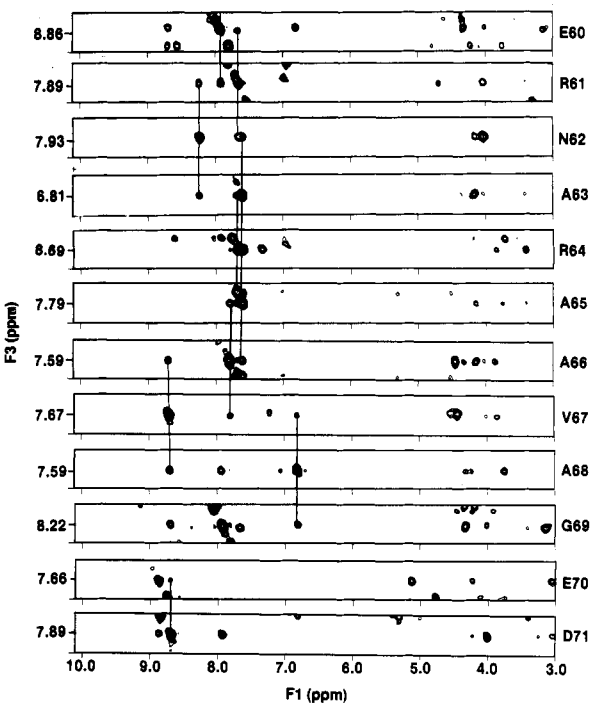

FIGURE 2: Portions of the ^{15}N-edited 3D NOESY-HMQC spectrum illustrating dNN connectivities among residues of helix α_2. Each panel represents a F_1 (^{1}H)–F_3 (^{1}H) plane taken at the F_2 (^{15}N) frequency of the backbone nitrogen atom of the residue indicated at the right of each panel. The lines connect amide proton autopeaks of residue i with $d\mathrm{N}_i\mathrm{N}_{i-1}$ and $d\mathrm{N}_i\mathrm{N}_{i+1}$ cross-peaks.

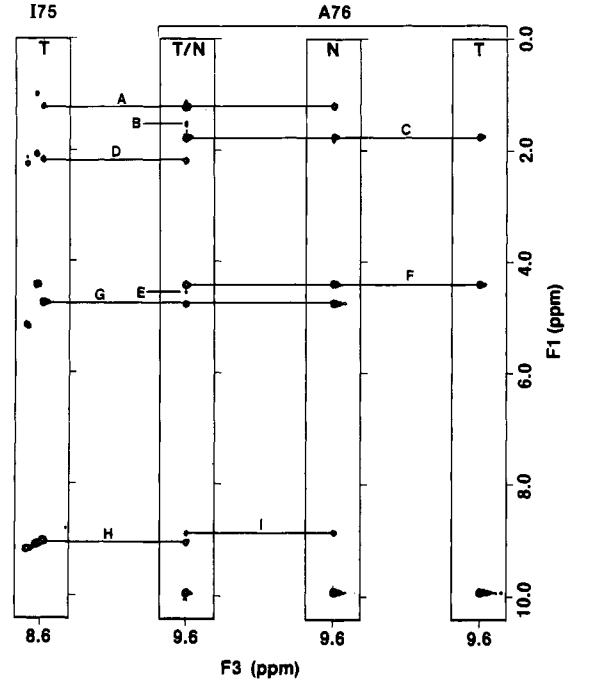

FIGURE 3: Portions of the ^{15}N-edited 3D spectra illustrating the sequential connectivities between Ile_{75} and Ala_{76}. The F_2 positions at which the F_1–F_3 planes are plotted are indicated by the "I_{75}" (132.1 ppm) and "A_{76}" (137.1 ppm) notations above the panels. The type of ^{15}N-edited 3D spectrum is indicated by T (HMQC-TOCSY), N (HMQC-NOESY), and T/N (HSQC-TOCSY-NOESY) as the top of each panel. The horizontal lines connect cross-peaks having the same F_1 ppm values. The letters A–I indicates the following $^{1}H/^{1}H$ connectivities involving residues Ile_{75} ($i-1$), Ala_{76} (i), and Gly_{77} ($i+1$): A, $d\gamma_{m(i-1)}\mathrm{N}_{(i)}$; B, $d\gamma_{(i-1)}\mathrm{N}_{(i)}$; C, $d\beta_{m(i)}\mathrm{N}_{(i)}$; D, $d\beta_{m(i-1)}\mathrm{N}_{(i)}$; E, $d\alpha_{(i+1)}\mathrm{N}_{(i)}$; F, $d\alpha_{(i)}\mathrm{N}_{(i)}$; G, $d\alpha_{(i-1)}\mathrm{N}_{(i)}$; H, $d\mathrm{N}_{(i-1)}\mathrm{N}_{(i)}$; I, $d\mathrm{N}_{(i)}\mathrm{N}_{(i+1)}$.

for three contiguous segments of the polypeptide. The break points between the segments represented the two proline residues in hnRNP C RDB. Side-chain assignments for easily

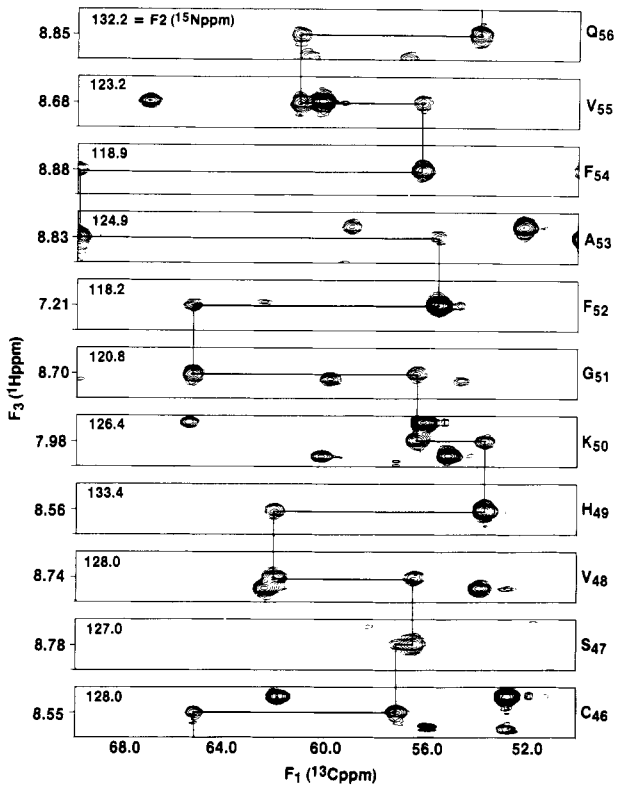

FIGURE 4: Portions of the 3D HNCA spectrum of doubly ^{13}C–^{15}N-labeled hnRNP C RBD. Each panel shows a F_1 (^{13}C)–F_3 (^1H) plane taken at the F_2 (^{15}N) frequency indicated at the top-left corner of each panel. The lines connect each intraresidue $^1H_{HNi}/^{13}C\alpha_i$ cross-peak with the sequential $^1H_{HNi}/^{13}C\alpha_{i-1}$ and $^1H_{HNi+1}/^{13}C\alpha_i$ cross-peaks. The residues giving rise to the cross-peaks are indicated to the right of each panel.

identified spin systems such as glycine, alanine, and valine were used to align these segments with the primary sequence.

Triple-Resonance Experiments. The subsequent availability of the doubly labeled ^{13}C–^{15}N K94 sample made it possible to collect a set of triple-resonance experiments (Kay et al., 1990a). The HNCA experiment correlates the N_i, $C\alpha_i$, and the HN_i resonances. In addition, the small but finite N_i–$C\alpha_{i-1}$ coupling provides a sequential correlation to the $C\alpha_{i-1}$ resonance. This spectrum confirmed the backbone sequential assignments made on the basis of the ^{15}N-correlated spectra and provided the $^{13}C\alpha$ resonances. Panels representing 2D F_1–F_3 slices of the 3D HNCA spectrum corresponding to the F_2 (^{15}N) positions of residues 46–56 are shown in Figure 4, and a tracing of the sequential connectivities is included.

Backbone assignments were completed by assignment of the ^{13}C carbonyl resonances. This was achieved by correlating the carbonyl resonance of residue i with the amide proton and nitrogen resonances of residue i+1 in the HNCO experiment (data not shown). Ambiguities were resolved by referring to the HCACO experiment (not shown) where intraresidue carbonyl, α-carbon, and α-proton resonances were correlated.

After the backbone assignments were established, the side-chain ^1H and ^{13}C assignments were obtained by analysis of 3D HCCH-COSY and 3D HCCH-TOCSY spectra. Two dimensional F_1–F_3 planes from the 3D HCCH-TOCSY spectra corresponding to the F_2 (^{13}C) positions of the carbons of Lys_{39} are shown in Figure 5. Strong correlations between virtually all carbon-bound ^1H spins within this residue are seen in this spectrum. For other spin systems, correlations between all ^1H spins were not always observed. The mixing time dependence of the magnetization transfer to DIPSI-3 (Clore et al., 1990) accounted for nearly all of these cases. However,

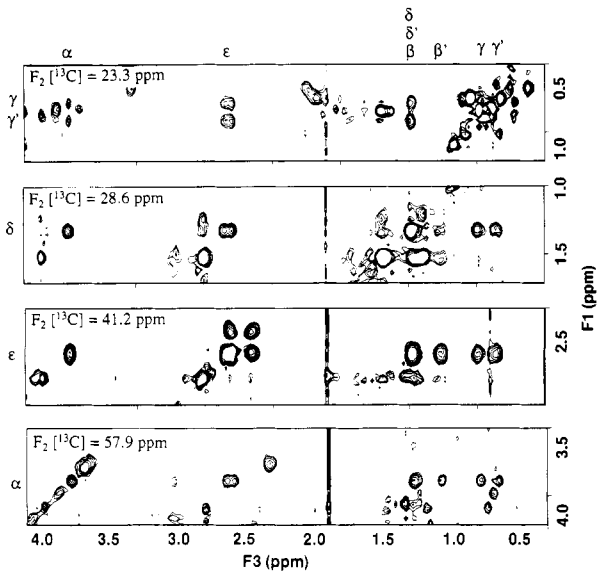

FIGURE 5: Portions of the 3D HCCH-TOCSY spectra illustrating the connectivities seen in the Lys_{39} spin system. The resonance positions of the Lys_{39} protons along F_3 are shown at the top of the figure. Relevant portions of the F_1 (^1H)–F_3 (^1H) planes are shown taken at the F_2 (^{13}C) frequency of the atoms indicated at the left.

correlations missing in the HCCH-TOCSY were found in the 3D HCCH-COSY experiment, allowing nearly complete assignments to be made.

The ^1H, ^{13}C, and ^{15}N assignments for both the backbone and side-chain resonances of hnRNP C RBD are listed in Table I. The assignments are essentially complete except for a few residues. The resonances corresponding to the side chains of Lys_{29} and Lys_{30} were not observed. The aromatic resonances of Phe_{19} and Phe_{52} were observed but not assigned due to overlap.

A summary of the sequential NOE connectivities is disgramed in Figure 6. Amides that were found to be protected from exchange with the solvent are also indicated in the diagram. The conserved RNP1 and RNP2 sequences (residues 50–57 and 18–23) have contiguous stretches of protected amides as well as strong $d\alpha_{i-1}N_i$ connectivities. This is consistent with the placement of these segments as the middle two strands of a β-sheet. Long-range NOEs representing proton interactions between adjacent strands of the β-sheet were observed and are indicated in the diagram of the four-stranded antiparallel β-sheet structure illustrated in Figure 7a.

In order to obtain additional NOEs, we collected two 4D heteronuclear edited ^1H NOESY spectra. A portion of a representative 2D plane of the 4D ^{13}C–^{13}C edited NOESY spectrum is shown in Figure 8. This part of the spectrum is a map of protons (F_4), edited by their attached carbon frequencies (F_3), which are involved in a NOE with the δ-methyl group of Ile_{20}. Data of this type were used to compile a small, selected set of long-range NOEs used in the preliminary structure calculations. In future studies, a more detailed analysis of these spectra will yield a much larger set of distance constraints in order to refine the structure.

Structure Calculations. A bounds file composed of qualitative NOE constraints and hydrogen-bond constraints was used to calculate preliminary structures in order to determine the global folding pattern of the domain. Of the resulting structures, those exhibiting the least NOE violations converged to the common fold illustrated in Figure 9. The $\beta\alpha\beta$–$\beta\alpha\beta$ structure is clearly evident, with the β-strands forming a four-stranded antiparallel β-sheet (see Figure 7B). The four β-strands are made up of residues Arg_{17}–Leu_{23} (β_1), Gly_{45}–

FIGURE 6: Summary of sequential connectivities and results of amide proton exchange experiments. The thickness of the lines indicates the relative strength of the NOE cross-peaks. $d\delta$N connectivities for proline residues are included with the dNN NOEs. Amide protons protected from ^2H_2O exchange are denoted by the circles. Secondary structure assignments are indicated above the primary sequences. The conserved RNP 2 (residues 18–23) and RNP 1 (residues 50–57) sequences are designated with boxes.

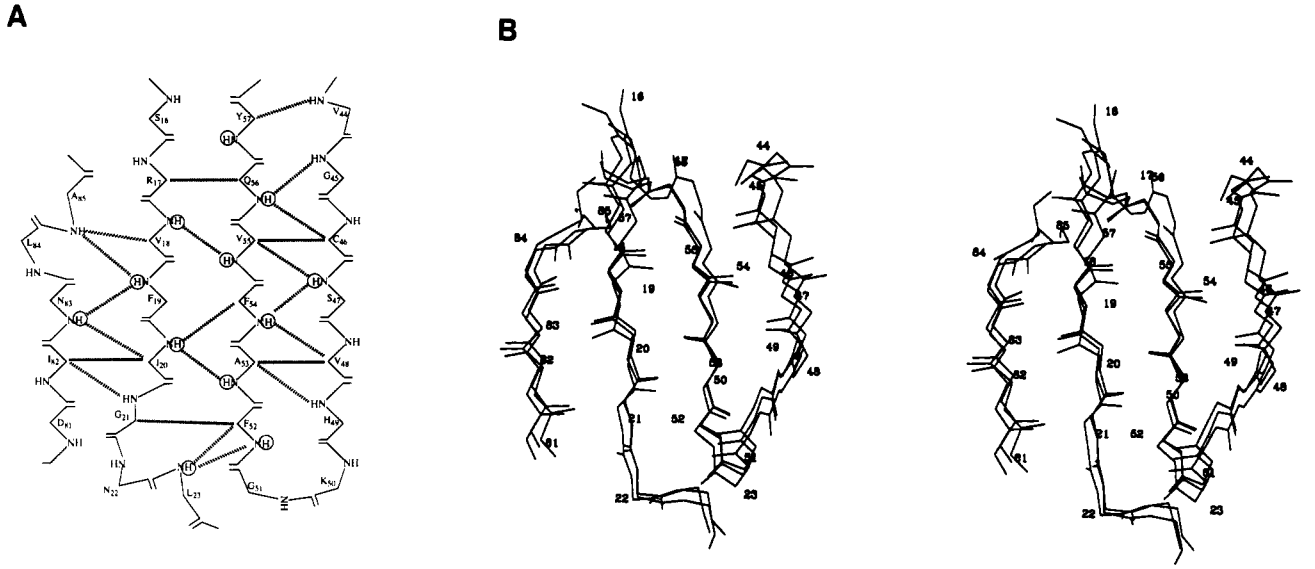

FIGURE 7: β-Sheet structure of the hnRNP C RBD. (A) Diagram indicating NOEs observed between strands of the β-sheet. Dashed lines represent NOEs. Circles indicate amide protons protected from ^2H_2O exchange. (B) Stereoview of the backbone atoms of the four strands of the β-sheet. The three calculated structures exhibiting the lowest NOE violations are shown superimposed with a best fit over all backbone atoms.

Val_{48} (β_2), Ala_{53}–Tyr_{57} (β_3), and Leu_{80}–Leu_{85} (β_4). Strands β_2 and β_3 are connected by a tight turn made up of residues His_{49}–Phe_{52}. The two inner strands of the sheet are the two highly conserved segments of homology, RNP 1 (residues 50–57) and RNP 2 (residues 18–23), seen in the RBDs of other members of this RNA-binding protein family.

The two α-helices lie on the other side of the domain structure, at nearly 90° angles to one another. The two α-helices are comprised of residues Lys_{29}–Tyr_{37} (α_1) and Asn_{59}–Asp_{71} (α_2). In helix α_2, Asn_{59} is likely to act as the N-cap residue (Richardson & Richardson, 1988) with a hydrogen bond between the side-chain oxygen atom of this residue and the amide proton Asn_{62}, accounting for the protection from solvent exchange observed for that amide proton. The positioning of the two α-helices relative to one another is established by connections to the β-sheet structure and a cluster of NOEs involving the aromatic side-chain resonances of Phe_{37}, Tyr_{40}, and Tyr_{57} with the methyl group of Ala_{66}. The loops connecting the elements of regular secondary structure are less well defined in the model of the hnRNP C RBD, and more precise positioning must await refinement of this set of preliminary structures. Note that residues 2–13 (N-terminus to the beginning of the β_1 strand) and 90–94 (end of the β_4 strand to the C-terminus) are disordered among the different

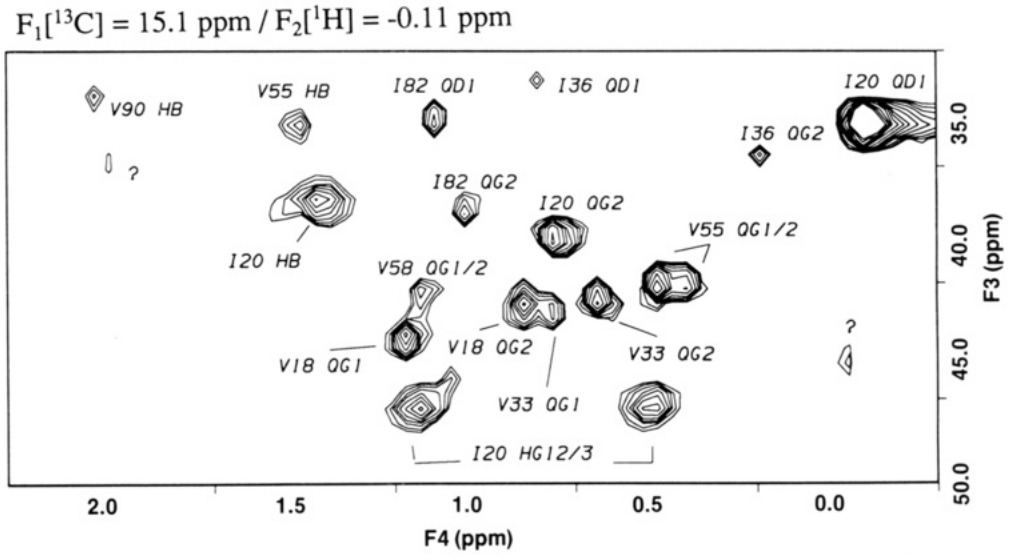

FIGURE 8: Portion of a selected F_3 (^{13}C)–F_4 (^{1}H) 2D plane of the 4D ^{13}C/^{13}C-edited NOESY spectrum. The F_1 (^{13}C) and F_2 (^{1}H) dimensions are locked at 15.1 and −0.11 ppm, respectively, which are the carbon and proton frequencies of the δ-methyl group of Ile_{20}. DIANA (Güntert et al., 1990) nomenclature is used for labeling the cross-peaks with atom names.

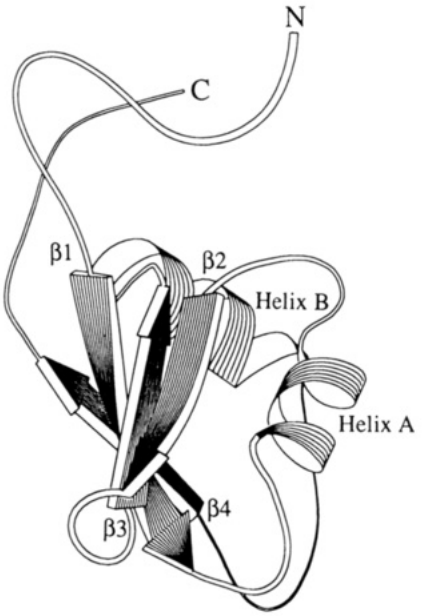

FIGURE 9: Schematic ribbon representation of the hnRNP C RBD (residues 2–94). The arrows indicate the four antiparallel β-strands, and the curled ribbons represent the two α-helices. The labels β1–β4 indicate the beginnings of the four strands of the β-sheet. N and C denote the amino and carboxy termini of the domain, respectively. The drawing was generated with the program RIBBON (J. Priestle) using the calculated structure having the lowest total NOE violations. Residues 2–13 and 90–94 are disordered among the set of calculated structures.

calculated structures, reflecting the lack long-range constraints for these residues.

Discussion

The backbone ^{1}H and ^{15}N resonances of hnRNP C RBD were initially assigned by utilizing a series of ^{15}N-edited 3D spectra using the uniformly labeled ^{15}N protein. Later, triple-resonance experiments on the doubly uniformly labeled ^{13}C–^{15}N K94 sample allowed assignment of the ^{13}C backbone atoms, confirmation of the sequential connectivities, and nearly complete ^{13}C as well as ^{1}H side-chain assignments to be carried out. The assignments allowed for the identification of 501 NOE interactions and 30 hydrogen bonds that were used as constraints for the calculation of a set of preliminary structures. The structures having the smallest constraint violations exhibited the same overall folding pattern. The triple-resonance experiments were essential for obtaining the complete set of assignments that will allow for a detailed refinement of these structures. However, the prior acquisition and analysis of the ^{15}N-edited spectra were sufficient to obtain the correct sequential assignments and, had the doubly labeled protein remained unavailable, these data would have provided enough information to generate structures of comparable quality and been sufficient to allow one to proceed with a more limited refinement of the model.

Except for the posttranslationally cleaved Met_1 residue, the N-terminus of the hnRNP C RBD fragment used in this study corresponds to the N-terminus of the intact hnRNP C1 and C2 proteins. On the basis of the 35% narrower line widths observed in the 2D HSQC ^{15}N–^{1}H correlation spectrum (Figure 1), residues from the N-terminus to Ser_{13} and residues 90–94 at the C-terminus appear to be considerably more mobile than the majority of the residues in the RBD domain. These observations are consistent with the near absence of long-range NOE interactions in these segments of the domain, giving rise to disordered N- and C-termini in the set of calculated structures.

While almost all resonances fall within the normal ranges seen for their residue types in folded proteins (Gross & Kalbitzer, 1988; Clore et al., 1990), a few resonances exhibit extreme chemical shift values. For example, one of the ^{1}Hβ resonances of His_{49} is shifted upfield to 0.94 ppm while the other ^{1}Hβ resonates at a more normal value of 2.77 ppm. Another example is the upfield-shifted ^{1}Hα resonance of Ala_{66} at 2.11 ppm. This extreme ^{1}Hα upfield shift is consistent with the close proximity of this proton to the three aromatic side chains of Phe_{37}, Tyr_{40}, and Tyr_{57} found in the hydrophobic core of the domain. NOEs among these four residues are important long-range interactions that help to define the overall global fold. The same type of upfield shift was observed for the ^{1}Hα resonance of the cognate Ala_{68} residue in the NMR study of the U1 snRNP A RBD (Hoffman et al., 1991).

In the hnRNP C RBD, the ^{13}Cα, ^{13}Cβ, and ^{13}C carbonyl chemical shifts are sensitive to secondary structure. Plots of the ^{13}Cα and ^{12}Cβ chemical shift deviations from random coil positions (secondary shift) versus primary sequence are shown in the panels A and B of Figure 10, while the deviations from

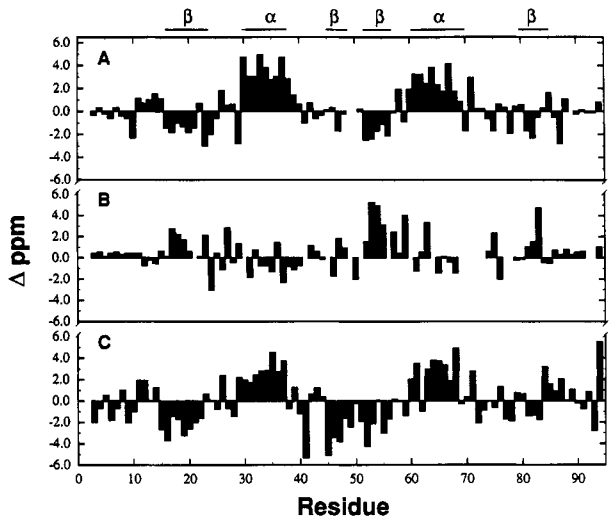

FIGURE 10: Deviations of ^{13}C chemical shifts. (A, B) The deviations in secondary chemical shift for $^{13}C\alpha$ (A) and $^{13}C\beta$ (B) resonances were calculated using random coil ^{13}C chemical shifts (Spera & Bax, 1991) and are plotted versus primary sequence of residues 2–94 for hnRNP C (K94). (C) Deviations of the ^{13}C carbonyl chemical shift from the average value observed in K94 (175.9 ppm) are plotted versus primary sequences of K94. The β and α designations at the top of the figure indicate the positions of the secondary structural elements. A positive value for Δ ppm indicates a downfield shift in chemical shift position.

the average chemical shift values for the ^{13}C carbonyl resonances versus primary sequence are plotted in panel C. There is a downfield chemical shift tendency for the $^{13}C\alpha$ and ^{13}C carbonyl resonances of residues involved in both helices α_1 and α_2. In addition, the residues involved in the β-sheet structure display upfield-shifted $^{13}C\alpha$ and ^{13}C carbonyl resonances and downfield-shifted $^{13}C\beta$ resonances. These observations are in agreement with reports which have documented a correlation between ^{13}C chemical shifts and backbone conformations for polypeptides in the solid state (Saito & Ando, 1989) and for a limited set of proteins for which both high-resolution crystal structures and $^{13}C\alpha$ and $^{13}C\beta$ NMR assignments from solution studies were available (Spera & Bax, 1991). The correlation for the β-sheet residues of the RBD is more pronounced for the two inner strands of the β-sheet, perhaps indicating that these strands are more rigid and/or adopt more regular extended β structures. Also, it may be significant that the ^{13}C carbonyl of Leu_{23} and the ^{13}C carbonyl and $^{13}C\beta$ resonances of Leu_{84} and Ala_{85} deviate from the predicted pattern since the backbone conformations of these residues appear to be distorted to accommodate the β-bulges in the sheet (see Figure 7 and below).

It has been noted that the $^{13}C\alpha$ chemical shifts of residues preceding proline residues often exhibit upfield secondary chemical shifts (Torchia et al., 1975; Clore et al., 1990). In the hnRNP C RBD, both Asp_{10} and Glu_{87} are upfield-shifted relative to their random coil positions.

The global folding patterns of hnRNP C RBD is essentially the same as that determined for the RBD of the A protein component of the U1 small nuclear ribonucleoprotein (U1 snRNP A RBD), as determined by X-ray crystallography (Nagai et al., 1990). 2D NMR spectroscopy has been used to assign the ^1H resonances of the U1 snRNP A RBD (Hoffman et al., 1991), and the global fold determined in that study, based on model building using a limited set of NOEs, is consistent with the X-ray structure and the model presented here.

Prior to the publication of the models based on X-ray and NMR results, the $\beta\alpha\beta$–$\beta\alpha\beta$ secondary structural pattern of the RBDs was correctly predicted on the basis of an analysis of the results of the application of secondary structure prediction algorithms to an alignment of 22 published RBD sequences (Ghetti et al., 1989). Subsequently, the correct overall global folding pattern of the RBD domain was predicted by model building (Ghetti et al., 1990) based on the structure of acylphosphatase (Saudek et al., 1989a), which displays a $\beta\alpha\beta$–$\beta\alpha\beta$ fold. An important detail not described in these studies is that although the overall global folds are the same for acylphosphatase and hnRNP C RBD structures (as well as that of the U1 snRNP A RBD), they differ in that the two α-helices are nearly antiparallel in acylphosphatase whereas they are at almost right angles to each other in the RBD structures. This appears to be due to the positioning along helix α_1 of two key aromatic residues, one helical turn apart, that are major structural components of the hydrophobic core. In the hnRNP C RBD structure, Phe_{37} and Tyr_{40} are located toward the C-terminal end of helix α_1. Side chains of these residues make contact with the methyl group of Ala_{66}, located in the middle of helix α_2. In contrast, acylphosphatase has the two aromatic residues from helix α_1 (Phe_{22} and Tyr_{25}) shifted one helical turn toward the N-terminal end of the helix. A hydrophobic interaction takes place between side chains of these residues and Trp_{64}, located at the C-terminus of helix α_2 (Saudek et al., 1989b). This results in an antiparallel helical arrangement for acylphosphatase whereas the hydrophobic interactions dictate a perpendicular helical orientation in the RBD structures. The perpendicular arrangement of the two helices is almost certainly a feature the other RBD structures since the position of the two aromatic residues in helix α_1, as well as the alanine residue in the middle of helix α_2, are conserved in nearly all RBD sequences (Bandziulis et al., 1989; Kenan et al., 1991). In addition, the conservation of the helical arrangement is supported by the extreme upfield chemical shifts of the $^1H\alpha$ resonances of the conserved alanine in helix α_2 of both RBDs for which NMR assignments are available (Hoffman et al., 1991; this paper).

The antiparallel β-sheet structure exhibits a pronounced right-handed twist. As can be seen in Figure 7, the sheet has a number of irregularities. For example, there is a β-bulge in strand β_4 at residues Asn_{83} and Leu_{84}. Also, a β-bulge exists in strand β_1 at residues Gly_{21} and Asn_{22}. This distortion has also been observed in the U1 snRNP-A RBD X-ray structure (Nagai et al., 1990), although Gly_{21} of the hnRNP C RBD is replaced by an asparagine in that protein. Since this is part of the conserved RNP2 sequence, it is likely to be a common feature of all of the RBDs.

While the overall global folding pattern is conserved between the hnRNP C and U1 snRNP A RBDs, there are a number of differences between the two structures. The major distinction is that the U1 snRNP A RBD sequence has an insertion of five residues just amino-terminal to RNP1 relative to hnRNP C. This insertion has the effect of extending the residues leading into and out of the β-turn connecting strands 2 and 3 of the β-sheet in the hnRNP C RBD structure into the extended loop structure seen in the U1 snRNP A RBD structure. The loop connecting the two β-strands of the U1 snRNP A RBD structure is thought to be important in mediating the binding to U1 RNA (Nagai et al., 1990; Scherly et al., 1990; Jessen et al., 1991). This is also the region where the different RBD sequences exhibit the greatest variability, in terms of length as well as residue type (Bandziulis et al., 1989; Keenan, 1991). Studies on the RNA-binding of the U1 snRNP A and hnRNP C proteins and structural information about the RBDs suggest that sequence and length differences

Table I: ^{1}H, ^{15}N, and ^{13}C Chemical Shift Assignments for hnRNP-C Fragment (residues 2–94) at 20 °C and pH 5.5

Ala	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$
A35	126.2	7.91	180.4	55.0	4.18	17.7	1.60
A53	124.9	8.83	173.8	49.9	5.02	24.2	1.16
A63	124.6	6.81	178.8	54.7	3.75	22.3	1.49
A65	128.1	7.79	179.5	54.6	4.14	17.6	1.68
A66	128.7	7.59	179.2	54.0	2.11	19.1	1.51
A68	122.8	7.59	180.8	54.0	4.16	17.6	1.46
A76	137.1	9.60	177.2	52.9	4.09	17.0	1.45
A85	132.8	9.13	177.4	53.9	4.20	18.5	1.50
A86	123.1	8.07	176.8	51.8	4.34	19.7	1.37

Arg	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$	$^{13}C\gamma$	$^{1}H\gamma$	$^{13}C\delta$	$^{1}H\delta$	$^{15}N\varepsilon$
R12	123.6	8.43	177.8	56.8	4.25	29.6	1.91, 1.91	26.8	1.73, 1.64	43.0	3.23, 3.23	119.8
R17	128.5	7.67	174.5	54.3	5.31	33.0	1.91, 1.91	27.8	1.48, 1.48	43.3	2.94, 2.94	121.8
R61	124.5	8.68	179.3	59.3	4.01	29.1	1.91, 1.78	26.9	1.72, 1.66	43.1	3.25, 3.25	120.2
R64	119.9	8.69	179.6	59.9	3.84			29.6	1.99, 1.79	42.8	3.27, 3.17	119.1
R73	126.8	8.04	175.0	56.3	4.22		2.10		1.70			
R92	127.2	8.49	176.7	56.0	4.34	30.3	1.66, 1.63	26.9	1.93, 1.80	43.0	3.21, 3.21	119.8

Asn	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$	$^{13}C\gamma$	$^{15}N\delta$	$^{1}H\delta$
N4	126.2	8.72	175.2	53.1	4.78	38.4	2.85, 2.81	58.7	117.4	7.64, 6.89
N7	125.9	8.46	175.2	53.1	4.71	38.4	2.85, 2.75	58.7	117.9	7.67, 6.98
N15	122.5	8.47	173.2	51.9	4.72	38.5	2.84, 2.77	57.9	116.3	7.53, 6.89
N22	122.8	8.86	174.3	53.5	4.47	38.0	3.88, 2.55		117.2	8.46, 6.84
N24	128.9	9.10	175.9	50.8	4.81	34.9	3.18, 2.49		115.1	8.07, 7.21
N59	117.2	7.38	174.5	51.9	5.12	41.9	3.05, 2.97		119.0	7.81, 6.93
N62	122.9	7.93	175.0	56.0	4.33	38.4	2.52, 2.38		114.3	7.92, 7.14
N83	124.3	8.87	174.2	52.3	5.02	42.6	2.91, 2.91	62.9	119.0	7.22, 6.81
N91	127.6	8.62	175.1	52.9	4.73	38.5	2.85, 2.76			

Asp	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$
D10	129.2	8.45	174.9	51.7	4.89	41.2	2.80, 2.65
D32	127.3	7.23	178.3	57.0	4.59	41.5	3.23, 2.72
D71	123.2	7.89	178.6	56.9	4.33	40.8	3.14, 2.67
D81	128.1	7.80	174.5	52.3	5.13	41.8	2.70, 2.58

Cys	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$
C46	128.0	8.55	172.5	57.2	5.15	27.2	2.93, 2.65

Gln	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{1}H\beta$	$^{13}C\gamma$	$^{1}H\delta$	$^{15}N\varepsilon$	$^{1}H\varepsilon$
Q56	132.2	8.85	174.2	54.0	4.80	2.30, 2.13	34.4	2.46, 2.30	121.4	7.31, 7.67
Q78	123.7	7.79	174.0	54.2	4.76	2.28, 2.06	54.4	2.11, 1.64	115.0	7.07, 6.69

Glu	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$	$^{13}C\gamma$	$^{1}H\delta$
E34	123.7	8.68	178.7	60.2	3.79	29.0	2.06, 2.10	37.6	2.55, 2.02
E60	127.8	8.86	177.9	58.3	4.21			35.7	2.62, 2.20
E70	119.3	7.66	176.2	54.7	4.65				1.52
E87	124.2	7.99	177.9	53.6	4.61	30.0	2.05, 1.90	35.7	2.28, 2.19

Gly	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$
G21	118.2	8.55	173.9	43.7	4.36, 3.78
G41	111.4	7.47	170.5	44.1	4.48, 3.90
G45	112.7	7.31	170.8	45.2	4.24, 3.89
G51	120.8	8.70	174.0	45.3	4.53, 4.42
G69	108.7	8.22	175.6	45.9	4.02
G72	119.1	8.37	173.8	45.3	4.21, 3.72
G77	106.6	8.55	174.2	45.4	4.22, 3.74
G93	115.1	8.51	173.1	45.0	3.97, 3.97

His	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$	$^{1}H\delta$	$^{1}H\varepsilon$
H49	133.4	8.56	173.5	53.8	4.63	29.3	2.77, 0.94	6.47	7.20

Ile	^{15}N	^{1}HN	$^{13}C'$	$^{13}C\alpha$	$^{1}H\alpha$	$^{13}C\beta$	$^{1}H\beta$	$^{13}C\gamma$	$^{1}H\gamma$	$^{13}C\gamma m$	$^{1}H\gamma m$	$^{13}C\delta$	$^{1}H\delta$
I20	132.2	8.35	173.3	59.5	4.04	38.6	1.39	27.1	1.13, 0.49	20.1	0.75	15.1	-0.11
I36	121.8	7.76	178.6	64.3	3.72	39.4	1.73	28.2	1.90, 0.96	16.5	0.16	13.1	0.80
I43	131.9	8.77	177.1	60.7	4.12	38.6	1.76	27.2	1.86, 1.86	17.6	0.68	13.8	0.80
I75	132.1	8.71	175.3	59.6	4.44	40.3	1.86	27.1	1.53, 1.19	16.7	0.92	12.3	0.87
I82	129.5	9.00	174.5	59.0	5.41	39.5	1.78	27.9	1.76, 1.20	19.1	0.99	14.9	1.08

Table I (Continued)

Leu	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{13}Cγ	^{1}Hγ	^{13}Cδ	^{1}Hδ	^{13}Cδ	^{1}Hδ
L23	123.6	7.41	176.5	52.1	4.03	44.4	1.31, 0.84			22.4	0.78	22.4	0.78
L26	126.0	7.74	178.2	56.9	4.34	41.2	1.98, 1.67	27.0	1.72	24.4	0.93	24.4	0.98
L80	133.9	9.10	176.5	55.7	4.46	42.2	2.21	23.4	2.21	25.3	1.06	23.0	0.83
L84	127.1	9.21	179.0	55.4	4.68	41.9	1.84, 1.68	27.6	1.84	25.1	0.98	24.1	0.87

Lys	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{13}Cγ	^{1}Hγ	^{13}Cδ	^{1}Hδ	^{13}Cε	^{1}Hε
K8	126.9	8.40	176.9	56.1	4.38	32.8	1.93, 1.78	24.6	1.46, 1.43	28.7	1.72, 1.72	41.7	3.02, 3.02
K29	129.5	9.02	178.0	53.7	4.60	33.8	2.07, 1.76						
K30	128.3	9.08	177.8	61.2	3.64								
K39	124.3	7.10	177,1	57.9	3.98	31.4	1.49, 1.29	23.3	1.00, 0.86	28.6	1.49, 1.49	41.2	2.85, 2.80
K42	122.3	8.04	176.5	57.2	4.21	33.6	1.86, 1.71	24.5	1.39, 1.39	28.7	1.68, 1.68	41.7	3.01, 3.01
K50	126.4	7.98	175.7	56.5	4.42	30.5	1.98, 1.79	24.5	1.46, 1.34	28.7	1.68, 1.68	41.7	2.99, 2.99
K89	126.5	8.38	176.9	56.5	4.28	32.8	1.80, 1.72	24.7	1.49, 1.43	29.1	1.75, 1.70	41.8	3.01, 3.01
K94	130.7	7.90	181.4	57.3	4.21	33.5	1.87, 1.72	24.6	1.40, 1.40	28.8	1.72, 1.72	41.8	3.02, 3.02

Met	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{13}Cγ	^{1}Hγ
M14	126.3	8.32	177.1	56.8	4.33	32.1	2.11, 2.03	32.2	2.65, 2.53
M74	129.8	8.75	175.9	54.7	4.92	33.2	2.07, 1.95	31.9	2.35, 2.19

Phe	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{1}Hδ1	^{1}Hε	^{1}Hζ
F19	132.3	9.29	172.7	56.7	4.75	40.7	3.07, 2.94			
F37	118.3	8.59	179.6	62.7	4.41	36.7	3.82, 2.91	8.04	6.99	7.11
F52	118.2	7.21	171.7	55.5	5.31	40.5	3.39, 2.66			
F54	118.9	8.88	175.9	56.3	5.90	43.9	2.93, 2.88	7.31	7.47	6.93

Pro	^{15}N	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{13}Cγ	^{1}Hγ	^{13}Cδ	^{1}Hδ
P11		177.8	64.2	4.43	32.1	2.37, 2.03	27.1	2.06, 2.06	50.6	3.95, 3.95
P88		175.9	64.2	4.06	32.5	2.26, 2.26	21.8	1.12, 1.08		

Ser	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ
S3		8.78	173.9	57.9	4.53	63.6	3.88, 3.88
S13	121.0	8.14	175.9	59.2	4.37	63.0	3.93, 3.87
S16	113.2	7.59	172.2	56.8	4.49	63.3	3.89, 3.40
S31	116.0	8.64	177.6	61.2	4.07	61.4	3.86
S38	126.6	8.95	175.2	61.0	4.86	62.4	4.25, 4.14
S47	127.0	8.78	172.1	56.5	4.76	65.0	3.82, 3.73

Thr	^{15}N	HN	^{13}C'	^{13}Cα	Hα	^{13}Cβ	Hβ	^{13}Cγ	^{1}Hγ
T6	122.6	8.36	174.1	61.5	4.35	69.6	4.22	21.3	1.18
T9	119.8	8.29	173.9	61.5	4.32	69.6	4.18	21.3	1.14
T25	122.7	8.30	175.2	63.2	4.33	69.6	4.33	21.3	1.18

Tyr	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{1}Hδ1	^{1}Hε
Y40	119.2	7.01	174.7	58.7	4.30	38.1	3.39, 2.67	6.80	7.20
Y57	131.2	8.76	176.0	58.0	5.28	41.2	3.72, 3.11	7.04	6.68

Val	^{15}N	^{1}HN	^{13}C'	^{13}Cα	^{1}Hα	^{13}Cβ	^{1}Hβ	^{13}Cγ	^{1}Hγ	^{13}Cγ	^{1}Hγ
V5	125.2	8.25	176.4	62.1	4.18	32.2	2.10	21.0	0.93	20.1	0.93
V18	129.8	9.63	174.2	61.3	4.56	34.3	1.98	22.6	1.12	22.6	0.84
V27	119.4	6.89	175.2	62.8	4.12	34.9	1.78	20.8	0.91	20.8	0.91
V28	126.6	8.31	174.5	62.9	3.94	31.7	1.72	21.1	0.90	21.1	0.90
V33	124.9	7.86	178.6	67.2	3.55	31.4	2.26	23.4	0.76	22.6	0.64
V44	129.7	8.86	176.3	62.0	4.20	32.0	2.00	20.1	0.88	21.1	0.73
V48	128.0	8.74	174.3	62.1	4.11	33.0	1.93	22.1	0.97	21.3	0.73
V55	123.2	8.68	172.9	61.2	4.37	35.2	1.45	22.0	0.46	22.0	0.38
V58	122.0	9.45	175.8	64.2	4.05	32.5	2.26	21.6	1.11	21.6	1.09
V67	121.3	7.67	177.7	66.4	3.42	31.7	2.16	22.7	1.05	21.3	1.00
V79	128.3	8.70	176.6	62.7	4.10	31.9	1.95	22.0	0.99	20.7	0.92
V90	126.1	8.21	175.7	61.8	4.09	32.6	2.03	20.4	0.91	20.4	0.91

in the β-turn/loop region of the sheet provide at least part of the structural basis for the differences in the RNA-binding specificities between these two proteins. We have found that the addition of RNA oligomers to the hnRNP C RBD causes very large chemical shift perturbations on the backbone ^{1}H and ^{15}N resonances of residues making up the β-sheet while leaving resonances from the α-helical segments unaffected (Görlach et al., 1992), suggesting that RNA interactions with the RBD are mediated by contacts on the β-sheet side of the domain.

The structural model presented here was calculated with only a relatively small set of NOE constraints with the purpose

of determining the global folding pattern. Only a small percentage of the information content of the 4D NOESY spectra was utilized for the calculation of a set of preliminary structures. Full utilization of the NOEs observed in these spectra, as well as angle constraints derived from coupling constants, will allow extensive refinement of the model. This will form the basis for further studies involving RBD interactions with RNA.

ACKNOWLEDGMENTS

We thank Dr. Steven Carr (Smith-Kline Beecham) for the mass spectroscopy analysis of the purified protein and the Protein Chemistry Group of the Department of Biochemistry and Biophysics, University of Pennsylvania, under the direction of Dr. Lim Tung, for the amino-terminal peptide sequencing. We also thank Dr. Mark McCoy for help with the shaped pulses and Dr. Robert Beckman and Dr. Keith Constantine for critical reading of the manuscript.

REFERENCES

Adam, S. A., Nakagawa, T. Y., Swanson, M. S., Woodruff, T., & Dreyfuss, G. (1986) *Mol. Cell. Biol. 6*, 2932–2943.

Bandziulis, R. J., Swanson, M. S., & Dreyfuss, G. (1989) *Genes Dev. 3*, 431–437.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M., & Kay, L. E. (1990) *J. Magn. Reson. 87*, 620–627.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185–189.

Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166–169.

Brünger, A. T. (1990) X-PLOR *Version 2.1 User Manual*, Yale University, New Haven, CT.

Burd, C. G., Swanson, M. S., Görlach, M., & Dreyfuss, G. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9788–9792.

Burd, C. G., Matunis, E. L., & Dreyfuss, G. (1991) *Mol. Cell. Biol. 7*, 3419–3424.

Choi, Y. D., Grabowski, P., Sharp, P. A., & Dreyfuss, G. (1986) *Science 231*, 1534–1539.

Chung, S. Y., & Wooley, J. (1986) *Proteins 1*, 195–210.

Clore, G. M., & Gronenborn, A. M. (1991) *Science 252*, 1390–1399.

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.

Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., & Wilson, S. H. (1986) *J. Biol. Chem. 261*, 3536–3543.

Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., & Wilson, S. (1988) *J. Biol. Chem. 263*, 1063–1071.

Dreyfuss, G. (1986) *Annu. Rev. Cell Biol. 2*, 459–498.

Dreyfuss, G., Swanson, M. S., & Piñol-Roma, S. (1988) *Trends Biochem. Sci. 13*, 86–91.

Driscoll, P. C., Gronenborn, A. M., Beress, L., & Clore, G. M. (1989) *Biochemistry 28*, 2188–2198.

Economidis, I. V., & Pederson, T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1599–1602.

Fakan, S., Leser, G., & Martin, T. E. (1986) *J. Cell. Biol. 103*, 1153–1157.

Farmer, B. T., II, Venters, R., Spicer, L., Wittekind, M., & Mueller, L. (1992) *J. Biomol. NMR 2*, 195–202.

Ghetti, A., Bolognesi, M., Cobianchi, F., & Morandi, C. (1990a) *FEBS Lett. 277*, 272–276.

Ghetti, A., Padovani, C., Di Cesare, G., & Morandi, C. (1990b) *FEBS Lett. 257*, 373–376.

Görlach, M., Wittekind, M., Beckman, R., Mueller, L., & Dreyfuss, G. (1992) *EMBO J.* (in press).

Gross, K.-H., & Kalbitzer, H. R. (1988) *J. Magn. Reson. 76*, 87–99.

Güntert, P. (1990) DIANA *Version 1.0 User's Manual and Instructions*, Eidgenössische Technische Hochschule-Hönggerberg, Zürich, Switzerland.

Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517–530.

Hoffman, D. W., Query, C. Q., Golden, B. L., White, S. W., & Keene, J. W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495–2499.

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

Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990a) *J. Magn. Reson. 89*, 496–514.

Kay, L. E., Clore, G. M., Bax, A., & Gronenborn, A. M. (1990b) *Science 249*, 411–414.

Kenan, D. J., Query, C. C., & Keene, J. D. (1991) *Trends Biochem. Sci. 16*, 214–220.

Kumar, A., Williams, K. R., & Szer, W. (1986) *J. Biol. Chem. 261*, 11266–11273.

Lutz-Freyermuth, C., Query, C. C., & Keene, J. D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6393–6397.

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., Ikura, M., & Bax, A. (1989b) *J. Magn. Reson. 84*, 425–430.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989c) *Biochemistry 28*, 6150–6156.

Mattaj, I. W. (1989) *Cell 57*, 1–3.

Matunis, E. L., Matunis, M. J., & Dreyfuss, G. (1992c) *J. Cell Biol. 116*, 257–269.

Matunis, M. J., Michael, M. W., & Dreyfuss, G. (1992a) *Mol. Cell. Biol. 12*, 164–171.

Matunis, M. J., Matunis, E. L., & Dreyfuss, G. (1992b) *J. Cell Biol. 116*, 245–255.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., & Williams, K. R. (1988) *J. Biol. Chem. 263*, 3307–3313.

Mueller, L., & Ernst, R. R. (1979) *Mol. Phys. 38*, 963–992.

Mueller, L., Campbell-Burk, S., & Domaille, P. (1992) *J. Magn. Reson. 96*, 408–415.

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

Nakagawa, T. Y., Swanson, M. S., Wold, B. J., & Dreyfuss, G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2007–2011.

Nietfeld, W., Mentzel, H., & Pieler, T. (1990) *EMBO J. 9*, 3699–3705.

Nilges, M., Clore, G. M., & Gronenborn, A. M. (1988) *FEBS Lett. 229*, 317–324.

O'Farrel, P. Z., Goodman, H. M., & O'Farrel, P. H. (1977) *Cell 12*, 1133–1142.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J., & Wüthrich, K. (1990) *EMBO J. 9*, 3085–3092.

Piñol-Roma, S., Choi, Y. D., Matunis, M. J., & Dreyfuss, G. (1988) *Genes Dev. 2*, 215–227.

Piñol-Roma, S., Swanson, M. S., Gall, J. G., & Dreyfuss, G. (1989) *J. Cell Biol. 109*, 2575–2587.

Powers, R., Gronenborn, A. M., Clore, G. M., & Bax, A. (1991) *J. Magn. Reson. 94*, 209–213.

Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Query, C. C., Bentley, R. C., & Keene, J. D. (1989) *Cell 57*, 89–101.
Sachs, A. B., Bond, M. W., & Kornberg, R. D. (1986) *Cell 45*, 827–835.
Saito, H., & Ando, I. (1989) *Annu. Rep. Nucl. Magn. Reson. Spectrosc. 21*, 251–263.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
Saudek, V., Williams, R. P. J., & Ramponi, G. (1989a) *FEBS Lett. 242*, 225–232.
Saudek, V., Atkinson, R. A., Williams, R. P. J., & Ramponi, G. (1989b) *J. Mol. Biol. 205*, 229–239.
Scherly, D., Boelens, W., van Venrooji, W. J., Dathan, N. A., Hamm, J., & Mattaj, I. W., (1989) *EMBO J. 8*, 4163–4170.
Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., & Mattaj, I. W. (1990) *Nature 345*, 502–506.
Shaka, A. J., Keeler, J., Frenkiel, T., & Freeman, R. (1983) *J. Magn. Reson. 52*, 335–338.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274–293.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490–5492.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubbendorf, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Swanson, M. S., & Dreyfuss, G. (1988a) *Mol. Cell. Biol. 8*, 2237–2241.
Swanson, M. S., & Dreyfuss, G. (1988b) *EMBO J. 11*, 3519–3529.
Swanson, M. S., Nakagawa, T. Y., LeVan, K., & Dreyfuss, G. (1987) *Mol. Cell. Biol. 7*, 1731–1739.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405–410.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387–2391.

Pulsed EPR Studies of the Type 2 Copper Binding Site in the Mercury Derivative of Laccase†

Jinfeng Lu,‡,§ Christopher J. Bender,‡ John McCracken,‡,∥ Jack Peisach,*,‡ John C. Severns,⊥ and David R. McMillin*,⊥

Department of Molecular Pharmacology, Albert Einstein College of Medicine of Yeshiva University, 1300 Morris Park Avenue, Bronx, New York 10461, and Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received March 2, 1992; Revised Manuscript Received April 27, 1992

ABSTRACT: The nuclear modulation effect in pulsed EPR spectroscopy was used to study the type 2 copper binding site in the mercury derivative of laccase (MDL) in which the type 1 copper is substituted by Hg(II). By comparing the three-pulse electron spin–echo modulations and Fourier transform spectra of MDL and several model compounds, we conclude that the imidazole groups of two histidyl amino acid residues are equatorially coordinated to Cu(II) in the type 2 site. Computer simulations of these data suggest that the remote nonbonding nitrogens of the two imidazoles possess nuclear quadrupole parameters e^2qQ = 1.47 MHz and η = 0.83. A_{iso} values of these two nitrogens are not identical, being 1.5 and 2.0 MHz. We have also used samples of the enzyme exchanged with D_2O to examine the coordination of the water to the type 2 copper site. The deuterium modulation that is resolved by taking the ratio of the time domain ESEEM data from native and D_2O-exchanged enzyme indicates that there is an equatorial water ligand, and further data show that this water is displaced by azide.

Laccase is an intensely blue, copper-containing protein found in plants and fungi (Reinhammar, 1984). It catalyzes the oxidation of aromatic diamines and diphenols, yielding water as the final product of dioxygen reduction. The intense blue color arises from a thiolate to Cu(II) charge transfer at one of the metal sites (McMillin et al., 1974; Solomon et al., 1976a,b). This site, termed type 1 (Malmstrom et al., 1968), has an unusually small nuclear hyperfine coupling constant, $A_\parallel$, in the EPR.[1] On the basis of optical and magnetic properties, it resembles some low molecular weight, mononuclear copper proteins where the copper is bound to a cysteinyl sulfur, two histidyl imidazoles, and a methionyl sulfur (Colman et al., 1978; Adman et al., 1978; Norris et al., 1981). Imidazole coordination is demonstrated by pulsed EPR methods (Mondovi et al., 1977; Avigliano et al., 1981). The

† This work supported by National Institute of General Medical Sciences Grant GM 40168 and NIH Grant RR-02583 (J.P.) and GM 22764 from the U.S. Public Health Service (D.M.).

* Authors to whom correspondence should be addressed.
‡ Einstein College of Medicine.
§ Present address: National Laboratory of Natural and Biomimetic Drugs, Beijing Medical University, Beijing 100083, People's Republic of China.
∥ Present address: Department of Chemistry, Michigan State University, East Lansing, MI 48824-1322.
⊥ Purdue University.

[1] Abbreviations: EPR, electron paramagnetic resonance; ESEEM, electron spin–echo envelope modulation; ENDOR, electron–nuclear double resonance; MDL, mercury derivative of laccase; LEFE, linear electric field effect; NQI, nuclear quadrupole interaction; EXAFS, extended X-ray absorption fine structure spectroscopy.

0006-2960/92/0431-6265$03.00/0 © 1992 American Chemical Society