Determination of the Secondary Structure and Folding Topology of an RNA Binding Domain of Mammalian hnRNP A1 Protein Using Three-Dimensional Heteronuclear Magnetic Resonance Spectroscopy[†]

Daniel S. Garrett,[‡] Patricia J. Lodi,[‡] Yousif Shamoo,[§] Kenneth R. Williams,[§,⊥] G. Marius Clore,[*,‡] and Angela M. Gronenborn[*,‡]

Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, and Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University Medical School, New Haven, Connecticut 06510

ABSTRACT: The secondary structure and folding topology of the first RNA binding domain of the human hnRNP A1 protein was determined by multidimensional heteronuclear NMR spectroscopy. The 92 amino acid long domain exhibits a $\beta\alpha\beta\beta\alpha\beta$ folding pattern, arranged in a four-stranded antiparallel β-sheet flanked by two α-helices, which is very similar to that found for other members of this family. Regions of marked variation between the structurally characterized RNA binding proteins of this class to date are mainly localized in the loops connecting the secondary structure elements.

Newly synthesized mRNA in the nucleus of eukaryotic cells is generally associated with a large number of RNA binding proteins in the form of heterogeneous nuclear ribonucleoprotein (hnRNP) particles [for a recent review see Dreyfuss et al. (1993)]. Several of these proteins play an important role in transcript-specific packaging, alternative splicing, and transport of mRNA (Beyer et al., 1977; Chung & Wooley, 1986; Mayrand et al., 1981; Ge & Manley, 1990; Piñol-Roma & Dreyfuss, 1992; Choi & Dreyfuss, 1984). Among human hnRNPs, protein A1 is one of the best characterized. It consists of a two-domain structure (Cobianchi et al., 1986). The N-terminal region contains two homologous 92 amino acid long nucleic acid binding domains (Merrill et al., 1986), whereas the C-terminal domain is characterized by a high content of glycine residues (Cobianchi et al., 1986). Each approximately 90 amino acid long nucleic acid binding domain contains the octamer and hexamer consensus sequences (RNP1 and RNP2 motifs) found in a large number of RNA binding proteins from *Escherichia coli*

to humans (Adam et al., 1986; Sachs et al., 1986; Bandziulis et al., 1989; Kenan et al., 1991). In order to gain insight into the detailed features responsible for the interaction between the RNA binding domain and RNA, it is necessary to carry out structural studies by X-ray crystallography and NMR on the isolated components as well as the complexes. The results of such studies will aid our understanding of specific recognition between proteins and RNA in much the same way as the wealth of structural data on DNA–protein complexes has already achieved [see Steitz (1990), Harrison (1991), Pabo and Sauer (1992), and Wolberger (1993) for reviews]. At present, very limited structural information on RNA binding proteins is available. The X-ray structure of the RNA binding domain of the U1A protein, which is a component of the human U1 small nuclear ribonucleoprotein particle (snRNP), has been determined (Nagai et al., 1990), and an NMR investigation on the same molecule has been carried out (Hoffman et al., 1991). In addition, NMR studies on the RNA binding domain of hnRNP C are under way (Wittekind et al., 1992; Görlach et al., 1992). In this paper we describe an NMR investigation of the first RNA binding domain of the human A1 protein and discuss our findings in the light of current knowledge with respect to the structures of the U1A and hnRNP C RNA binding domains. ^{1}H, ^{13}C, and ^{15}N resonance assignments for A1 (1–92) are presented, the secondary structure is delineated, and similarities to and differences from the other RNA binding domains are pointed out.

EXPERIMENTAL SECTION

Sample Preparation. Uniformly (>95%) ^{15}N and ^{15}N/^{13}C labeled hnRNP A1 RNA binding domain (1–92) was prepared from *E. coli* using the expression vector pYS45 (Y. Shamoo and K. R. Williams, manuscript in preparation) which is derived from plasmid pEX11 containing the entire A1 coding sequence (Cobianchi et al., 1988). A1 protein (1–195) was expressed in *E. coli* BL21 DE3 in minimal media using either [^{15}N]NH$_4$Cl (Aldrich) or [^{15}N]NH$_4$Cl and [^{13}C$_6$]D-glucose (Isotec) as sole nitrogen and carbon sources. Cells were grown in a Labline/SMS Hi-Density fermentor at 35 °C to an A_{590} of 2.8 before induction with 0.6 mM isopropyl β-D-thiogalactopyranoside (IPTG). Growth was continued after in-
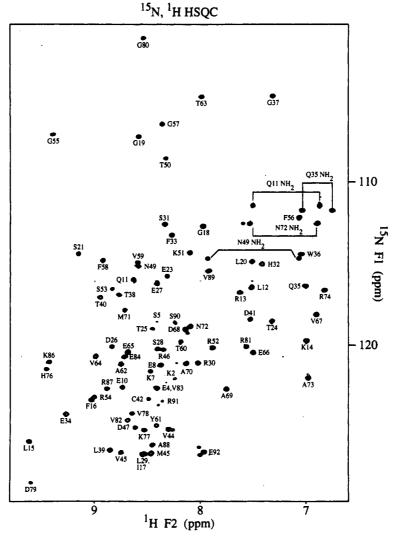
[1] Abbreviations: hnRNP, heterogeneous nuclear ribonucleoprotein; snRNP, small nuclear ribonucleoprotein particle; A1 (1–92), first RNA binding domain of hnRNP A1 (note that the numbering scheme used here includes the N-terminal methionine as residue 0); NMR, nuclear magnetic resonance; CD, circular dichroism; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; CBCANH, C$^\beta$/C$^\alpha$ carbon to nitrogen to amide proton correlation; CBCA(CO)NH, C$^\beta$/C$^\alpha$ carbon to nitrogen (via carbonyl) to amide proton correlation; HBHA(CO)NH, C$^\beta$H/C$^\alpha$H proton to nitrogen (via carbonyl) to amide proton correlation; HNHA, amide proton to nitrogen to C$^\alpha$H proton correlation; HCCH-COSY, proton–carbon–carbon proton correlation using carbon correlated spectroscopy; HCCH-TOCSY, proton–carbon–carbon–proton correlation using carbon total correlated spectroscopy; 2D, two-dimensional; 3D, three-dimensional.

FIGURE 1: (A, top) ^{15}N (F_1 axis)–NH (F_2 axis) region of the 2D ^{15}N–^1H Overbodenhausen correlation spectrum of 1.0 mM ^{15}N-labeled A1 (1–92) in 58 mM sodium phosphate buffer, pH 6.0. Several small resonances are observed for residues at the C-terminus, arising from heterogeneity of the sample due to the proteolytic cleavage. (B, bottom) Selected $^{13}C(F_1)$–$^1H(F_3)$ planes at $\delta[^{15}N(F_2)]$ = 117.0 ppm of the 3D CBCA(CO)NH and CBCANH spectra of A1 (1–92). Sequential connectivities for several residues are indicated.

Table 1: ^{15}N, ^{13}C, and ^{1}H Resonance Assignments for hnRNP A1 at pH 5.7 and 25 °C[a]

residue	N	C$^\alpha$	C$^\beta$	other
M0	* (*)	58.1 (4.14)	31.2 (2.04, 1.87)	C$^\gamma$, 36.5 (2.19, *)
S1	119.5 (8.38)	58.2 (4.53)	64.0 (3.92, *)	
K2	121.7 (8.36)	56.4 (4.41)	33.3 (1.87, 1.80)	C$^\gamma$, 24.8 (1.46, *); C$^\delta$, 29.0 (1.69, *); C$^\epsilon$, 41.9 (3.0)
S3	117.4 (8.46)	58.4 (4.46)	63.8 (3.85, *)	
E4	122.7 (8.42)	56.2 (4.39)	30.7 (2.06, 1.92)	C$^\gamma$, 35.9 (2.26, *)
S5	118.5 (8.42)	56.6 (4.75)	63.2 (3.90, 3.81)	
P6	– (–)	63.3 (4.46)	32.2 (2.36, 1.91)	C$^\gamma$, 27.6 (2.06, *); C$^\delta$, 50.8 (3.87, 3.72)
K7	121.5 (8.49)	56.4 (4.31)	33.4 (1.85, 1.79)	C$^\gamma$, 25.0 (1.53, *); C$^\delta$, 29.0 (1.71, *); C$^\epsilon$, 43.3 (3.11)
E8	121.1 (8.36)	54.5 (4.67)	29.6 (1.96, *)	C$^\gamma$, 35.9 (2.35)
P9	– (–)	63.7 (4.36)	32.4 (2.42, 1.91)	C$^\gamma$, 27.7 (2.10, *); C$^\delta$, 50.8 (3.91, 3.68)
E10	122.5 (8.74)	59.6 (3.74)	29.9 (2.03, 1.96)	C$^\gamma$, 36.6 (2.19, *)
Q11	116.0 (8.65)	58.4 (4.03)	28.2 (2.06)	C$^\gamma$, 33.6 (2.43, 2.36); N$^\epsilon$, 110.9 (7.50, 6.87)
L12	116.4 (7.52)	56.0 (4.31)	42.4 (1.71, 1.69)	C$^\gamma$, 27.5 (1.68); C$^{\delta 1}$, 25.3 (0.99); C$^{\delta 2}$, 22.9 (0.91)
R13	116.8 (7.64)	56.2 (4.46)	31.0 (2.62)	C$^\gamma$, * (*); C$^\delta$, * (*)
K14	119.7 (7.00)	55.4 (5.23)	35.9 (1.69, *)	C$^\gamma$, 24.7 (1.49, 1.15); C$^\delta$, 29.5 (1.31, *); C$^\epsilon$, 41.2 (2.40)
L15	125.8 (9.65)	53.2 (5.06)	45.0 (1.56, 1.11)	C$^\gamma$, 26.6 (1.62); C$^{\delta 1}$, 25.2 (0.77); C$^{\delta 2}$, 24.6 (0.82)
F16	123.3 (9.06)	57.3 (4.62)	40.7 (3.04, 2.95)	
I17	126.7 (8.52)	59.3 (4.06)	37.0 (1.55)	C$^{\gamma 1}$, 26.6 (1.00, 0.85); C$^{\gamma 2}$, 19.7 (0.64); C$^\delta$, 12.6 (0.17)
G18	112.7 (8.00)	43.6 (4.48, 3.79)		
G19	107.2 (8.58)	46.6 (3.97, 3.80)		
L20	115.0 (7.53)	54.2 (4.06)	42.9 (1.29, 0.96)	C$^\gamma$, 27.0 (1.34); C$^{\delta 1}$, 26.6 (0.51); C$^{\delta 2}$, 23.9 (0.67)
S21	114.4 (9.14)	57.6 (4.16)	63.1 (3.96, *)	
F22	128.0 (9.11)	60.3 (4.35)	37.6 (3.35, 3.10)	
E23	115.8 (8.33)	56.7 (4.29)	30.0 (2.15, 1.84)	C$^\gamma$, 36.6 (2.22, *)
T24	118.5 (7.32)	64.9 (4.15)	68.7 (4.02)	C$^\gamma$, 23.7 (1.22)
T25	119.1 (8.46)	59.3 (4.52)	73.1 (4.78)	C$^\gamma$, 21.7 (1.34)
D26	120.1 (8.84)	58.2 (3.95)	40.0 (2.80, 2.65)	
E27	116.2 (8.47)	60.3 (4.02)	29.2 (2.12, 1.99)	C$^\gamma$, * (*, *)
S28	120.2 (8.42)	61.7 (4.34)	62.2 (3.91, *)	
L29	126.5 (8.54)	58.2 (4.10)	42.6 (1.75, 1.18)	C$^\gamma$, 26.8 (1.54); C$^{\delta 1}$, 26.1 (0.61); C$^{\delta 2}$, 25.3 (0.97)
R30	121.2 (8.07)	59.7 (3.89)	30.7 (2.27, 1.92)	C$^\gamma$, 28.1 (1.46, *); C$^\delta$, 43.3 (3.46, 3.13)
S31	112.6 (8.36)	61.3 (4.16)	62.9 (4.01, 3.97)	
H32	114.7 (7.43)	60.1 (4.44)	29.3 (3.32, 3.16)	
F33	113.2 (8.29)	64.2 (4.57)	38.5 (3.69, 2.84)	
E34	124.3 (9.30)	58.0 (4.99)	28.7 (2.28, 2.16)	C$^\gamma$, 36.5 (2.59, 2.44)
Q35	116.4 (6.98)	58.0 (3.76)	27.6 (0.70)	C$^\gamma$, 34.3 (2.01, 1.60); N$^\epsilon$, 110.8 (7.03, 6.77)
W36	114.3 (7.05)	61.6 (4.51)	29.0 (3.60, 3.08)	N$^\epsilon$, 130.9 (10.16)
G37	104.8 (7.34)	45.5 (4.56, 3.97)		
T38	116.6 (8.67)	63.0 (4.27)	69.9 (4.18)	C$^\gamma$, 21.5 (1.18)
L39	126.3 (8.87)	53.0 (4.98)	41.8 (1.68, 1.14)	C$^\gamma$, 26.9 (1.59); C$^{\delta 1}$, 28.4 (0.85); C$^{\delta 2}$, 24.5 (0.15)
T40	117.0 (8.95)	62.8 (4.23)	67.7 (*)	C$^\gamma$, * (*)
D41	118.4 (7.53)	54.2 (4.66)	44.3 (2.61, 2.41)	
C42	123.4 (8.50)	59.5 (4.88)	26.4 (2.81, 2.68)	
V43	126.4 (8.77)	60.2 (4.62)	36.6 (1.81)	C$^{\gamma 1}$, 21.2 (0.89); C$^{\gamma 2}$, * (*)
V44	125.2 (8.31)	62.5 (3.75)	33.1 (1.84)	C$^{\gamma 1}$, 22.7 (0.58); C$^{\gamma 2}$, 20.6 (0.71)
M45	126.6 (8.48)	54.8 (4.48)	30.4 (1.45, 1.05)	C$^\gamma$, 33.3 (2.60, 2.29)
R46	120.3 (8.36)	54.3 (4.87)	34.0 (1.40, 1.20)	C$^\gamma$, * (*, *); C$^\delta$, * (*, *)
D47	124.9 (8.62)	52.1 (4.73)	42.9 (3.05, 2.56)	
P48	– (–)	64.6 (4.34)	32.3 (2.33, 1.96)	C$^\gamma$, 27.2 (2.11, 2.03); C$^\delta$, 50.9 (3.98, 3.87)
N49	115.2 (8.59)	55.4 (4.77)	39.5 (2.97, 2.85)	N$^\delta$, 114.0 (7.93, 7.07)
T50	108.5 (8.33)	62.1 (4.32)	70.7 (4.37)	C$^\gamma$, 20.9 (1.22)
K51	114.3 (8.12)	58.0 (3.95)	29.3 (2.21, 2.08)	C$^\gamma$, 25.2 (1.37, 1.30); C$^\delta$, 28.6 (1.64, 1.59); C$^\epsilon$, 42.4 (3.00)
R52	120.1 (7.89)	56.3 (4.20)	31.1 (1.80, 1.64)	C$^\gamma$, 27.8 (1.68, 1.45); C$^\delta$, 43.2 (3.15, *)
S53	116.6 (8.86)	58.6 (4.21)	63.8 (4.04, 3.85)	
R54	123.2 (9.02)	56.1 (4.43)	31.4 (2.25, 0.93)	C$^\gamma$, * (*, *); C$^\delta$, * (*, *)
G55	107.1 (9.42)	45.7 (4.14, 3.13)		
F56	112.1 (7.07)	54.9 (5.23)	41.8 (3.19, 2.73)	
G57	106.5 (8.37)	45.3 (4.18, 3.95)		
F58	114.8 (8.93)	56.5 (5.69)	44.1 (2.87, 2.60)	
V59	114.9 (8.59)	59.5 (4.58)	34.7 (1.71)	C$^{\gamma 1}$, 22.3 (0.54); C$^{\gamma 2}$, 20.1 (0.32)
T60	119.8 (8.21)	60.9 (5.12)	70.2 (3.77)	C$^\gamma$, 22.6 (1.08)
Y61	124.9 (8.41)	58.4 (5.38)	42.3 (3.77, 2.80)	
A62	121.2 (8.77)	55.3 (4.09)	19.1 (1.68)	
T63	104.8 (7.99)	59.1 (4.94)	73.2 (4.73)	C$^\gamma$, 21.6 (1.22)
V64	120.6 (8.99)	65.7 (3.70)	31.6 (2.06)	C$^{\gamma 1}$, 21.8 (0.96); C$^{\gamma 2}$, • (*)
E65	120.5 (8.72)	60.4 (4.08)	28.9 (2.11, 1.99)	C$^\gamma$, 36.8 (2.47, 2.37)
E66	120.4 (7.51)	59.3 (3.90)	29.2 (2.31, 1.78)	C$^\gamma$, 36.7 (2.44, 2.33)
V67	118.1 (6.91)	66.4 (3.19)	31.2 (2.45)	C$^{\gamma 1}$, 24.4 (1.29); C$^{\gamma 2}$, 22.3 (1.02)
D68	119.0 (8.15)	57.4 (4.16)	39.9 (2.76, 2.55)	
A69	122.6 (7.76)	55.2 (4.11)	18.2 (1.73)	
A70	121.1 (8.14)	55.4 (3.00)	18.2 (1.40)	
M71	117.7 (8.72)	56.3 (4.44)	30.7 (2.24, 2.04)	C$^\gamma$, 33.1 (2.56, 2.50)
N72	118.6 (8.06)	54.3 (4.66)	38.6 (2.88, 2.80)	N$^\delta$, 112.0 (7.54, 6.90)
A73	122.3 (6.98)	51.3 (4.36)	19.0 (0.92)	
R74	116.7 (6.94)	54.9 (3.71)	30.4 (1.86)	C$^\gamma$, 27.3 (1.84, *); C$^\delta$, 43.6 (3.37, *)
P75	– (–)	61.8 (4.50)	35.1 (2.41, 1.89)	C$^\gamma$, 31.1 (1.97, *); C$^\delta$, 50.1 (3.63, 3.49)

Table 1 (Continued)

residue	N	C^{α}	C^{β}	other
H76	121.2 (9.46)	56.2 (4.56)	32.1 (2.99, 2.38)	
K77	125.0 (8.51)	54.7 (5.01)	33.5 (1.61, 1.48)	C^{γ}, 25.4 (1.06, *); C^{δ}, 29.4 (1.58, *); C^{ϵ}, 41.6 (2.81)
V78	123.9 (8.66)	60.8 (4.01)	33.4 (1.76)	$C^{\gamma 1}$, 21.7 (0.76); $C^{\gamma 2}$, 21.0 (0.19)
D79	128.4 (9.59)	56.0 (4.27)	39.8 (2.79, 2.47)	
G80	101.4 (8.54)	45.5 (4.19, 3.49)		
R81	120.0 (7.58)	53.2 (4.68)	32.5 (1.82, 1.65)	C^{γ}, 26.1 (1.60, *); C^{δ}, 42.7 (3.21, *)
V82	124.5 (8.69)	63.2 (4.53)	31.0 (2.00)	$C^{\gamma 1}$, 22.0 (0.96); $C^{\gamma 2}$, * (*)
V83	122.5 (8.44)	59.6 (4.73)	32.8 (2.22)	$C^{\gamma 1}$, 22.5 (0.82); $C^{\gamma 2}$, 19.2 (0.70)
E84	120.9 (8.74)	53.0 (5.12)	31.4 (2.02, 1.82)	C^{γ}, 35.7 (2.19, 2.15)
P85	– (–)	61.6 (5.34)	32.8 (2.07, 1.86)	C^{γ}, * (2.18, 1.69); C^{δ}, 51.0 (4.03, 3.98)
K86	120.8 (9.40)	55.0 (4.70)	36.7 (1.92, 1.79)	C^{γ}, 24.6 (1.52); C^{δ}, 29.3 (1.66, *); C^{ϵ}, 42.0 (2.89)
R87	122.6 (8.88)	57.3 (4.39)	30.3 (2.04, 1.69)	C^{γ}, 27.1 (1.59, *); C^{δ}, 43.3 (3.31, 3.18)
A88	126.0 (8.46)	53.7 (4.30)	18.9 (1.49)	
V89	115.4 (7.94)	61.6 (4.21)	33.3 (2.06)	$C^{\gamma 1}$, 21.1 (0.89); $C^{\gamma 2}$, 20.2 (0.89)
S90	118.6 (8.23)	58.2 (4.46)	64.0 (3.85, *)	
R91	123.3 (8.36)	56.2 (4.33)	31.0 (1.87, 1.71)	C^{γ}, 26.9 (1.59, *); C^{δ}, 43.2 (3.10, *)
E92	127.0 (8.00)	58.0 (4.11)	31.1 (2.03, 1.86)	C^{γ}, * (*)

a In each column, ^{15}N and ^{13}C shift precede the corresponding ^{1}H shift give in parentheses. ^{1}H and ^{13}C chemical shifts are reported relative to 3-(trimethylsilyl)propionic-d_4 acid and ^{15}N shifts relative to external liquid NH$_3$. An asterisk indicates that the chemical shift was not determined.

duction for an additional 16 h. Cells were harvested by centrifugation and lysed by sonication. Cellular debris was removed by centrifugation prior to passage of the crude cell extract at a flow rate of 1 mL/min over a DE52 (Whatman) column (4.5 × 20 cm) equilibrated against 20 mM phosphate (pH 6.0), 0.1 mM EDTA, 0.1 mM dithiothreitol (DTT), and 0.1 mM phenylmethanesulfonyl fluoride (PMSF) followed by affinity chromatography on single-stranded DNA–cellulose to yield recombinant purified A1 (1–195). The purified protein was digested with *Staphyloccocus aureus* protease V8 (Pierce), which was immobilized via covalent attachment to an Affigel 10 resin (Bio-Rad). Digests were stopped by ultrafiltration of the solution to remove the protease-containing resin, followed by the addition of diisopropyl phosphorofluoridate. The released peptides 1–92 and 93–195 were separated over a Mono-S column. The partially pure 1–92 protein was resuspended in 4 mL of H$_2$O and loaded onto a G-50 gel filtration column (1.8 × 60 cm) at a flow rate of 0.1 mL/min. The 1–92 protein elutes at approximately 56 mL and is more than 95% free of protein or nucleic acid contaminants as judged by a combination of SDS-PAGE and relative absorbance at 280 and 260 nm. Incorporation of label was monitored using mass spectrometry of protein produced with and without [^{15}N]NH$_4$Cl/[^{13}C$_6$]D-glucose as sole nitrogen and carbon sources (W. M. Keck Foundation Biotechnology Resource Laboratory at Yale University). Molecular masses of the ^{15}N/^{13}C-labeled protein versus unlabeled protein suggested that the incorporation was between 96% and 98% with a theoretical limit of 98.8% based on the reported atom percent from the [^{15}N]NH$_4$Cl and [^{13}C$_6$]D-glucose manufacturers. The pooled 1–92 protein was then exhaustively dialyzed versus 0.5 mM phosphate (pH 6.0) prior to being lyophilized for storage until use in NMR studies. Samples for NMR contained 2 mM ^{15}N- or ^{15}N/^{13}C-labeled A1 (1–92) (pH 5.7), dissolved in either 90% H$_2$O/10% D$_2$O or 99.996% D$_2$O, respectively.

NMR Spectroscopy. All NMR spectra were recorded at 25 °C on a Bruker AM600 spectrometer equipped with a triple-resonance ^{1}H, ^{15}N, ^{13}C probe. The following 3D spectra were recorded: CBCANH (Grzesiek & Bax, 1992a), CBCA-(CO)NH (Grzesiek & Bax, 1992b), HBHA(CO)NH (Grzesiek & Bax, 1993), HNHA (Vuister & Bax, 1993), ^{15}N-separated HOHAHA (Clore et al., 1991), ^{15}N-separated NOESY (Marion et al., 1989), ^{13}C-separated NOESY (Ikura et al., 1990), HCCH-COSY (Bax et al., 1990a), and HCCH-TOCSY (Bax et al., 1990b).

Quantitative $^{3}J_{\mathrm{HN\alpha}}$ couplings were obtained from a 3D HNHA spectrum as described by Vuister and Bax (1993).

Slowly exchanging NH protons were identified by recording a series of ^{15}N–^{1}H Overbodenhausen correlation spectra (Bodenhausen & Ruben, 1980; Bax et al., 1990c; Norwood et al., 1990) over a period of ~24 h after dissolving an unexchanged sample of lyophilized protein in D$_2$O.

All spectra were processed on a Sun Sparc Workstation using in-house routines for Fourier transformation (F. Delaglio, unpublished results) and linear prediction (Zhu & Bax, 1990). Analysis of the 3D spectra and peak picking was carried out using the in-house programs CAPP and PIPP (Garrett et al., 1991), and PEAK-SORT (R. Powers, unpublished results).

RESULTS AND DISCUSSION

General Assignment Strategy. Sequential assignment was accomplished by means of double- and triple-resonance 3D NMR experiments (Ikura et al., 1990; Clore & Gronenborn, 1991; Bax & Grzesiek, 1993). Specifically, the CBCANH experiment was used to establish intraresidue $C^{\beta}(i)$, $C^{\alpha}(i)$–N(i)–NH(i) and interresidue $C^{\beta}(i-1)$,$C^{\alpha}(i-1)$–N(i)–NH(i) correlations which could then be unambiguously distinguished by means of the CBCA(CO)NH experiment which only detects the interresidue $C^{\beta}(i-1)$,$C^{\alpha}(i-1)$–N(i)–NH(i) correlations. The HBHA(CO)NH experiment was then used to establish C_{β}H($i-1$),C^{α}H($i-1$)–N(i)–NH(i) correlations which were complemented by the intraresidue C^{α}H(i)–N(i)–NH(i) correlations detected in the HNHA and ^{15}N-separated HOHAHA experiments. Finally, side chain spin systems were identified using the HCCH-COSY and HCCH-TOCSY experiments to establish intraresidue H(j)–C(j)–C($j\pm1$)–H($j\pm1$) and H(j)–C(j)–C($j\pm n$)–H($j\pm n$) correlations, respectively. By this means we were able to obtain essentially complete ^{1}H, ^{15}N, and ^{13}C assignments which are listed in Table 1. Examples of the quality of the data are provided by the ^{1}H–^{15}N correlation spectrum shown in Figure 1A, and by selected planes of the CBCA(CO)NH and CBCANH experiments illustrated in Figure 1B.

Secondary Structure Determination. Elements of regular secondary structure can be reliably delineated from a qualitative analysis of NOE data involving the backbone protons, together with information on NH exchange rates and $^{3}J_{\mathrm{HN\alpha}}$ coupling constants (Wüthrich, 1986; Clore & Gronenborn, 1989). Specifically, a stretch of consecutive sequential NH-(i)–NH(i+1) NOEs in conjunction with C^{α}H(i)–NH-
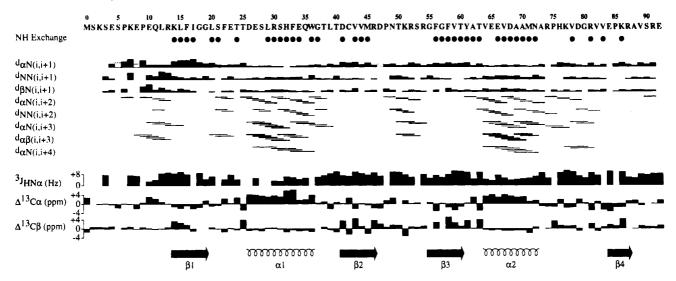
0   5   10   15   20   25   30   35   40   45   50   55   60   65   70   75   80   85   90
MSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHKVDGRVVEPKRAVSRE

NH Exchange

$d_{\alpha N(i,i+1)}$
$d_{NN(i,i+1)}$
$d_{\beta N(i,i+1)}$
$d_{\alpha N(i,i+2)}$
$d_{NN(i,i+2)}$
$d_{\alpha N(i,i+3)}$
$d_{\alpha\beta(i,i+3)}$
$d_{\alpha N(i,i+4)}$

$^3J_{HN\alpha}$ (Hz)

$\Delta^{13}C\alpha$ (ppm)

$\Delta^{13}C\beta$ (ppm)

β1   α1   β2   β3   α2   β4

FIGURE 2: Summary of the sequential and medium-range NOEs involving the NH, C$^\alpha$H, and C$^\beta$H protons, the amide exchange and $^3J_{HN\alpha}$ coupling constant data, and the $^{13}C^\alpha$ and $^{13}C^\beta$ secondary chemical shifts observed for A1. The thickness of the lines reflects the strength of the NOEs. Amide protons which are still present after 10 h after taking up a lyophilized sample in D_2O are indicated by closed circles. The secondary structure deduced from the data is indicated at the bottom of the figure.

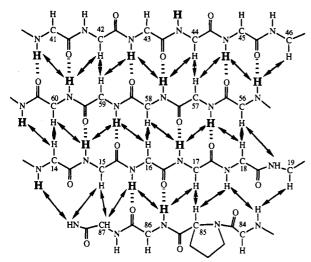FIGURE 3: Antiparallel β-sheet structure of A1 (1–92) as derived from the NOE, as well as the amide exchange data. Interstrand NOEs are indicated by arrows, and slowly exchanging amide protons are shown in bold print. Hydrogen bonds compatible with the NOE and exchange data are indicated by broken lines.

(i+2,3,4), NH(i)–NH(i+2), and C$^\alpha$H(i)–C$^\beta$H(i+3) NOEs and $^3J_{HN\alpha}$ coupling constants of less than 6 Hz are characteristic of an α-helix. A stretch of strong sequential C$^\alpha$H-(i)–NH(i+1) NOEs in combination with interstrand NOEs involving the NH and C$^\alpha$H protons, as well as $^3J_{HN\alpha}$ coupling constants of >8 Hz, is characteristic of a β-sheet. Slowly exchanging amide protons are indicative of hydrogen bonding and thus provide further evidence on the presence and nature of the secondary structure. A summary of the sequential NOE data, NH exchange, $^3J_{HN\alpha}$ coupling constants, and $^{13}C^{\alpha/\beta}$ chemical shift information is presented in Figure 2. NOEs were identified from 3D ^{15}N- and ^{13}C-separated NOESY spectra, and quantitative $^3J_{HN\alpha}$ coupling constants were obtained from a 3D HNHA spectrum. In addition, deviations of $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts from their random coil values are diagrammed at the bottom of the figure. These secondary ^{13}C chemical shifts are strongly correlated with secondary structure elements such that residues located in α-helices exhibit positive secondary shift values for C$^\alpha$ resonances and negative ones for C$^\beta$ resonances, while those in β-sheets exhibit negative ones for C$^\alpha$ resonances and positive ones for C$^\beta$

resonances (Spera & Bax, 1991; Wishart et al., 1991). Qualitative interpretation of all the data presented in Figure 2 clearly establishes the secondary structure for the A1 domain: $\beta\alpha\beta\beta\alpha\beta$. This folding pattern is essentially the same as that found for U1A by X-ray crystallography (Nagai et al., 1990) and NMR (Hoffman et al., 1991) and the RNA binding domain of hnRNP C (Wittekind et al., 1992; Görlach et al., 1992). In this regard it is interesting to point out that the correct topological fold for the second RNA binding domain of A1 was predicted by model building based on the known three-dimensional structure of acylphosphatase (Ghetti et al., 1990).

The four β-strands in the A1 (1–92) domain form an antiparallel β-sheet. The alignment of the strands within the sheet as well as the hydrogen-bonding pattern is illustrated in Figure 3. Again, the central part of the β-sheet is very similar to the one found in hnRNP C (Wittekind et al., 1992) and U1A (Nagai et al., 1990). In contrast to hnRNP C where β-strands 2 and 3 are connected by a β-hairpin, the A1 domain possesses an eight amino acid long insertion at this location (residues 47–54). This stretch of the polypeptide chain is not part of the β-sheet but forms an irregular loop, possibly containing a small segment of the 3–10 helix (residues 49–54) characterized by NH(i)–NH(i+1), NH(i)–NH(i+2), C$^\alpha$H-(i)–NH(i+2,3), and C$^\alpha$H(i)–C$^\beta$H(i+3) NOE connectivities (cf. Figure 2). Interestingly, the U1A domain also contains an insertion at this location, albeit a slightly shorter one of five amino acids (Nagai et al., 1990), and indeed, the most extensive length variability leading to gaps in the alignment of RNA recognition domains is found in this region.

The two α-helices comprise residues 24–36 (α1) and 63–72 (α2). These two α-helices are also very similar in length and location to those found in hnRNP C (Wittekind et al., 1992) and U1A (Nagai et al., 1990). Although α1 in hnRNP C has been reported to only extend from Lys-29 to Phe-37 (Wittekind et al., 1992), it could well be that it is actually one turn longer, extending up to Tyr-40. This suggestion is based on the reported secondary ^{13}C chemical shifts (Wittekind et al., 1992) as well as the fact that this would make all three proteins structurally more alike. It is intriguing to note that both helices contain amino acids at their N-terminal end which have been proposed to constitute capping boxes (Harper & Rose, 1993), such as TXXE or TXXS for helix α1 and TXXE for helix α2.

```
                    0     5    10    15    20    25    30    35       40    45    50    55    60    65    70    75    80    85    90
hnRNP-A1(1-92)    MSKSESPK-EP-EQLRKLFIGGLSFE-TTDESLRSHFEQWGTL···TDCVVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHK-VDGRVVEPKRAVSRE

hnRNP-A1(93-183)  --DSQRPG-AH-LTVKKIFVGGIKED-TEEHHLRDYFEQYGKI···EVIEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHT-VNGHNGEVRKALSKQ

hnRNP-C1          MASNVTNKTDPRSMNSRVFIGNLNTLVVKKSDVEAIFSKYGKI···VGCSVHK········GFAFVQYVNERNARAAVAGEDGRMIAGQVLDINLAAEPK

U1-snRNP A        MAVPETRPNHTIYINNLNEKIKKDELKKSLYAIFSQFGQILDILVSRSL···KMRGQAFVIFKEVSSATNALRSMQGFPFYDKPMRIQYAKTDS
```

FIGURE 4: Amino acid sequence alignment for hnRNP A1 (1–92), hnRNP C1, and snRNP U1A based on the conserved secondary structure elements. Also included is the sequence of the second RNA binding domain of A1, namely, the hnRNP A1 (93–183) domain.

Inspection of the pattern of secondary ^{13}C chemical shifts for the C^α and C^β resonances of residues at the beginning of the helices reveals the interesting observation that one residue exhibits a marked negative deviation for the C^α chemical shift and a positive deviation for the C^β chemical shift compared to the respective random coil values. This striking pattern of secondary chemical shifts arises from the relationship between ϕ and ψ dihedral angles and the C^α and C^β chemical shifts (Spera & Bax, 1991), and it is mainly the unusual ψ value of 167° ± 5° found for N-cap residues (Harper & Rose, 1993) that gives rise to this pattern. Indeed, it is possible to identify such capping boxes from inspection of secondary C^α and C^β chemical shift values (A. M. Gronenborn and G. M. Clore, unpublished observation). The pattern in helix $\alpha 1$ clearly identifies Thr-25 as the N-cap residue although from the sequence Thr-24 would have been equally possible.

Structure-based alignment of the amino acid sequences for the three RNP proteins whose secondary structures and topologies are available is presented in Figure 4. This alignment differs from a previous one which was mainly based on the U1A structure (Kenan et al., 1991). In particular, residues in β-strands 2 and 4 are shifted in the A1 (1–92) sequence. Thus, Cys-42 in A1 (1–92) is equivalent to Ile-43 in U1A, and Pro-85 in A1 (1–92) corresponds to Ile-84 in U1A. Also included in Figure 4 is the sequence of the A1 (93–183) RNA binding domain which has 32% sequence identity to the A1 (1–92) domain and can be aligned without deletions or insertions. On the basis of the available secondary structure information for A1 (1–92) presented here, it is clear that the three-dimensional structure will be very similar to that of U1A (Nagai et al., 1990) and the hnRNP C domain (Wittekind et al., 1992). All the major hydrophobic core residues are either conserved or conservatively substituted, resulting in almost identical packing of the secondary structure elements. Major differences exist only in the loop regions. Loop 1 which connects β-strand 1 with helix $\alpha 1$ is one residue shorter in the A1 domain than in the two other proteins, whereas loop 2 which connects $\alpha 1$ with β-strand 2 is three residues shorter in both the A1 (1–92) and hnRNP C domains relative to U1A. The largest deviation occurs in loop 3 between β-strands 2 and 3; in the hnRNP C domain both β-strands are connected by a β-hairpin, so that no loop is present. In the A1 (1–92) domain, on the other hand, an eight amino acid insertion is present. In the U1A structure the loop is of intermediate length, comprising five amino acids. The last loop connecting helix $\alpha 2$ with β-strand 4 in the A1 (1–92) domain is again one residue shorter than in the other two proteins.

On the basis of structural and mutational data for the U1A system, a model has been suggested for the interaction between the protein and nucleic acid (Nagai, 1992). A general feature of this and other proposals (Görlach et al., 1992) is based on the notion that the four-stranded β-sheet containing the RNP

sequence motifs constitutes a platform for general RNA binding. In addition, basic amino acids within the loop regions are also thought to play a major role (Jessen et al., 1991). For the A1 (1–92) domain the involvement of the β-sheet region in RNA binding has been clearly established, given the fact that the solvent-exposed aromatic side chains of Phe-15 and Phe-57, both of which are located in the RNP motifs, could be cross-linked to d(T)$_8$ (Merrill et al., 1988). Preliminary NMR evidence also identifies the solvent-exposed side of the β-sheet, as well as the N- and C-terminal arms, as being in contact with the RNA (D. S. Garrett, P. J. Lodi, Y. Shamoo, K. R. Williams, G. M. Clore, and A. M. Gronenborn, unpublished observation). Any detailed description of the molecular interactions between the RNA binding domains and RNA will have to await the completion of the three-dimensional structures of such complexes by either X-ray crystallography or NMR. The latter is under investigation for the A1 domain in our laboratory.

REFERENCES

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., & 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., & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131–138.

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

Bax, A., Clore, G. M., & Gronenborn, A. M. (1990b) *J. Magn. Reson. 88*, 425–431.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A., & Tschudin, R. (1990c) *J. Magn. Reson. 86*, 304–318.

Beyer, A. L., Christensen, M. E., Walker, B. W., & LeStourgeon, W. M. (1977) *Cell 11*, 127–138.

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

Choi, Y. D., & Dreyfuss, G. (1984) *J. Cell. Biol. 99*, 1997–2004.

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

Clore, G. M., & Gronenborn, A. M. (1989) *CRC Crit. Rev. Biochem. Mol. Biol. 24*, 479–564.

Clore, G. M., & Gronenborn, A. M. (1991) *Prog. Nucl. Magn. Reson. Spectrosc. 23*, 43–92.

Clore, G. M., Bax, A., & Gronenborn, A. M. (1991) *J. Biomol. NMR 1*, 13–22.

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. H. (1988) *J. Biol. Chem. 263*, 1063–1071.

Dreyfuss, G., Matunis, M. J., Pinõl-Roma, S., & Burd, C. G. (1993) *Annu. Rev. Biochem. 62*, 289–321.

Garrett, D. S., Powers, R., Gronenborn, A. M., & Clore, G. M. (1991) *J. Magn. Reson. 95*, 214–220.

Ge, H., & Manley, J. L. (1990) *Cell 62*, 25–34.

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.

Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 99*, 201–207.

Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291–6293.

Grzesiek, S., & Bax, A. (1993) *J. Biomol. NMR 3*, 185–204.

Harper, E. T., & Rose, G. D. (1993) *Biochemistry 32*, 7605–7609.

Harrison, S. C. (1991) *Nature 353*, 715–719.

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., Kay, L. E., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 86*, 204–209.

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

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

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

Mayrand, S., Setyono, B., Gronenberg, J. R., & Pederson, T. (1981) *J. Cell Biol. 90*, 380–384.

Merrill, B. M., LoPresti, M. B., Stone, K. L., & Williams, K. R. (1986) *J. Biol. Chem. 261*, 878–883.

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

Nagai, K. (1992) *Curr. Opin. Struct. Biol. 2*, 131–137.

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

Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990) *J. Magn. Reson. 87*, 488–510.

Pabo, C. O., & Sauer, R. T. (1992) *Annu. Rev. Biochem. 61*, 1053–1095.

Piñol-Roma, S., & Dreyfuss, G. (1992) *Nature 355*, 730–732.

Sachs, A. B., Bond, M. W., & Kornberg, R. D. (1986) *Cell 45*, 827–835.

Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490–5492.

Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205–280.

Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc.* (in press).

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

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

Wolberger, C. (1993) *Curr. Opin. Struct. Biol. 3*, 3–10.

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

Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405–410.