Two RNA-binding Domains Determine the RNA-binding Specificity of Nucleolin*

Guillaume Serin, Gérard Joseph, Laurence Ghisolfi, Marielle Bauzan, Monique Erard, François Amalric, and Philippe Bouvet‡

From the Laboratoire de Biologie Moléculaire Eucaryote, Institut de Biologie Cellulaire et de Génétique du CNRS, UPR 9006, 118 route de Narbonne, 31062 Toulouse Cedex, France

Nucleolin is an abundant nucleolar RNA-binding protein that seems to be involved in many aspects of ribosome biogenesis. Nucleolin contains four copies of a consensus RNA-binding domain (CS-RBD) found in several other proteins. In vitro RNA-binding studies previously determined that nucleolin interacts specifically with a short RNA stem-loop structure. Taken individually, none of the four CS-RBDs interacts significantly with the RNA target, but a peptide that contains the first two adjacent CS-RBDs (R12) is sufficient to account for nucleolin RNA-binding specificity and affinity. The full integrity of these two domains is required, since N- or C-terminal deletion abolishes the specific interaction with the RNA. Mutation of conserved amino acids within the RNP-1 sequence of CS-RBD 1 or 2 drastically reduces the interaction with the RNA, whereas mutation of the analogous residues in CS-RBDs 3 and 4 has no effect in the context of the R1234G protein (which corresponds to the C-terminal end of nucleolin). Our results demonstrate that nucleolin RNA-binding specificity is the result of a cooperation between two CS-RBDs (RBDs 1 and 2) and also suggests a direct or indirect involvement of the RNP-1 consensus sequence of both CS-RBDs in the recognition of the RNA target.

Many aspects of the post-transcriptional regulation of gene expression involve specific RNA-protein interactions (1). Selective RNA recognition by RNA-binding proteins is achieved through a variety of conserved protein motifs (2). One of the most well characterized RNA-binding motifs is the ribonucleoprotein consensus sequence, also called RNA recognition motif or consensus RNA-binding domain (CS-RBD). The CS-RBD is found in proteins implicated in heterogeneous RNA processing (3), pre-mRNA splicing (4, 5), as components of pre-ribosomal proteins (6) in poly(A) tail synthesis and maturation (7, 8), translational control (9), and mRNA stability (10).

Despite the limited number of well characterized RNA binding sites for CS-RBD-containing proteins, two classes can be distinguished with respect to their RNA target structures. In the first class, CS-RBDs interact specifically with single-stranded RNA sequences. These proteins include hnRNP A1 (11), hnRNP C1 (12), poly(A)-binding protein (13), splicing factors ASF/SF2, SC35 (14), and sex-lethal protein (15, 16). The minimal single-stranded RNA required ranges from 5 nucleotides (poly(A)-binding protein (PABP); Ref. 13) to 10 nucleotides (ASF; Ref. 14) as determined by in vitro selection-amplification procedure (SELEX). In the second class, CS-RBDs interact with RNA stem-loop structures and include U1-snRNP A protein (17, 18), U2-snRNP B″ protein (19), U1-snRNP (70 kDa) (20), and nucleolin (21).

A typical CS-RBD contains 80–90 amino acid residues with two highly conserved sequences, the RNP-1 octapeptide (R/K/G/F/Y/G/A/F/Y/NX/F/Y) and the RNP-2 (L/I)(F/Y)(V/I)(G/K/G/N)L hexapeptide motifs (22–24). Crystal structure determination (25) and NMR spectroscopy (26) of the U1A-snRNP A protein (17, 18) demonstrated that this domain is constituted of four antiparallel β-strands packed against two perpendicularly oriented α-helices. This structure appears to be well conserved in other members of the CS-RBD gene family (27–30). The RNP-1 and RNP-2 motifs are located in two adjacent β-strands, and conserved amino acids within these two motifs have been shown to be implicated in direct nucleotide contacts (31–34). This was conclusively demonstrated with the crystal structure of the U1A protein complexed with a variant of hairpin II of U1 snRNA (35). The 10-nucleotide RNA loop binds to the surface of the β-sheet and interacts extensively with the conserved RNP-1 and RNP-2 motifs and with the C-terminal domain of the protein (31, 36–38). However, since the RNP-1 and RNP-2 motifs are very conserved in a large number of CS-RBD proteins, they probably do not distinguish between different RNA target sequences; thus, determinants of RNA-binding specificity must reside in the more variable region or outside the CS-RBD motif per se (39). Indeed, RNA-binding specificity determinants were found in the sequence adjacent to the CS-RBD for hnRNP C1, U1-snRNP (70 kDa), and Tra2 proteins (12, 22, 40), or in the CS-RBD variable loop 3 region of the U1 snRNP A and U2-snRNP B″ proteins (17, 19, 41).

Although the structure of CS-RBD motifs seems to be highly conserved, a general mechanism of how they interact with specific RNA sequences or structures cannot be drawn, and detailed studies of the interaction of other CS-RBD motifs with their cognate RNA are needed. An interesting situation is found with proteins that contain multiple CS-RBDs like Sxl (16), ASF/SF2 (14, 42), hnRNP A1 (11), and the poly(A)-binding protein (43, 44). For these proteins, the RNA-binding specificity of the full-length protein seems to require the cooperation of several CS-RBD domains.

Nucleolin, which contains four CS-RBDs (45–47), is thought to be a multifunctional protein implicated in the control of pre-ribosomal RNA synthesis and processing (48, 49), in transcriptional repression (50), as a nuclear matrix attachment.
region DNA-binding protein (51), and in packaging of nascent pre-ribosomal RNA (21, 46, 52). We showed previously that mouse and human nucleolin interact specifically with pre-rRNA and with *in vitro* selected RNAs that contain a hexanucleotide motif U/GCCCGA within a short stem-loop structure (21). In this study, we determine the minimal domain of nucleolin responsible for this specific interaction. We show that two CS-RBDs (CS-RBDs 1 and 2) are necessary and sufficient to account for the specific interaction of nucleolin with its RNA target. A recombinant protein corresponding to these 2 domains has the same binding properties as full-length nucleolin. Mutagenesis of the conserved amino acid residues within the RNP-1 motifs of domain 1 or 2, or deletion at either the N- or C-terminal extremities of domain 1 and 2, respectively, abolished the specific interaction with the RNA target. Furthermore, isolated CS-RBDs 1 and 2 possess no significant binding activities. Our results demonstrate that the RNA-binding specificity of nucleolin arises from a cooperation between two CS-RBDs.

**MATERIALS AND METHODS**

**Oligonucleotides**—The sequences were as follows: R1, CCCCATATGCACTTTAGGAAA; R2, CCCCATATGACACTTTAGGAAA; R3, CCCCATATGCACTTTAGGAAA; R4, CCCCATATGCACTTTAGGAAA; R1(D,D)S, TATATTATTGAATTTAAGTC; R1(D,D)NS, AATTATTTAAGCAAGCCCTTTACTCTTC- TGTACC; R2(L,Y)NS, AAT(T/A)AAAGCAA(G/T)CCCTTTACTCTTC- TGTACC; R2(L/I,L/F)NS, AAT(T/A)AAAGCAA(G/T)CCCTTTACTCTTC- TGTACC; R2(L/I,L/F)S, AAAGGG(C/A)TTGCTTT(T/A)ATTGAATTTA- TTTTGAGTC; R1(L/F,L/F)NS, CCAC(T/A)AAACC(T/A)AATTTCCTATT- GCTGATATTGAATTTAAGTC; R2(D,D)NS, AATATCAGCATCCCCT- TCAAGGGTCTG; Nco pET15b, TATACCATGGGCAGCAGCCATC; R1C, CCCGGATCCTGGTTTTT- TGAATCTGTTCATTGGAAAC; R2N, CCCCATATGACACTTTAGCA- TACTA; R12R4N, CGAATTGCTGGTTTTT- TGAATCTGTTCATTGGAAAC; RGGC, GAAAC; R3C, CCCGGATCCGGTACCTTGTAACTCCAACCTGAT; R4C, CCCGGATCCGGTACCGGCCCAGTCCAAAGTAAC; RGGC, GAAAC; R3C, CCCGGATCCGGTACCTTGTAACTCCAACCTGAT; R4C, CCCGGATCCGGTACCGGCCCAGTCCAAAGTAAC.

**Introduction of the RNP-1 mutations in the R12 peptide was achieved by PCR site-directed mutagenesis with the following oligonucleotides:**

- For Mut RBD 1, R3(L,L)NS and R3(L,L)S for 1LL and 2LL.
- For Mut RBD 4, R4(L,L)NS and R4(L,L)S oligonucleotides were used. All recombinant plasmids were sequenced to confirm the presence of the introduced mutations.

**Expression and Purification of Recombinant Proteins**—BL21(DE3) pLysS was transformed with each recombinant plasmid. Cells grown at 37 °C in LB (100 mg/liter ampicillin, 20 mg/liter chloramphenicol, and 1 g/liter methicillin) were induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 4 h. Harvested cells were suspended in buffer A (50 mM sodium phosphate, pH 8, 300 mM NaCl, 10 μg/ml lysozyme), sonicated, and centrifuged (30 min at 10,000 × g). The supernatant was recovered and dialyzed against ice-cold buffer C for 1 h at 4 °C and after addition of 1 μl of Ni^2+^-nitrilotriacetic acid resin (Qiagen)/ml of initial culture. After three washes with buffer A and two with buffer B (50 mM sodium phosphate, pH 6, 300 mM NaCl, 10% glycerol), tagged peptide was eluted with buffer C (buffer B + 0.5 M imidazole). The supernatant was applied on a Q-sepharose column (Pharmacia Biotech Inc.) equilibrated with 100 mM KCl and 10 mM Tris-HCl, pH 7.5. Concentrations were estimated with Bradford reagent (Bio-Rad protein assay) and checked by SDS-polyacrylamide gel electrophoresis.

**RNA Transcription**—The nucleolin recognition element (NRE) and non-specific (NS) sequences cloned between the XbaI and HindIII sites of pSP64PA plasmid (Promega) encode a 68-nucleotide-long RNA (21). 1 μg of plasmid linearized by HindIII was transcribed with 10 μCi of l-[ω-32P]CTP using T7 RNA polymerase (Promega), according to the instructions of the manufacturer. RNA was purified by ammonium acetate (3 M) precipitation, and its integrity was checked by electrophoresis on a 6% acrylamide, 8 M urea gel, and [ω-32P]CTP incorporation was quantified to estimate RNA concentration. A new RNA preparation was made every week or every other week. We found that a step of denaturation-renaturation does not improve or modify the binding affinity of the protein.

**RNA Binding Assay**—Filter binding assays were performed essentially as described previously (53). Nucleolin was prepared from Chinese hamster ovary cells (21). For gel retardation assays, 10 fmol of labeled RNA was incubated in 10 μl of TMKC buffer (20 mM Tris-HCl, pH 7.4, 4 mM MgCl₂, 200 μM KCl, 20% glycerol, 1 mM diithiothreitol, 0.5 mg/ml trNA, 4 μg/ml bovine serum albumin) with the indicated amount of protein for 15 min at room temperature. The mixture was loaded directly on a 8% polyacrylamide gel (acrylamide:bisacrylamide 19:1, on the basis of a gel volume) and run for 2 h at 100 V. The RNA bound to the protein was detected by autoradiography. Gel shift experiments were performed with at least two independent protein and RNA preparations.

**Circular Dichroic Measurements**—Circular dichroic spectra were recorded at 20 °C with a Jobin–Yvon dichrograph V1. A cell of 1 mm optical path length was used to record spectra of polypeptides in the far-ultraviolet region (190–260 nm) at a concentration of 0.2 mg/ml in TMKC buffer (20 mM Tris-HCl, pH 7.4, 4 mM MgCl₂, 0.2 mM KCl). The glycerol was omitted to avoid the formation of microbubbles that could interfere with spectra recording. We previously determined that glycerol concentration had no effect on the RNA-protein interaction. The results are presented as molellar ellipticity values in degrees·dmol⁻¹·cm⁻², on the basis of an amino acid mean residue mass of 110 Da. A cell of 1 cm optical path length was used to record spectra of RNA and polypeptide/ RNA complexes in the near-ultraviolet region (200–320 nm) at a concentration of 20 μg/ml in RNA in TMKC buffer. The results are presented as molar ellipticity values in degrees·cm²·dmol⁻¹, on the basis of a nucleotide mean residue mass of 330 Da. CD spectra were recorded with at least two different protein and RNA preparations.

**RESULTS**

**Identification of Nucleolin Domains Involved in Recognition of a Specific RNA Sequence**—In previous studies we demonstrated that nucleolin interacts specifically with a short RNA stem-loop (21, 54). To characterize this interaction, we expressed several nucleolin peptides in *Escherichia coli* as histidine-tagged proteins and purified them by affinity chromatography on Ni²⁺-nitrilotriacetic acid columns. After purification to near homogeneity (Fig. 1, A and B), these peptides were used for 2DD; R1(L/F,L/F)S and R1(L/F,L/F)NS for 1FF, 1FL, 1FL, and 1LL; R2 (L/F,L/F)S and R2(L/F,L/F)NS for 2IF, 2IL, 2FL, and 2LL; R1(L,Y)S and R1(L,Y)NS for 1LY; R2(L,Y)S and R2(L,Y)NS for 2LY. To generate the series of RNP-1 mutations within the R1234G protein, the same method of PCR site-directed mutagenesis with the corresponding mutagenizing oligonucleotides for Mut RBD 1 and 2, the same oligonucleotides were used as for the 1LL and 2LL mutants. For Mut RBD 3, R3(L,L)NS and R3(L,L)S, and for Mut RBD 4, R4(L,L)NS and R4(L,L)S oligonucleotides were used. All recombiant plasmids were sequenced to confirm the presence of the introduced mutations.
drastically reduced the binding affinity of nucleolin for this RNA (recognition. A single mutation, G to A, in the loop of the NS RNA interacts with the same binding affinity and specificity as full- sequence. Gel shift experiments indicate that the R1234G protein which contains the four RNA-binding motifs and the RGG tested if the C-terminal region of nucleolin (R1234G peptide), within the consensus selected sequence drastically reduces nucleolin with high affinity (21). A single mutation (G to A) at the C terminus of nucleolin and R1234G peptides sym- bolizes the RGG domain. In rat nucleolin, 12, 21, and 14 amino acid residues are located between RBDs 1–2, 2–3, and 3–4, respectively. Linker length and sequence are relatively well conserved between mouse, rat, hamster, and human, but differ substantially in chicken and Xenopus laevis nucleolin. B, Coomassie staining of purified nucleo- lin protein mutants. 1 μg of each protein was loaded on a 15% SDS-polyacrylamide gel. C, partial sequence of the 68-nucleotide NRE and NS RNAs. The NRE RNA was isolated by SELEX, and has been de- scribed in detail previously (21). The stem-loop structure represented in gel shift experiments with two RNAs (Fig. 1C). The NRE RNA is a 68-nucleotide-long RNA isolated by SELEX that binds nucleolin with high affinity (21). A single mutation (G to A) within the consensus selected sequence drastically reduces nucleolin interaction and gives rise to the NS RNA. We first tested if the C-terminal region of nucleolin (R1234G peptide), which contains the four RNA-binding motifs and the RGG domain, was able to interact specifically with the NRE sequence. Gel shift experiments indicate that the R1234G protein interacts with the same binding affinity and specificity as full- length nucleolin (K_d of 20–50 nM with the NRE sequence and 250 nM for the mutated (NS) RNA; Fig. 2, A and B). Deletion of the C-terminal region of this peptide (RGG domain) in the R1234 protein does not modify significantly the binding affinity toward the NRE RNA but significantly reduces the nonspecific interaction with the NS RNA (K_d of 1–2 μM). This is in agreement with other studies that show that this RGG domain is able to interact nonspecifically with RNA (53, 55).

We then performed a series of N- and C-terminal deletions on the R1234 protein, removing one RBD at a time (Fig. 1A). Deletion of the fourth or of the two last RBD domains gave rise to proteins (R123 and R12) that retained the ability to interact with high affinity with the specific RNA sequence (K_d of 20–50 nM) and not with the control NS RNA (K_d > 2.5 μM); conversely, deletion of the first (R234) or of the first two RBDs (R34, Fig. 2A) completely abolished the interaction with the NRE. These results indicate that the determinants of specificity must reside within the first two RBDs (R12). To determine which one of the two RBDs was responsible for this interaction, each isolated domain was tested for binding activity with the NRE sequence. Surprisingly, none of these two domains was able to interact significantly with the NRE RNA. Only at very high concentration (>10 μM) could a nonspecific interaction with both RNAs (NRE and NS) be observed with these single domains (data not shown). Isolated RBDs 3 and 4 were also unable to interact with the NRE or NS RNAs (K_d > 10 μM) (data not shown). These results therefore demonstrate that the RNA-binding specificity of nucleolin is contained within the first two RBDs.

Deletion Analysis of the R12 Peptide—The isolated RBDs described in Fig. 1 were designed according to the definition of a minimal RNA-binding domain (2, 22), which spans the begin- ning of the β1 strand to the end of β4. However, for the U1A, U1 (70 kDa), and hnRNP C proteins, the β4 strand C-terminal region is crucial to confer a high affinity and specificity of RNA binding (12, 22, 31, 56). To see if the C- or N-terminal regions of RBD 1 and RBD 2 were required to confer the RNA-binding specificity and affinity, several deletions were performed on the R12 peptide (Fig. 3). In the protein R12ΔN, the first 50 amino acids of RBD 1 have been deleted. This left the αB helix and β4 strand of RBD 1 intact at the N terminus of RBD 2. Gel shift experiments (Fig. 3B) with the NRE sequence clearly show that this protein is unable to interact with this RNA. In the R12C protein, the C-terminal region of RBD 2 (β3-αB-β4) has been deleted. This truncated protein is also unable to interact with high affinity with the NRE. However, at high concentration of protein, an RNA-protein complex could be detected with the NRE (>1 μM) and NS (>2.5 μM) RNAs (data not shown). This indicates that the region at the C terminus of RBD 1 is able to restore part of the RNA binding activity to the isolated domain. To determine if this interaction involved specific sequences within the second RBD, a new chimeric protein was con- structed (R14) where RBD 2 was replaced by nucleolin RBD 4. Gel shift analysis shows that R14 interacts with the NRE (Fig. 3B) and NS sequences (data not shown) only at high concentra- tion (K_d > 2.5 μM). This interaction is significantly lower that for the R12 peptide. Therefore, these results suggest that the region at the C-terminal end of RBD 1 (potentially the linker region between RBDs 1 and 2 or sequences within RBD 2) is important to confer a basal binding activity to RBD 1. However, the low binding affinity and the inability to discrim- inate between specific and nonspecific RNAs show that RBD 1 alone cannot account for the RNA-binding specificity of full- length nucleolin. Therefore, the integrity of both RBD 1 and RBD 2 is required for a specific and high affinity interaction with the NRE sequence.

Role of the RNP-1 Motifs of RBDs 1 and 2 in RNA Binding— The most characteristic feature of RBD proteins is the RNP-1
octamer motif, which contains conserved aromatic residues involved in RNA binding (2, 35, 39). The presence of these aromatic residues is absolutely required to confer an RNA binding activity to most of the RBDs (32–34, 58). To further investigate the role of the individual RBDs 1 and 2 in the specific interaction with the NRE sequence, we introduced a series of point mutations into the RNP-1 motif of these domains (Fig. 4A) in the context of the R12 protein. We choose to mutate two of the most conserved residues at positions 3 and 5 of the RNP-1 motif (Phe-Tyr in RBD 1 and Ile-Tyr in RBD 2). Mutation of Phe-Tyr to Phe-Phe in RBD 1 does not change significantly the binding affinity and specificity of the chimeric protein (Kd ~ 50 nM for the NRE versus 1 μM for NS; Fig. 4, B and C). In contrast, a mutation to Phe-Leu, Leu-Leu, or Asp-Asp drastically reduced the interaction with the NRE (Kd > 2 μM for 1FL and >10 μM for 1LL and 1DD, where single-letter amino acid code is used), suggesting that the presence of an aromatic residue at position 5 within the RNP-1 motif of RBD 1 was absolutely required for an interaction with the NRE RNA. This was further confirmed with mutants 1LF and 1LY, which partially restored the binding affinity (Kd ~ 200 nM). The same qualitative results that for RBD 1. Mutation of Ile-Tyr to Phe-Leu, Leu-Leu, or Asp-Asp drastically reduced the interaction with the NRE RNA. The same analysis performed on the RNP-1 motif of RBD 2 provided the same qualitative results that for RBD 1. Mutation of Ile-Tyr to Phe-Leu, Leu-Leu, or Asp-Asp reduced moderately the binding affinity for the NRE RNA (Kd ~ 250 nM). The 2LL mutant interacts weakly with the specific RNA target (Kd ~ 500 nM), and the interaction was drastically reduced with the mutations 2LL and 2DD (Kd > 1 μM). The binding affinity obtained with the 2LL mutant is also in agreement with the binding affinity of the R12ΔC protein (Fig. 3), where the RNP-1 motif of the RBD 2 has been deleted. It is interesting to note that the mutation Ile-Tyr to Leu-Leu within RBD 2 seems to be less deleterious than the analogous mutation in RBD 1 (Phe-Tyr to Leu-Leu) (Kd > 10 μM for 1LL and around 1 μM for 2LL; see Fig. 4B and also Fig. 5B), and might indicate that RBD 1 has a preponderant role for NRE recognition. This mutational analysis of the RNP-1 motif of RBDs 1 and 2 within the R12 protein clearly indicates the importance of the aromatic residue at position 5 within the RNP-1 motif of both RBDs and suggests a direct or indirect involvement of both RNP-1 consensus sequences in the recognition of the NRE sequence.

To further confirm that RBDs 1 and 2 were necessary and sufficient to confer RNA binding affinity and specificity to nucleolin, we introduced the RNP-1 mutations within the R1234G protein (which corresponds to the C-terminal end of nucleolin; see Fig. 1) (Fig. 5A). As shown above (Fig. 2), the R1234G protein interacts with the NRE sequence with the same affinity (Kd ~ 20–50 nM) and specificity as native nucleolin. Introduction of the conservative mutation Phe-Tyr to Leu-Leu in RBD 1 and Ile-Tyr to Leu-Leu in RBD 2, respectively, severely reduced the interaction with the NRE sequence (Kd ~ 500 nM for Mut RBD 1 and 200 nM for Mut RBD 2). The binding affinity of Mut RBD 1 is in fact close to the binding affinity of the R12ΔC protein for the NS RNA (Fig. 2), and results mainly from the nonspecific interaction of the RGG domain with RNA (Fig. 2) (53). Interestingly, the mutations Phe-Tyr to Leu-Leu in RBD 1 and Ile-Tyr to Leu-Leu in RBD 2 confer the same binding properties to the resulting chimeric proteins when they are introduced either in the R12 or R1234G proteins, i.e: the 2LL and Mut RBD 2 proteins interact with a
slight but reproducibly higher affinity with the NRE sequence than the 1LL and Mut RBD 1 proteins. Again, this small difference in binding affinity suggests a preponderant role of RBD 1 over RBD 2 for selective NRE recognition. In contrast, the same mutation within RBD 3 (Tyr-Phe to Leu-Leu) or RBD 4 (Phe-Phe to Leu-Leu) produced no detectable effect on the binding affinity of the resulting chimeric proteins for the NRE sequence compared with wild-type protein (Kd of 20–50 nM). Therefore, these results confirm the deletion (Figs. 2 and 3) and mutational analysis of the RNP-1 motif of RBDs 1 and 2 in the context of the R12 protein (Fig. 4) and show that RBDs 1 and 2 are both necessary and sufficient for the interaction with the NRE sequence.

**Circular Dichroism Analysis**—We used circular dichroism as a means to assess the folding of the R12 polypeptide and to evaluate the effect of the various amino acid substitutions on this folding. The spectrum of R12 (Fig. 6, A and B) is very similar to a simulated spectrum calculated on the basis of the CD signal (data not shown). Furthermore, the R12 spectrum is analogous to the reported spectra of other RBDs (57, 59). This suggests an autonomous folding of each domain within R12 into an overall structure common to most RBDs (Fig. 7). The spectra of the RBD 1 RNP-1 double mutants (Fig. 6A) differ only slightly from the spectrum of R12, which means that the substitutions of the two aromatic residues have not grossly affected the R12 folding. Subtle alterations of the wild-type spectrum are visible (Fig. 6A; see arrows) especially in the case of the 1DD and 1LL spectra, and correspond to a few percent loss or gain in α-helix content, respectively. Modeling of R12 and its variants (under current investigation) indeed predicts a possible destabilization (for 1DD) or stabilization (for 1LL) of the nearby helices due to the corresponding side-chain substitutions in RNP-1. The CD spectra of the RBD 2 RNP-1 double mutants also reveal some similarity in secondary structure elements to the R12 spectrum (Fig. 6B), which again suggests that these mutants also retain a correct folding overall. In this case, however, alterations in the spectra of 2DD and 2LL (Fig. 6B, arrows), albeit of the same type as for 1DD and 1LL, are more pronounced. Moreover, in addition to these fluctuations in α-helix content, there might be a qualitative modification of the β component in the case of 2DD, as indicated by the significant shifts in the positions of the two CD minima.

**CD** is a useful tool to analyze conformational changes induced by formation of a protein/nucleic acid complex, since the composite spectrum is dominated by the nucleic acid component in the case of 2DD, as indicated by the significant increase in the value of the RNA's CD maximum, which is consistent with a certain degree of RNA condensation (61) mediated by a small proportion of R12 oligomerization. It should be noted that such a spectroscopic change probably masks the weak decrease in CD induced by the unstacking of a few bases upon R12 binding.

**DISCUSSION**

We have analyzed the interaction between mouse nucleolin and an RNA target. We previously identified an RNA (NRE), which binds with high affinity to nucleolin (21, 54), and our present purpose is to determine the origin of the RNA-binding specificity of nucleolin. Binding studies with the four isolated nucleolin CS-RBDs (each individual domain is defined here from the beginning of the β1 strand to the C terminus of the β4 strand) indicate that none of them was able to interact significantly with the nucleolin RNA target (Kd > 10 μM; Fig. 2). The low binding affinity of these isolated CS-RBDs seems to be common to most CS-RBDs (63, 64). To identify the minimal region of nucleolin involved in specific RNA recognition, we performed N- and C-terminal deletions on the R1234G protein (Fig. 2), which retains the RNA-binding specificity of full-length protein. A region of the protein containing two CS-RBDs
(RBDs 1 and 2) is required for the specific interaction with the NRE RNA sequence (Fig. 2). A recombinant protein, which contains these two CS-RBDs separated by a 12-amino acid linker, interacts with the same affinity and specificity as full-length protein. Deletions at either end of this bi-RBD peptide completely abolish the high affinity interaction (Fig. 3). Our results strongly suggest that these two domains cooperate in the recognition of the RNA sequence. A cooperation between two CS-RBDs has also been suggested for other multiple RBD-containing proteins, like hnRNPA1 (11, 57), ASF/SF2 (14, 42), Sxl (16), and PABP (43, 64). Comparison of the poly(A)-binding protein with nucleolin is interesting since both proteins have four CS-RBDs. As in the case of nucleolin, it has been proposed that the RNA-binding specificity of PABP arises from the first two RBDs (43, 64). RBD 1 of PABP has no RNA binding activity by itself, but improves the discriminatory ability in combination with the other domains (44). However, it should be pointed out that the RNA-binding specificity of PABP seems to be less restricted to these two RBDs than in nucleolin since RBDs 3 and 4 of PABP interact also with the specific RNA target (poly(A) sequence) with only a 10-fold decreased binding affinity compared with RBDs 1 and 2 (44). In contrast, nucleolin RBDs 3 and 4 (R34, Fig. 2A) do not interact significantly with the NRE sequence at the protein concentration used in this study. Compared with RBDs 1 and 2, the binding affinity of RBDs 3 and 4 for the NRE sequence is decreased at least 1000-fold.

Effect of Mutations within the RNP-1 Motifs on RNA Binding—The three-dimensional structures of the hnRNP C, Sxl, and of the U1-SnRNP A N-terminal CS-RBDs suggest that solvent-exposed aromatic residues in RNP-1 and RNP-2 sequences contribute to RNA binding through ring-stacking interactions with the bases (16, 25, 27). Mutation of residues at positions 3 and 5 of the RNP-1 motif has been shown to dras-
tically reduce the binding affinity of the corresponding CS-RBD (32–34, 42, 58). To determine which nucleolin CS-RBD was involved in the specific recognition of NRE RNA, we introduced several mutations at these positions within each of the four CS-RBDs (Fig. 5). Binding studies with these chimeric proteins show that mutation of the RNP-1 motif within the CS-RBDs 3 and 4 does not affect the binding affinity and specificity, which demonstrates that these domains are not involved in recognition of this specific RNA. It is not clear now if these two domains are involved in other RNA-specific interactions, since a SELEX experiment with the full-length nucleolin identified only the consensus sequence (NRE) (21) used in this study. In contrast, substitution of the analogous residues within CS-RBDs 1 and 2 drastically reduced the binding affinity of the mutated protein. When the same substitutions were introduced within the R12 peptide (Fig. 4), we also observed that mutation of the RNP-1 sequences of RBD 1 or RBD 2 abolished the specific high affinity interaction with the NRE RNA sequence. This result clearly demonstrates the involvement of both CS-RBDs in the specific recognition of NRE RNA.

We also observed that mutations in CS-RBD 1 (Figs. 4 and 5) had a more drastic effect than the analogous mutations in CS-RBD 2, suggesting a preponderant role for RBD 1 in RNA recognition. Binding studies with N- and C-terminal deletion mutants of the R12 protein (Fig. 3) indicate that RBD 1 displays a higher basal binding affinity ($K_d$ of 1 $\mu$M) than RBD 2 ($K_d$ > 10 $\mu$M). These isolated RBDs are not able to discriminate specific versus nonspecific RNA sequences. We therefore suggest that RBD 1 binds first to the RNA, then this interaction is stabilized further by the second RBD. This stabilization most likely involves specific contacts between both RBDs and the RNA sequence. This mechanism of interaction of a bi-RBDs peptide with a single RNA target could also explain how the hnRNPA1 and Sxl proteins interact with their targets. We also tested whether the binding specificity and affinity of the R12 peptide for the NRE sequence could be obtained when both individual RBDs are mixed together (data not shown). Under these conditions, no interaction could be detected even at high peptide concentration (>5 $\mu$M), showing that these two domains must be present on the same peptide to interact with the RNA target. The requirement for a $\text{cis}$ arrangement of the 2 CS-RBDs of Sxl has also been suggested for their site-specific interaction (16).

The affinity of the individual RBDs for the RNA target is significantly lower, and they lose the ability to discriminate between the specific and nonspecific RNA sequence. However, the nonspecific binding activity of the first RBD is about 100-fold higher than for RBD 2, suggesting also that RBD 1 will be the first one to interact with the RNA target. An analogous situation is found in hnRNP A1 where two CS-RBDs are required for a strong specific interaction (11, 32, 57). Cross-linking studies (34) and mutational analysis of the RNP-1 motif of each domain (32) also suggested a preponderant role for the first RBD. An interesting characteristic of the binding properties of a 2-RBD peptide versus the corresponding isolated domains is that the binding affinity of the bi-RBD protein is within the R12 peptide (Fig. 4), we also observed that mutation of the RNP-1 sequences of RBD 1 or RBD 2 abolished the specific high affinity interaction with the NRE RNA sequence. This result clearly demonstrates the involvement of both CS-RBDs in the specific recognition of NRE RNA.

FIG. 6. Circular dichroism analysis of the R12 protein and several variants. A, CD spectra of the R12 polypeptide (WT) and its variants mutated within the RNP-1 of the first RBD. B, CD spectra of the R12 polypeptide (WT) and its variants mutated within the RNP-1 of the second RBD. C, CD spectra of NRE RNA either free or in complex with the R12 peptide (1:1 protein/RNA stoichiometry). The key to the different curves is shown on each panel.

FIG. 7. Homology modeling of nucleolin RBDs 1 and 2. Models were generated using the Biosym Technologies modules InsightII, Discover, and Homology (version 230), run on a Silicon Graphics Indigo workstation. Computer modeling of the two separate domains was performed by homology with the first domain of U1A (Protein Data Bank code 1URN). The backbones of the two respective RNP-1 motifs have been labeled in black. The conserved aromatic and hydrophobic residues, which are the main sites of mutation, have been represented in full. One of the potential structures of the flexible linker (under current investigation) appears both at the RBD-1 C terminus and RBD-2 N terminus.

We also observed that mutations in CS-RBD 1 (Figs. 4 and 5) had a more drastic effect than the analogous mutations in CS-RBD 2, suggesting a preponderant role for RBD 1 in RNA recognition. Binding studies with N- and C-terminal deletion mutants of the R12 protein (Fig. 3) indicate that RBD 1 displays a higher basal binding affinity ($K_d$ of 1 $\mu$M) than RBD 2 ($K_d$ > 10 $\mu$M). These isolated RBDs are not able to discriminate specific versus nonspecific RNA sequences. We therefore suggest that RBD 1 binds first to the RNA, then this interaction is stabilized further by the second RBD. This stabilization most likely involves specific contacts between both RBDs and the RNA sequence. This mechanism of interaction of a bi-RBDs peptide with a single RNA target could also explain how the hnRNPA1 and Sxl proteins interact with their targets. We also tested whether the binding specificity and affinity of the R12 peptide for the NRE sequence could be obtained when both individual RBDs are mixed together (data not shown). Under these conditions, no interaction could be detected even at high peptide concentration (>5 $\mu$M), showing that these two domains must be present on the same peptide to interact with the RNA target. The requirement for a $\text{cis}$ arrangement of the 2 CS-RBDs of Sxl has also been suggested for their site-specific interaction (16).

The affinity of the individual RBDs for the RNA target is significantly lower, and they lose the ability to discriminate between the specific and nonspecific RNA sequence. However, the nonspecific binding activity of the first RBD is about 100-fold higher than for RBD 2, suggesting also that RBD 1 will be the first one to interact with the RNA target. An analogous situation is found in hnRNP A1 where two CS-RBDs are required for a strong specific interaction (11, 32, 57). Cross-linking studies (34) and mutational analysis of the RNP-1 motif of each domain (32) also suggested a preponderant role for the first RBD. An interesting characteristic of the binding properties of a 2-RBD peptide versus the corresponding isolated domains is that the binding affinity of the bi-RBD protein is...
significantly less than the product of the binding affinity of the two isolated domains (32, 57, this study). For hnRNPA1, this difference could be explained by theoretical models in which the two independent binding domains are connected by a flexible linker (57, 65). In this model, when the RBD with the highest binding affinity is bound to the RNA target, the local concentration of the second domain is lower than that required for its efficient interaction with the RNA sequence. These theoretical considerations (57, 62) could also explain why the binding affinity of the R12 peptide for the NRE sequence (Kd of 20–50 nM) is significantly less than the product of the affinities of RBD 1 and RBD 2 (Kd > 10 μM for each domain). Our observation that the RNP-1 motif of each RBD is involved in RNA interaction, together with the fact that binding of RBDs 1 and 2 is nonadditive, suggests that the simultaneous binding of both RBDs to the RNA target is required for a high affinity and specific interaction.

CD analysis has allowed us to address two main concerns: the independent folding of the two RBDs (Fig. 7) and the lack of major restructuration after mutation in their respective RNP-1 sequences. The fact that the overall structure of the R12 peptide was maintained permits an unbiased diagnosis of the role of the mutated aromatic residues in RNA binding. That both sets of aromatic residues appear essential to RNA binding should be more fully investigated. As mentioned above, it is likely that the RNP-1 aromatic residues from RBD 1 play a preponderant role in the first step of RNA binding. What would then be the role of the RNP-1 aromatic residues from the second RBD? Preliminary modeling studies of the R12 peptide indicate a spatial proximity of both sets of aromatic residues. Although undetected in our conditions of CD experiments, which are mostly sensitive to the secondary structure of the polypeptide, the hydrophobic core formed by these aromatic residues, could play a major role in holding in close vicinity determinants of specificity coming from both RBDs.

Acknowledgments—We thank Y. de Preval for synthesis of oligonucleotides and L. Poljak for reading the manuscript.

REFERENCES

1. McCarthy, J., and Kollmus, H. (1995) Trends Biochem. Sci. 20, 191–197
2. Burd, C. G., and Dreyfuss, G. (1994) Science 265, 615–621
3. Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993) Annu. Rev. Biochem. 62, 289–321
4. Amrein, H., Gorman, M., and Noethiger, R. (1988) Cell 55, 1025–1035
5. Zamore, P. D., Patton, J. G., and Green, M. R. (1992) J. Biol. Chem. 267, 3307–3313
6. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1994) J. Mol. Biol. Rep. 9, 39–47
7. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K., and Dreyfuss, G. (1986) Mol. Cell. Biol. 6, 2932–2943
8. Sachs, A. B., Bond, M. W., and Kernberg, R. D. (1996) Cell 87, 827–835
9. Narada, T., Strong, W. B., Menaya, J., Faberi, B. J., and Hershey, J. W. B. (1994) J. Biol. Chem. 269, 14465–14472
10. Zhang, W., Wagner, B. J., Ehrenman, K., Schaefer, A. W., Demaria, C. T., Crater, D., Dehaven, K., Long, L., and Brewer, G. (1993) Mol. Cell. Biol. 13, 7652–7665
11. Burd, C. G., and Dreyfuss, G. (1994) EMBO J. 13, 1197–1204
12. Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994) J. Biol. Chem. 269, 23074–23079
13. Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994) Exp. Cell Res. 211, 400–407
14. Tacke, R., and Manley, J. L. (1995) EMBO J. 14, 3540–3551
15. Sakashita, E., and Sakamoto, H. (1994) Nucleic Acids Res. 22, 4082–4086
16. Kanaar, R., Lee, A., Rudner, D. Z., Wemmer, D. E., and Ro, D. C. (1995) EMBO J. 14, 4530–4539
17. Scherly, D., Boelens, W., van Veenrooj, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. (1989) EMBO J. 8, 4163–4170
18. van Gelder, C. W. G., Gundersun, S. I., Jensen, E. R. J., Boelens, W. C., Polycarpou-Scharz, M., Mattaj, I. W., and van Veenrooj, W. J. (1993) EMBO J. 12, 5191–5200
19. Scherly, D., Boelens, W., Dathan, N. A., van Veenrooj, W. J., and Mattaj, I. W. (1990) Nature 345, 502–506
20. Query, C. C., Bentley, R. C., and Keene, J. D. (1989) Mol. Cell. Biol. 9, 4872–4881
21. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
22. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
23. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
24. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
25. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
26. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93
27. Ghisolfi, L., Joseph, G., Puvion-Dutilleul, F., Amalric, F., and Bouvet P. (1996) J. Mol. Biol. 260, 84–93