THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 38, Issue of September 22, pp. 22167–22175, 1995

Printed in U.S.A.

The UUAG-specific RNA Binding Protein, Heterogeneous Nuclear Ribonucleoprotein D0

COMMON MODULAR STRUCTURE AND BINDING PROPERTIES OF THE 2xRBD-Gly FAMILY*

(Received for publication, April 13, 1995, and in revised form, July 12, 1995)

Yasuko Kajita‡§, Jun-ichi Nakayama‡, Masuo Aizawa§, and Fuyuki Ishikawa‡¶

From the ‡Department of Life Science and §Department of Bioengineering, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226, Japan

Human cDNA clones encoding the UUAG-binding heterogeneous nuclear ribonucleoprotein (hnRNP) D0 protein have been isolated and expressed. The protein has two RNA-binding domains (RBDs) in the middle part of the protein and an RGG box, a region rich in glycine and arginine residues, in the C-terminal part ("2xRBD-Gly" structure). The hnRNP A1, A2/B1, and D0 proteins, all possess common features of the 2xRBD-Gly structure and binding specificity toward RNA. Together, they form a subfamily of RBD class RNA binding proteins (the 2xRBD-Gly family). One of the structural characteristics shared by these proteins is the presence of several isoforms presumably resulting from alternative splicing. Filter binding assays, using the recombinant hnRNP D0 proteins that have one of the two RBDs, indicated that one RBD specifically binds to the UUAG sequence. However, two isoforms with or without a 19-amino acid insertion at the N-terminal RBD showed different preference toward mutant RNA substrates. The 19-amino acid insertion is located in the N-terminal end of the first RBD. This result establishes the participation of the N terminus of RBD in determining the sequence specificity of binding. A similar insertion was also reported with the hnRNP A2/B1 proteins. Thus, it might be possible that this type of insertion with the 2xRBD-Gly type RNA binding proteins plays a role in "fine tuning" the specificity of RNA binding. RBD is supposed to bind with RNA in general and sequence-specific manners. These two discernible binding modes are proposed to be performed by different regions of the RBD. A structural model of these two binding sites is presented.

Ribonucleoproteins have been found in many macromolecular complexes that have vital biological roles, such as heterogeneous nuclear ribonucleoprotein (hnRNP)[1] (1), small nuclear ribonucleoprotein (snRNP), ribosomes, and signal recognition particles. These complexes are composed of RNAs and proteins, many of which show RNA binding activities. One of the most common groups of RNA binding proteins is the RBD class proteins (2). They possess a CS-RBD (consensus sequence-RNA binding domain) motif, which is typically 80–90 amino acids. Two short sequences, RNP 2 octamer and RNP 1 hexamer, have been found to be conserved among different RBDs. Several RBDs are commonly found in tandem within one molecule. It is also common to find an auxiliary RNA-binding motif present in addition to RBDs within the same molecule. Thus, RBD class RNA binding proteins typically possess several RNA-binding domains as modules. It has not been well studied, however, how these modular domains participate together in binding with RNA.

hnRNP proteins are a subset of proteinaceous components found in hnRNP, which is a large complex formed by the nascent pre-mRNA and proteins (1, 3). More than 20 proteins have been identified as hnRNP proteins on two-dimensional protein gel electrophoresis. Although structures of all of these proteins are not known, many contain RBDs, which are the regions responsible for interaction with RNA. Some hnRNP proteins have been implicated in the processing of pre-mRNA. Anti-hnRNP C protein antibody inhibited pre-mRNA splicing *in vitro* (4, 5). Several hnRNP proteins were reported to be associated in spliceosomal complexes (6, 7). Finally, the hnRNP A1 and A2/B1 have been shown to influence the splice site selection (8–10). These observations suggest that hnRNP proteins may have a role in specific RNA processing reactions by virtue of sequence-specific RNA binding in addition to nonspecific general RNA binding. In spite of this expectation, only a small number of hnRNP proteins have been shown to bind to RNA in a sequence-specific manner.

In a previous study, we showed that several different proteins from the HeLa cell nuclear extract specifically bind to single-stranded $d(TTAGGG)_4$ and $r(UUAGGG)_4$ oligonucleotides (11). These proteins have apparent molecular masses of 26, 28, 37, 39, 41, 50, and 55 kDa. Amino acid sequencing of the purified proteins indicated that the 26-, 28-, and 50-kDa proteins are the hnRNP A1 protein, A2/B1 protein, and nucleolin, respectively. The 39- and 41-kDa proteins were immunoreactive to anti-hnRNP D monoclonal antibodies. On two-dimensional gel electrophoresis, they migrated as spots near, but separate from, the hnRNP D protein. We suggested that the 39- and 41-kDa proteins are identical or closely related to the hnRNP D protein. Similarly, the 37-kDa protein was suggested to be identical or closely related to the hnRNP E protein and was referred to as hnRNP E0. In this study, we will refer to the 39- and 41-kDa hnRNP D-like proteins having UUAGGG-binding activity as hnRNP D0 proteins.

The hnRNP A1, A2/B1, D0, and E0 proteins bound to

* This work was supported by a Grant-in-aid for Specially Promoted Research from the Ministry of Education, Science and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) D55671 (human mRNA for hnRNP D protein, cDx4), D55672 (human mRNA for hnRNP D protein, cDx7), D556713 (human mRNA for hnRNP D protein, cDx8), and D55674 (human mRNA for hnRNP D protein, cDx9).

¶ To whom correspondence should be addressed. Fax: 81-45-924-5771; Tel.: 81-45-924-5703; E-mail: fishikaw@bio.titech.ac.jp.

[1] The abbreviations used are: hnRNP, heterogeneous nuclear ribonucleoprotein; snRNP, small nuclear ribonucleoprotein; PCR, polymerase chain reaction; bp, base pair(s); RBD, RNA binding domain; MES, 2-(*N*-morpholino)ethanesulfonic acid.

UUAGGG repeats but not to single base-substituted oligoribonucleotides, such as CUAGGG-, UCAGGG-, UUGGGG-, or UUAAGG repeats. Thus, their binding to these substrates is exceptionally sequence-specific compared with other hnRNP proteins. This feature offers an opportunity to study the molecular interaction between RBD and RNA. In this study, we first examined the cDNA structure of the hnRNP D0 proteins. Results revealed that the hnRNP D0 protein has a modular structure in common with the hnRNP A1 and A2/B1 proteins. We next examined the RNA binding properties of each modular domain of the hnRNP D0 proteins. A model for molecular interaction between the protein and RNA is proposed based upon these structural and functional analyses.

MATERIALS AND METHODS

Oligonucleotides Used in the PCR Reactions and Binding Assays—Oligonucleotides were synthesized either by an Applied Biosystems 380B synthesizer or by a Perceptive Expedite System 8900. All oligonucleotides were purified by denaturing acrylamide gel electrophoresis and Sep-pak C18 cartridges (Waters). The sequences are as follows: S1, 5′-d(CTGAATTCCATGGGAACGACACTCTGAAGCA)-3′; S2, 5′-d(CAGTCGACGAATTCAACCGGCTCTTTTGTT)-3′; S3, 5′-d(CTGAATTCCATGGCCATGAAAACAAAAGAGCC)-3′; S4, 5′-d(CAGTCGACGAATTCTTGCTGTTGCTGATATT)-3′; P1, 5′-d(CGGATCCAAATGTCGGAGGAGCAGTT)-3′; P3, 5′-d(AGGATCCAAGCCAGTAAGAACGAGGA)-3′; P4, 5′-d(CGAATTCTCAGGCTTTGGCCCTTTTAG)-3′; P5, 5′-d(CGGATCCAAGCCATGAAAACAAAAGA)-3′; P6, 5′-d(CGAATTCTCACGACATGGCTACTTTTA)-3′; rH4, 5′-r(UUAGGG)$_4$-3′; rH4X1, 5′-r(UUGGGG)$_4$-3′ and rECGF, 5′-r(GCAGCCUUGAUGACCUCGUGAACC)-3′.

Isolation of cDNA Clones—Two DNA fragments of human E2BP cDNA were prepared by PCR using two primer sets of S1, S2 and S3, S4. A 292-bp fragment generated by S1 and S2 (corresponding to positions 210–501 of GenBank™ M94630) and a 281-bp fragment generated by S3 and S4 (corresponding to positions 477–759 of GenBank™ M94630) were obtained. They were ^{32}P-labeled and were used to screen the cDNA library.

A HeLa cDNA library was constructed using 4 μg of HeLa poly(A)$^+$ RNA using λEXlox vector (Novagen). A total of 2×10^6 plaques were screened by the two E2BP cDNA-specific probes. Nine clones, cDx1–9, were identified as being positive by both probes. The clones were sequenced by an Autocycle Sequencing kit and an ALF. DNA Sequencer (Pharmacia Biotech Inc.).

Generation of cDNA Fragments Encoding Truncated hnRNP D0 Proteins—To obtain cDNA fragments encoding truncated mutant hnRNP D0 proteins, parts of cDNA were generated by reverse transcription-PCR amplification from HeLa poly(A)$^+$ RNA using different sets of PCR primers. Fragments encoding RBD-1 were generated by P3 and P4 primers. Two DNA fragments of different sizes were obtained. One RBD-1 with and one without the 19-amino acid insertion (see "Results"). A fragment encoding RBD-2 was generated by P5 and P6 primers. Fragments encoding RBD-1 and -2 were generated by P3 and P6 primers. Again, two DNA fragments of different sizes were obtained, corresponding to RBD-1 and -2 with and without the 19-amino acid insertion in RBD-1. To construct cDNAs encoding +/− and −/+ type isoform proteins (for explanations see "Results"), first, 5′-parts of cDNA coding for the N-terminal portion and RBD-1 with and without the 19-amino acid insertion, were prepared by reverse transcription-PCR using primers P1 and P4. P1 starts at the initiating codon of cDNA. 3′-parts of cDNA were derived from cloned cDNA, cDx4, and 7. As both PCR-derived 5′-parts and cDNA-derived 3′-parts have a common unique *Bgl*II site, they were digested at this restriction site and were ligated to give rise to all of the entire coding regions. All of these truncated cDNAs were subcloned into pGEX-5X-1 (Pharmacia).

In Vitro Expression of Recombinant Proteins—All hnRNP D0 protein fragments were expressed as fused proteins with glutathione *S*-transferase for easier purification. Each of the pGEX-5X-1 plasmids containing truncated cDNA fragments was transformed into *Escherichia coli* strain JM105 or BL21 (DE3) pLysS. Bacterial cultures grown in L-rich medium (2.5% tryptone, 0.75% yeast extract, 0.5% NaCl, 0.2% glucose) were treated by 0.2 mM isopropyl-1-thio-β-D-galactopyranoside for 4 h to induce protein expression when their A_{600} reached 0.4–0.6. Cells were pelleted, frozen at −70 °C, thawed in NTEN150 (20 mM Tris-HCl pH 8.0, 1 mM, 150 mM NaCl, 0.5% (v/v) Nonidet P-40), and sonicated. Extracts were centrifuged twice at $10,000 \times g$ for 15 min. All of the glutathione *S*-transferase fusion proteins were soluble. The supernatant was treated with glutathione-Sepharose 4B (Pharmacia) on ice for 30 min. Then the glutathione-Sepharose 4B was washed with NTEN100 (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 100 mM NaCl, 0.5% (v/v) Nonidet P-40) by centrifuging at $5000 \times g$, for 20 s at 4 °C three times. Proteins were eluted from glutathione-Sepharose 4B by an elution buffer (50 mM Tris-HCl, pH 8.0, 20 mM glutathione). Some fusion proteins were digested by Factor Xa to remove glutathione *S*-transferase before further purification. To cleave at the Factor Xa recognition site that is present between the glutathione *S*-transferase and the hnRNP D0 sequences, $CaCl_2$ was added to protein samples to a final concentration of 5 mM. Factor Xa was added to the solution, which was incubated at room temperature overnight. Then, to remove bacterial RNA in samples, 50 units of micrococcal nuclease was added. Samples were incubated at 37 °C for 10 min. The reaction was terminated by adding EGTA to 50 mM. To purify the proteins by ion exchange chromatography, an equal volume of 0.5 M MES, pH 5.0 was added to samples, which were loaded on a HiTrap SP column (Pharmacia). Recombinant proteins were eluted by 0.1 M NaCl, 20 mM Tris-HCl pH 7.5. Eluted samples were concentrated and suspended in 0.1 M NaCl, 10% glycerol, and 20 mM Tris-HCl, pH 7.5, by Centricon 10 (Amicon). Concentrated samples were loaded on a Sephadex 75 HR gel filtration chromatography (Pharmacia) equilibrated by 0.5 M NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10% glycerol, and 20 mM Tris-HCl, pH 7.4. Purified proteins were concentrated and suspended in a binding buffer, BB (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 50 mM NaCl 10% (v/v) glycerol) by Centricon 10. Protein concentrations were measured by a protein assay kit (Bio-Rad) with bovine serum albumin as a standard.

Filter Binding Assay—Recombinant hnRNP D0 proteins were diluted with BB immediately before use. They were incubated with 1–0.1 nM of ^{32}P-labeled RNA probes in 100 μl of BB. After incubation at room temperature for 20 min, reactions were filtrated through a nitrocellulose membrane (Schleicher & Schuell), and membranes were dried at 100 °C. The radioactivities were measured by liquid scintillation counting. About 5% of the input radioactivity was measured as background in the absence of any protein in the reaction mixture. This background count was subtracted from the measured counts to give rise to specific binding counts.

RESULTS

Primary Structure of the hnRNP D0 Proteins—Previously, we described amino acid sequences of five peptides obtained from the purified human hnRNP D0 proteins (11). They were identical, or nearly identical, to sequences that had been reported under several different protein names. These included the human hnRNP C protein, the rat hnRNP C-type protein, and the E2BP hepatitis B enhancer binding protein (12–14). It was highly possible that these proteins were derived from the same gene as the hnRNP D0 proteins. Although the reported cDNA sequences were closely related to each other, several base insertions, deletions, and substitutions that changed the open reading frames were noted. Therefore, we first isolated the cDNA clones and examined the primary structure.

The predicted amino acid sequences deduced from the E2BP cDNA revealed that this protein has two RBDs (14). Two sets of PCR primers, S1:S2 and S3:S4 were prepared according to the reported sequences of each RBD. Accordingly, two DNA fragments derived from E2BP cDNA were obtained by reverse transcription-PCR from the two primer sets using the total RNA of HeLa cells. A total of 200,000 clones of the HeLa cell cDNA library were prepared by oligo(dT)-priming and were screened by these E2BP-specific probes. Nine different clones were determined to be double-positive by the probes. The longest clone, cDx7, was sequenced completely, identifying a 1589-bp cDNA insert (Fig. 1). A long open reading frame, bound by TAG at 226–228 and TAA at 1204–1206, was identified. A polyadenylation signal sequence, AATAAA, was noted at 1541–1546. ATG at 286–288 was tentatively assigned as an initiating codon. It was predicted that the open reading frame encodes a 306-amino acid protein with a calculated molecular weight of 32,800. All five amino acid sequences identified in peptides, obtained from the purified hnRNP D0 proteins, were found in the predicted amino acid sequence, except that one amino acid substitution was noted (Fig. 1). As will be described

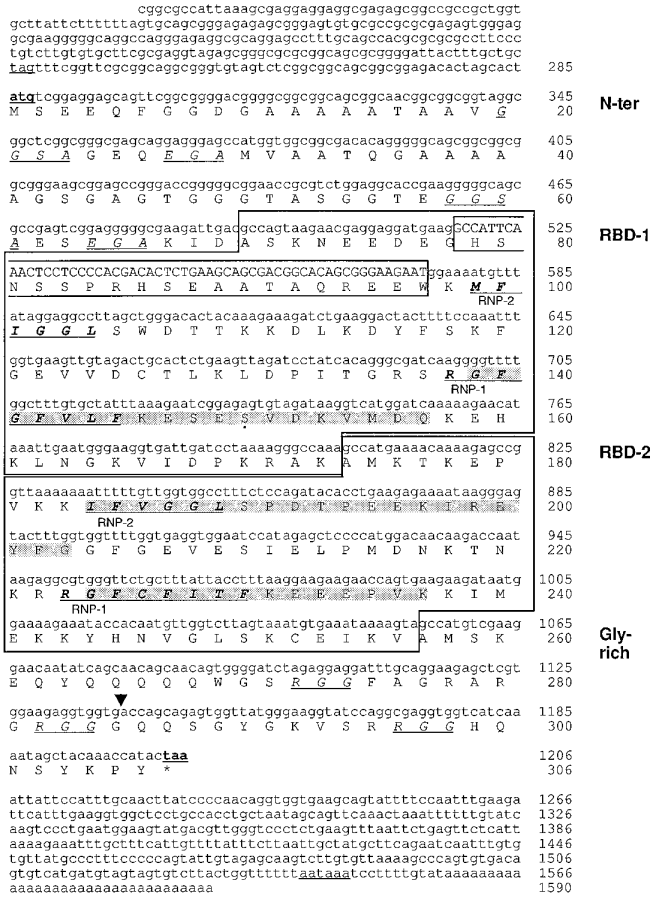

FIG. 1. **Nucleotide sequence of the hnRNP D0 cDNA, cDx7, and predicted amino acid sequence.** The nucleotide sequence is numbered from the 5′-end of cDx7 cDNA. A putative initiation codon ATG at 286–288 and the stop codon of the open reading frame TAA at 1204–1206 are shown by *boldface letters* and are *underlined*. An in-frame, upstream stop codon TAG at 226–228 and polyadenylation signal sequence aataaa at 1541–1546, are *underlined*. The amino acid sequences of tryptic peptides that were obtained from purified hnRNP D0 proteins (11) are indicated by *shading*. Serine at position 150 was identified as arginine in the previous study (11) (*dotted*). The two CS-RBDs, RBD-1 and RBD-2, are *boxed*. A 57-bp insertion that is missing in some isoforms is shown by a *box inside* the RBD-1. The RNP 1 and RNP 2 sequences are indicated by *boldface italics* and are *underlined*. The 5′-region encodes an amino acid sequence that is rich in alanine and glycine (indicated by *N-ter*). Two short motifs of GGSA and EGA, found repeatedly in tandem (amino acids 20–29 and 58–66), are indicated by *italics* and are *underlined.* The 3′-region encodes an amino acid sequence rich in glycine (indicated by *Gly-rich*). The three RGG motifs in this region are shown by *italics* and are *underlined*. The position at which the 147-bp insertion is found in other isoforms is indicated by an *arrowhead* (positions 1138–1139). The differences between the reported E2BP cDNA sequence (14) and this sequence are as follows: E2BP cDNA has a "t" insertion at positions 1202 and 1203, "gg" deletion at positions 1277 and 1278, and a 139-bp deletion at positions 1419–1557.

later, the recombinant protein of this cDNA is immunoreactive to an anti-hnRNP D monoclonal antibody and binds to the $d(TTAGGG)_4$ and $r(UUAGGG)_4$ oligonucleotides specifically. Therefore, we concluded that the cDNA clones we have isolated are for the hnRNP D0 proteins.

The nucleotide sequence of cDx7 is different from that of E2BP in several ways. cDx7 has a longer 5′ upstream sequence than E2BP, allowing us to locate the most probable initiating codon. Several nucleotides were missing or replaced by other nucleotides in E2BP, resulting in changes to the open reading frame and the predicted amino acids. The detail of discordance is presented in Fig. 1. These discordant sequences were repeatedly examined with cDx7 and with our other cDNA clones,

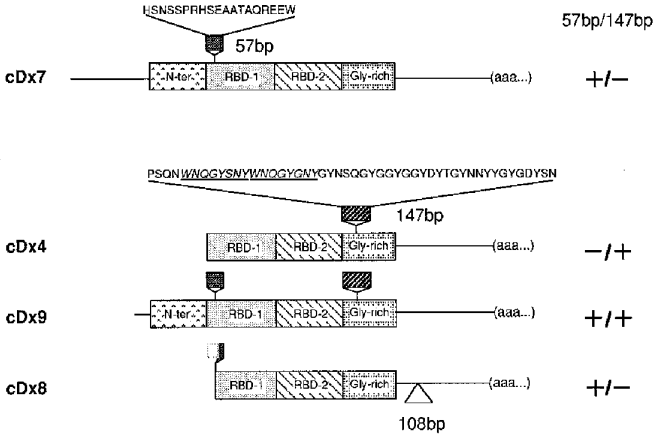

FIG. 2. **Structures of the hnRNP D0 cDNA isoforms.** Structures and the predicted coding regions of the four hnRNP D0 cDNAs; cDx7, -4, -9, and -8 are schematically shown. Deduced amino acid sequences comprise four domains: N-ter, RBD-1, RBD-2, and Gly-rich. A 57-bp insertion in RBD-1 that is present in cDx7, -9, and -8, is illustrated by *heavily shaded boxes*. A 147-bp insertion in the C-terminal Gly-rich region found in cDx4 and -9 is indicated by *hatched boxes*. Deduced amino acid sequences of these two insertions are also shown. A motif of eight amino acids was repeated in tandem twice (*italic* with *underline*). A 107-bp insertion found in the untranslated region of cDx8 is also shown. The nucleotide sequence of 107-bp insertion in the 3′-untranslated region is as follows: 5′-cgggaacttcattgcaggccctgtgtcgcgctgacttcagattctcacaggcccgctcaatgcggacagggtaacgagatgctccacgctctcgaatgctgcc-gtttg-3′.

giving the same results.

The predicted amino acid sequence of cDx7 can be divided into three parts. The N-terminal 69 amino acids forms an acidic region that is unique to this protein. Alanine and glycine are abundant in this region (27 and 29%, respectively). Two short motifs of GGSA and EGA are found repeatedly in tandem (amino acids 20–29 and 58–66). The Chow and Fasman algorithm predicts that this region contains four α-helices. The second portion, occupying the central and major part of the protein, consists of two typical RBDs. Two RBDs are arranged in tandem (amino acids 70–173 and 174–256) without any apparent spacer sequence between them. Further analysis of the structure of this portion will be presented later. The third portion, the C-terminal third of the protein, starts after a short repeat of glutamine (amino acids 262–268) and is characterized by high contents of glycine (32% of amino acids 269–306). In this region, three repeats of RGG are noted (amino acids 272–274, 282–284, and 334–336). RGG has been found in several RBD class RNA binding proteins (15). It has been suggested that it is an auxiliary motif responsible for protein-protein interaction or nonspecific nucleic acid binding (16).

Different cDNA Isoforms Resulting from Possible Alternative Splicing—Restriction mapping of other cDNA clones revealed the presence of several isoforms of cDNAs (Fig. 2). In summary, they can be classified into three types. One class is represented by the clone cDx4. Nucleotide sequencing of this cDNA reveals a 57-bp deletion (nucleotides 518–574 of cDx7, Fig. 1) in the 5′-coding region, and a 147-bp insertion in the 3′-coding region (between nucleotides 1138 and 1139 of cDx7). These variations result in a 19-amino acid deletion in the N-terminal portion of RBD-1 and a 49-amino acid insertion in the C-terminal Gly-rich region. The primary open reading frame is not affected by these deletions and insertions. The inserted 49-amino acid sequence revealed a unique feature. The sequence consists primarily of Gly, Tyr, and Asn (69% of 49-amino acid sequence). A motif of GY(G/N) repeatedly appears in this sequence. A stretch of eight amino acids, WNQGY(S/G)NY, appears in tandem twice. The second class is represented by cDx9. This clone

has both 57- and 147-bp insertions in the 5′- and 3′-region. The third class is represented by cDx7 with the 57-bp insertion but without the 147-bp insertion. When insertion or deletion is shown as + or − in the order of the 57- and 147-bp sequences, +/+ is cDx2, -5, and -9; +/− is cDx1, 6, 7, and 8; and −/+ is cDx4. Thus, it is suggested that +/+ and +/− classes are equally abundant and that clones having the 57-bp deletion in the 5′-portion are relatively minor. cDx8 is characterized by another insertion of 108 bp in the 3′-untranslated region.

Several hnRNP genes have been shown to produce variant mRNAs resulting from alternative splicing. This mechanism expands the complexity of hnRNP proteins. The differences found in our cDNA clones most likely comes from alternative splicing as well, although at present we do not have any direct evidence for it. We have not isolated cDNA of the −/− type. Thus, we could expect at least three different isoforms of mRNAs with or without the 57- and 147-bp insertions. The shortest +/− type encodes 306 amino acids with a molecular mass of 32.8 kDa. The intermediate −/+ type encodes 336 amino acids with a molecular mass of 36.2 kDa. Finally the longest +/+ mRNA predicts 355 amino acids with a molecular mass of 38.4 kDa. A previous SDS-polyacrylamide gel electrophoresis analysis identified proteins of apparent molecular masses of 41 kDa (possibly doublet) and 39 kDa as anti-hnRNP D monoclonal antibody-immunoreactive proteins in a TTAGGG-binding protein preparation (11). The presence of isoform mRNAs described above may explain the presence of native proteins with different apparent molecular masses. The proteins' mobility on SDS-polyacrylamide gel electrophoresis was slower than expected from the calculated molecular mass values. This may be in part due to the basic nature of these proteins (the calculated pI is about 8.8).

The hnRNP D0 Proteins as Members of the 2xRBD-Gly Family—A homology search of GenBank™ (release 87) indicated that many RNA-binding proteins have significant homology with the hnRNP D0 proteins: the DNA binding protein E2BP (14), the hnRNP C type protein (12), the A+U-rich RNA binding protein AUF1 (17, 18), the hnRNP type A/B protein (19), the CArG box binding protein (20), the D-box binding protein (21), the hrp40 proteins produced by *Drosophila squid* gene (22, 23), the hnRNP A1 protein, the hnRNP A2/B1 proteins, and the *Xenopus* hnRNP A2 family proteins. Among them, E2BP, the hnRNP C type protein, and AUF1 show an almost identical amino acid sequence with the hnRNP D0 proteins and thus are most likely derived from the same gene. Other genes like the hnRNP A1, hnRNP A2/B1, and hnRNP type A/B proteins are obviously distinct from, but homologous with, hnRNP D0. Finally, the mouse CArG box binding protein, the chicken D-box binding protein, and the *Drosophila squid* gene are derived from different species, and it is not known at present whether they are the counterparts of the hnRNP D0 gene of these species or not.

All of these proteins are characterized as having two RBDs in tandem in the N terminus (hereafter referred to as RBD-1 and RBD-2 from the N terminus) and a Gly-rich region, which typically contains the RGG motif, in the C terminus. The term "2xRBD-Gly group RNA binding protein" was coined to designate these proteins on the basis of their common structural organization (1). A compilation of an additional number of proteins, including hnRNP D0, is shown in Fig. 3, and these new members support the idea of the presence of this group of proteins. The RBD generally consists of about 90 amino acids. Two short stretches of sequence, RNP 1 and RNP 2 (eight and six amino acids, respectively) are highly conserved among the different RBD class RNA-binding proteins. Regions other than RNP 1 and 2 are less conserved. Significantly, proteins listed in

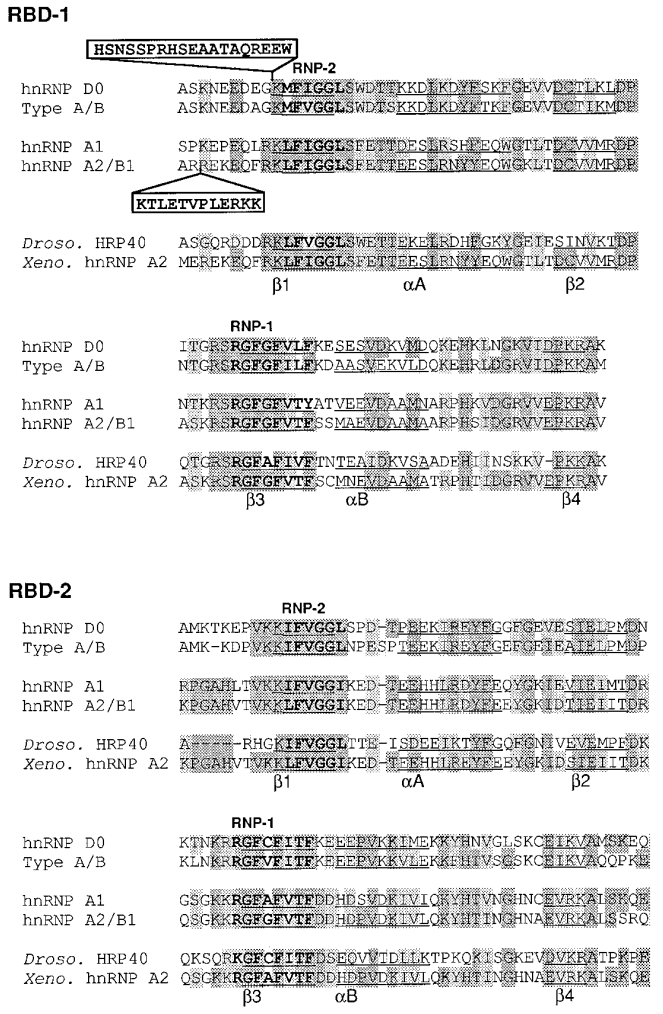

FIG. 3. **Alignment of the amino acid sequences of two RBDs of proteins having the 2xRBD-Gly structure.** The amino acid sequences of RBD-1 and RBD-2 of human hnRNP D0 (this study), type A/B hnRNP (19), hnRNP A1 (39), hnRNP A2/B1 (29), *Drosophila* HRP40 (22, 23), and *Xenopus* hnRNP A2 (40) are aligned manually. Identical and conserved amino acids among these proteins are marked by *heavy* and *light shading*, respectively. Positions of secondary structure are deduced from the study of hnRNP A1 (24) and indicated by *underlines*. RNP 2 hexamer and RNP 1 octamer are shown in *boldface letters*. Insertion of short peptides found in isoforms of hnRNP D0 and A2/B1 are shown using *boxes*.

Fig. 3 have conserved amino acid sequences, not only in RNP 1 and 2 but throughout the RBD. This long range conservation of amino acid sequences, along with a common structural organization, reinforces the presence of the 2xRBD-Gly group RNA binding proteins.

Recently, an NMR study of the N-terminal RBD of the human hnRNP A1 was reported (24). The study indicated that the hnRNP A1 RBD also forms four-stranded anti-parallel β-sheets as reported repeatedly with other RBDs (25, 26). Because 2xRBD-Gly type proteins are so closely related to each other, we are able to tentatively assign the secondary structures determined with the hnRNP A1 to other members of this group of proteins (Fig. 3). According to it, the 19-amino acid insertion of the hnRNP D0 found in RBD-1 is located at the N terminus of β1 of RBD-1.

Binding Properties of Recombinant Proteins—One of the most notable features of the hnRNP D0 proteins is their very stringent binding specificity with single-stranded nucleic acids. A previous study showed that protein binding to $d(TTAGGG)_4$ or $r(UUAGGG)_4$ was abolished by a single base substitution at

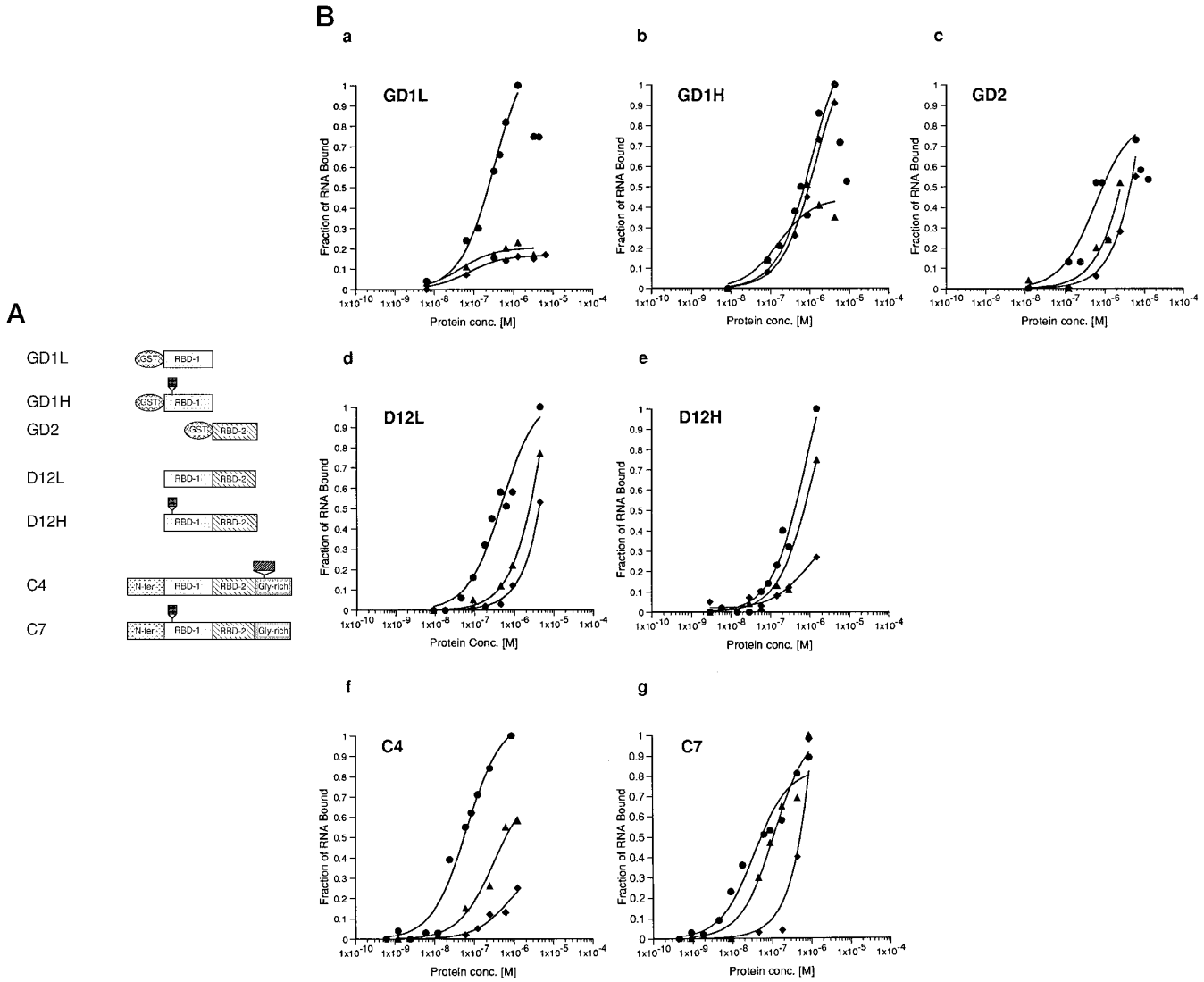

FIG. 4. **Binding properties of recombinant hnRNP D0 proteins.** *A*, structures of recombinant the hnRNP D0 proteins are schematically shown. Definition of the domain is the same as described in Figs. 1 and 2. *GST*, glutathione *S*-transferase, which is fused with RBD-1 and -2. GD1L and GD1H are RBD-1 fused to glutathione *S*-transferase, with (*H*) or without (*L*) the 19-amino acid insertion (*heavily shaded boxes*) at the N terminus of RBD-1. GD2 is RBD-2 fused with glutathione *S*-transferase. D12L and D12H are RBD-1 and -2, with (*H*) or without (*L*) the 19-amino acid insertion at the N terminus of RBD-1. C4 and C7 are different isoforms of the whole hnRNP D0 protein. The 19-amino acid insertion in RBD-1 is present in C7 but not in C4. A 49-amino acid insertion in the Gly-rich region is present in C4 but not in C7 (*hatched boxes*). *B*, the filter-binding assays were carried out and evaluated as described under "Materials and Methods" with GD1L (*part a*), GD1H (*part b*), GD2 (*part c*), D12L (*part d*), D12H (*part e*), C4 (*part f*), and C7 (*part g*). Oligoribonucleotide probes were as follows: J ●, rH4 (r(UUAGGG)4); H ▲, rH4X1 (r(UUGGGG)4); F ♦, rECGF r(GCAGCCUUGAUGACCUCGUGAACC).

each of the first four bases of repeat units. Thus, the proteins bind to $r(UUAGGG)_4$ but do not bind to $r(CUAGGG)_4$, $r(UCAGGG)_4$, $r(UUGGGG)_4$, or $r(UUAAGG)_4$, for example. Because the hnRNP D0 proteins exhibit the modular structure of 2xRBD-Gly, it is important to know the contribution of each domain to specific or nonspecific single-stranded DNA binding. To investigate, we constructed a series of truncated cDNAs having one or several domains.

Fig. 4*A* schematically depicts the structure of mutant recombinant proteins. GD1H and GD1L are RBD-1 fused to glutathione *S*-transferase, with (*H*) or without (*L*) insertion of the 19-amino acid sequence, respectively. GD2 is RBD-2 with glutathione *S*-transferase. These clones, having only one domain of RBD, were used as glutathione *S*-transferase fusion proteins because a single RBD is too small to be analyzed by filter binding assay. D12H and D12L are RBD-1 and RBD-2 with (*H*) or without (*L*) insertion of the 19-amino acid sequence. C4 and C7 are full-length recombinant proteins expressed from cDx4 (−/+ type) and cDx7 (+/− type), respectively.

Immunoblotting analysis of the recombinant proteins with an anti-hnRNP D monoclonal antibody 5B9 showed that GD2 is immunoreactive but that GD1H and GD1L are not (data not shown). This result supports the conclusion that the clones we isolated are for the hnRNP D0 and suggests that the epitope for the monoclonal antibody 5B9 is present in RBD-2.

Recombinant proteins were subjected to a filter binding assay to analyze their binding activities. Binding experiments were carried out by incubating variable amounts of recombinant proteins with constant amounts of oligonucleotides. Under these conditions, oligonucleotide concentrations (typically 1–10 nM) were always much lower than protein concentrations. The apparent K_d of binding reactions was estimated by the concentration of proteins at which half maximum binding was obtained. The oligoribonucleotide probes used in these assays

were rH4 (r(UUAGGG)$_4$), rH4X1 (r(UUGGGG)$_4$), and rECGF (r(GCAGCCUUGAUGACCUCGUGAACC)). rECGF was used as an unrelated sequence having the same length as rH4. Our previous study indicated that the purified HeLa cell proteins bind to rH4 but not to rH4X1 or rECGF. The following results were also obtained with DNA versions of these oligonucleotides, although the binding affinity was lower than that of RNA oligonucleotides (data not shown).

First, mutant recombinant proteins, having only one of the two RBDs, were examined. GD1L bound to rH4 with high binding affinity (the K_d is about 200 nM). In contrast, GD1L bound to either rH4X1 or rECGF much less efficiently (Fig. 4*B* (*part a*)). This specificity found between rH4 and rH4X1 indicated that a single RBD can strictly discriminate a single base change in the oligonucleotide. A recombinant protein of only glutathione *S*-transferase, excluding hnRNP D0, did not show any binding activity (data not shown). This result further confirms that the cDNA clones we isolated are for the UUAG-specific binding protein hnRNP D0.

Unexpectedly, sequence-specific binding observed with GD1L was detected in a somewhat different manner with GD1H (Fig. 4*B* (*part b*)). GD1H bound to rH4 with a K_d of about 1.1 μM. GD1H also bound to rH4X1 with nearly the same efficiency. Binding to rECGF was more efficient than GD1L showed. The major difference between GD1H and GD1L is the presence or absence of the 19-amino acid sequence at the N terminus of RBD-1. This result suggests that the presence of this insertion changes the preference of sequences to which hnRNP D0 proteins bind in a sequence-specific manner. GD2 showed intermediate binding properties between GD1L and GD1H. GD2 bound to rH4 with a K_d of about 320 nM. It bound to rH4X1 and rECGF to some extent, although the specificity discriminating between rH4 and rH4X1 was higher than that of GD1H (Fig. 4*B* (*part c*)).

The implication that the 19-amino acid insertion at RBD-1 may have a role in "sequence preference" was also suggested by the results of other recombinant proteins (Fig. 4*B* (*parts d–g*)). D12L, D12H, C4, and C7 bound to rH4 at a K_d of about 490, 880, 60, and 34 nM, respectively. No significant difference in the K_d of binding between rH4 and proteins was observed in the presence or absence of the 19-amino acid insertion. However, recombinant proteins with the insertion, C7 and D12H, also bound to rH4X1 as tightly as to rH4. In contrast, proteins without the 19-amino acid insertion, C4 and D12L bound to rH4X1 less efficiently. Therefore, all binding results are compatible with the idea that the 19-amino acid insertion modifies the sequence preference of hnRNP D0 protein resulting in the accommodation of rH4X1 as well as rH4.

Concerning proteins with several RNA-binding domains, it is of special interest to know whether or not one molecule of ligand bound to several domains simultaneously. From binding experiments, rH4 binds to one RBD with a K_d of 0.2–1 μM, and to two RBDs with a K_d of 0.5–0.9 μM. If both RBD-1 and -2 can bind to rH4 at the same time, the K_d for this binding should be much less than that of a single RBD. However, the K_d values were almost the same. Thus, it was concluded that RBD-1 and -2 of the hnRNP D0 protein cannot bind to rH4 simultaneously (numerical treatment for this discussion is available on request).

DISCUSSION

We have examined the structure of the hnRNP D0 protein cDNA and have studied the binding properties of recombinant proteins. Results showed that this protein is a member of the 2xRBD-Gly type RNA binding proteins. The notion of grouping the 2xRBD-Gly family is not based simply upon mere resemblance of the proteins but upon detailed structural and functional analysis as discussed below. A comparison of several cDNA clones revealed the presence of different isoforms of proteins, which are presumably derived from alternative splicing. One type of these different isoforms was due to a 19-amino acid insertion at the N terminus of RBD-1. Recombinant proteins having one or more combinations of modular domains were expressed. A filter binding assay of these mutant recombinant proteins with oligonucleotides clearly showed that a single RBD can bind to RNA sequence-specifically. In addition, "sequence preference" of the binding was found to be influenced by the presence or absence of the amino acid insertion in RBD-1.

Common Structure of the 2xRBD-Gly Family—RBD-class RNA binding proteins are very interesting, not only because many biologically important proteins are included in this class of proteins but also because these proteins typically have different types of RNA binding domains, which give rise to modular structures. 2xRBD-Gly type RNA binding proteins have two RBDs arranged in tandem in the N-terminal half, and a region rich in arginine and glycine (RGG box) in the C-terminal half. Regions rich in glycine, asparagine, and arginine were noted in several RNA-binding proteins and have been proposed to have a role in RNA-binding and protein-protein interaction, leading to cooperative binding (15, 16). The mutation study with the hnRNP U protein, which lacks the canonical RBD, clearly identified the C-terminal glycine-rich region as responsible for RNA binding. Since then, the region has been referred to as the RGG box (27). Although the RGG box has been identified in different contexts, the concurrent presence of two RBDs arranged in tandem in the N terminus and the RGG box in the C terminus in some proteins was noted (28, 29). The hnRNP A1 and A2/B1 proteins were thus first found to be members of this type of proteins. Recently, a detailed analysis of the structures of the hnRNP A1 and A2/B1 genes was reported (30). According to it, the exon/intron organization is conserved between these two genes. In particular, the presence of introns in the N-terminal region of the first RBD and RGG box was pointed out as a common feature shared by these genes. Moreover, alternatively spliced exons were reported to be present in the intron of the first RBD of the hnRNP A2/B1 gene and in the intron of the RGG box of the hnRNP A1 gene (29, 31). In this report, we show that the hnRNP D0 protein has a 2xRBD-Gly structure. HnRNP D0 differs from hnRNP A1 and A2/B1 in that it has a longer N-terminal region that shows no obvious homology with hnRNP A1 or A2/B1. However, the hnRNP D0 cDNAs show the characteristic features identified in hnRNP A1 and A2/B1. First, a 57-bp nucleotide insertion resulting in a 19-amino acid insertion in the N terminus of the first RBD was noted in some isoforms. This correlates with the alternative exon coding for the 12-amino acid insertion found in the N-terminal RBD of the hnRNP A2/B1 protein (Fig. 3). Second, some hnRNP D0 isoforms showed a 147-bp nucleotide insertion resulting in a 49-amino acid insertion in the RGG box. This correlates with the alternative exon VII bis of hnRNP A1 resulting in a 52-amino acid insertion in the RGG box. Finally, the two insertions found in hnRNP A1 and D0 are commonly abundant in Gly, Tyr, and Asn.

Besides this macroscopic similarity among the hnRNP A1, A2/B1, and D0 proteins, nucleotide sequences of RBDs are also highly conserved (Fig. 3). RBD is made up of about 80–90 amino acids. However, only two relatively short amino acids sequences, RNP 1 and 2, are highly conserved among different RBD class RNA binding proteins (2, 32). Most of the RBD sequences other than RNP 1 and 2 are even far less conserved. From this point, it is remarkable that the comparison of RBDs derived from hnRNP A1, A2/B1, and D0 has shown a signifi-

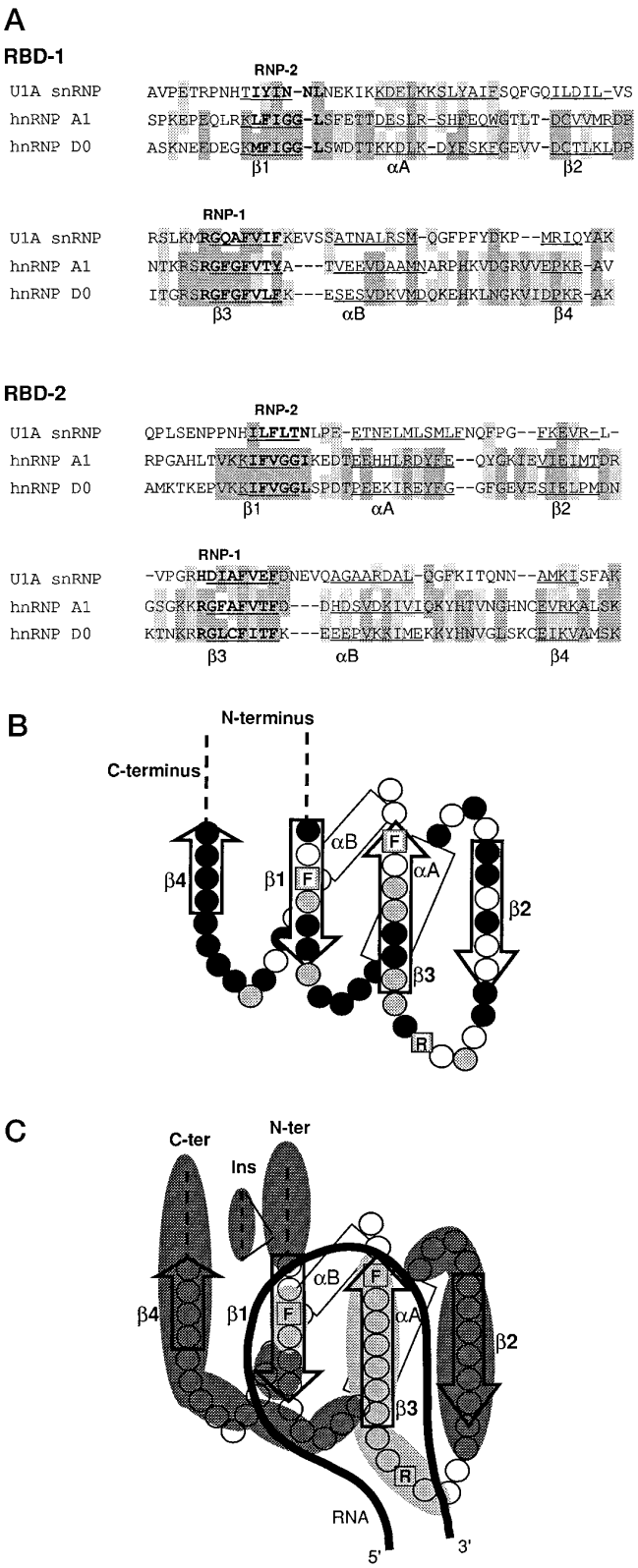

FIG. 5. **Mapping of general and specific RNA binding sites on a structural model of the hnRNP D0 protein.** *A*, a comparison of amino acid sequences of N-terminal (*RBD-1*) and C-terminal (*RBD-2*) RBDs of snRNP U1A protein (41), the hnRNP D0 protein, and the hnRNP A1 protein. Assignment of secondary structures is from Refs. 24 and 25). Positions of secondary structures are marked by *underlining*. The prediction of the secondary structure with the hnRNP D0 protein is based solely upon sequence similarity to the hnRNP A1 protein. RNP 2 hexamer and RNP 1 octamer are shown in *boldface letters*. *B*, the mapping of conserved amino acids on a structural model of the hnRNP D0 RBD-1. The four-stranded β-sheets model of the hnRNP D0 protein

cant identity throughout the RBD regions (Fig. 3; see Fig. 5*A* for a comparison between the 2xRBD-Gly and non-2xRBD-Gly proteins). The common structural organization, producing similar isoform proteins among hnRNP A1, A2/B1, and D0 and highly conserved amino acid sequences throughout RBDs, strongly suggests that these genes belong to a closely related gene family having a common and old ancestral gene. This notion is consistent with the fact that invertebrates like *Drosophila* contain only 2xRBD-Gly type hnRNP proteins (22, 33), whereas vertebrates have many different types of hnRNPs.

Common Binding Properties of the 2xRBD-Gly Family—Even though 2xRBD-Gly proteins show a common protein structure, it was not necessarily expected that they would have the same sequence specificity for binding. We purified human proteins that bind to d(TTAGGG)-repeats and identified hnRNP A1, A2/B1, D0, E0, and nucleolin (11). HnRNP E0 is very closely related to hnRNP D0 and is a member of the 2xRBD-Gly family.[2] Nucleolin, a 100-kDa ribosomal protein, consists of four RBDs arranged in tandem and a Gly-rich sequence in the C terminus (34). Therefore, nucleolin can be considered as a distantly related protein to the 2xRBD-Gly family. In summary, hnRNP A1, A2/B1, D0, E0, and nucleolin can be grouped together because they have similar protein structure and binding specificity.

Determinants of RNA Binding Specificity—In this study, we showed that recombinant proteins having only RBD-1 or -2 of hnRNP D0 (GD1 or GD2) have a similar binding affinity with proteins having both RBD-1 and -2 (D12) in binding to rHum4. This implies that two RBDs of the hnRNP D0 protein are not occupied by a particular 24-nucleotide oligonucleotide rHum4 at any one time. We do not exclude the possibility that a single-stranded nucleic acid binds to the two RBDs simultaneously. Rather, we think that it is highly possible that longer single-stranded nucleic acids are recognized by several RBDs, as suggested by others (35). However, 24 nucleotides is obviously shorter than that recognized by two RBDs. Alternatively, two RBDs may not bind to any RNA simultaneously as concluded by others (36). In any case, this study indicated that each of the two RBDs of hnRNP D0 specifically binds to rHum4. On the other hand, a methylation interference experiment revealed that only one out of four repeats of $d(TTAGGG)_4$ is recognized by the native hnRNP D0 protein.[3] These observations form the basis for the binding mode of the hnRNP D0 protein in that one RBD binds to one or few repeats of the UUAGGG sequence.

Recently, Oubridge *et al.* (37) reported a crystal structure of

[2] F. Ishikawa and Y. Kajita, unpublished observation.
[3] F. Ishikawa and T. R. Cech, unpublished observation.

RBD-1 is deduced from a NMR study of the hnRNP A1 protein (24), the hnRNP C protein (26), and an x-ray crystallography study of snRNP U1A protein (25). Each *circle* corresponds to an amino acid residue. Amino acids of hnRNP D0 that are found identical or conserved among all of the snRNP U1A, hnRNP D0, and hnRNP A1 proteins in *A*. are shown on a structure model by *stippled circles*. Amino acids that are found identical or conserved between the hnRNP D0 and hnRNP A1 proteins but not with snRNP U1A in *A*, are shown by *filled circles*. Highly conserved aromatic amino acids in β1 and β3 and a basic amino acid in loop β2β3 are indicated using *squares*. *F* and *R* represent phenylalanine and arginine, respectively. Amino acids present in α-helices A and B are not shown for clarity. *C*, a model of the distribution of general (*lightly shaded*) and specific (*heavily shaded*) binding sites of RBD. The positions of the N- and C-terminal portions of RBD (*N-ter and C-ter*), the amino acid insertion found in isoforms of hnRNP D0 (*Ins*), and the RNA substrate bound with RBDs (*thick line*) are also schematically shown. These positions are not experimentally determined with hnRNP D0 but rather deduced from those of the U1A protein (37) (see "Discussion").

the RBD of the snRNP U1A protein complexed with an oligoribonucleotide of U1 snRNA hairpin II. Fig. 5*A* shows the comparison among RBDs of U1A protein, the hnRNP A1 protein, and the hnRNP D0 protein. U1A protein recognizes specifically 7 nucleotides of the 5′-end of the 10-nucleotide loop, U1 hairpin II. Oubridge *et al.* (37) showed that the 7-nucleotide bases are extensively recognized by the surface of the β-sheets, maintaining intimate contact with the highly conserved RNP 2 and RNP 1 motifs. Because the overall structure and length of the RBD and its substrate are similar among U1A, hnRNP D0, and hnRNP A1, it may be possible to use the higher ordered structure reported by Oubridge *et al.* (37) as a starting point for constructing a model of binding between hnRNP D0 and an oligonucleotide.

It has been suggested that the flat surface of the four-stranded β-sheets of RBD binds to RNA in two different modes, one in specific and the other in nonspecific general binding. It has also been suggested that these functionally discernible types of binding are performed by molecularly different regions of the β-sheets (3, 38). Two highly conserved short sequences, RNP 2 and 1, which are located in the central two β-sheets (β1 and β3, respectively) are candidates for regions functioning in general binding. Regions variable among different RBDs, like the β2β3 loop and the C-terminal portion of RBD, are candidate regions for specific binding. However, further mapping of these two distinct binding sites has not been proposed.

HnRNP D0 and hnRNP A1 showed a similar binding specificity, which is different from that of U1A. Therefore, amino acid sequences in Fig. 5*A* that are conserved between hnRNP D0 and A1 but not with U1A may be responsible for specific binding. On the other hand, regions conserved among these three proteins in common may be candidate regions for general binding. In Fig. 5*A*, identical or conserved amino acid residues among three proteins are indicated. Fig. 5*B* is the result of mapping on a structural model of the two types of amino acids of the hnRNP D0 protein. The result shows interesting distributions of these residues. Residues conserved in three proteins are mostly located in β1 and β3, which form the central "umbilicus" of the platform (Fig. 5*B*, *stippled circles*). In contrast, amino acid residues conserved only in hnRNP A1 and D0 are distributed at the margin of the platform (*filled circles*). These include the start and end regions of β1; loops connecting β1 and αA, αA and β2; β2; loops connecting β2 and β3, αB and β4; and the entire β4. These regions, which are apparently distributed at intervals in the primary sequence, precisely trace the rim of the platform. The characteristic distributions of the "general" and "specific" amino acids (central *versus* marginal) predict the position of bound RNA on the RBD platform. For RNA to keep contact with both general and specific amino acid residues, it needs to be positioned by being fitted into the clefts formed by general β-sheets (β1 and β3) and specific β-sheets (β2 and β4). This is exactly what was found in the structural study with the U1A-RNA complex (37). They found that the U1 loop II binds to U1A as schematically shown in Fig. 5*C*. General β1 and specific β4 have contact with the ascending 5′-half of the loop. Aromatic amino acids conserved very well in RNP 1 and 2, which occupy the upper portions of general β3 and β1 in the orientation of Fig. 5, interact with the top of the loop (indicated by *squares*). The 3′-descending loop is recognized relatively loosely by specific β2. Finally, the highly conserved basic amino acid in the β2β3 loop (indicated by a *square*) interacts with the neck of the loop. It is remarkable that we found that both the loop between αB and β4 and β4 are composed mainly from 2xRBD-Gly-specific amino acid residues. In U1A, this region was shown to have a tight interaction with RNA in a sequence-specific manner. Thus, this region may be the major determinant of sequence specificity of RBD. In contrast, in β2, we mapped relatively few specific amino acids compared with β4. This correlates with the observation with U1A that β2 has relatively few contacts with RNA. Three nucleotides of the 3′-end of the loop that are positioned around here are known not to participate in sequence-specific binding. The loop connecting β2 and β3 shows a somewhat different nature from other regions, because this region is a mixture of general and specific amino acids. In U1A, this loop penetrates and opens up the RNA loop. Therefore, it is possible that amino acids present in this loop participate in both general and specific binding.

In this context, the observations that the N-terminal end (this study) and the C-terminal end (38) of RBD are concerned in sequence-specific binding can be easily understood, because these regions are presumably positioned at the platform rim (Fig. 5*C*). We propose a model, as shown in Fig. 5*C* for the map of the general and sequence-specific RNA binding sites of RBD. The margin of the RBD platform, including the N and C termini of RBD, β4, β2, and several loops (shown by *heavy shading* in Fig. 5*C*) interacts with the RNA sequence specifically. The central part of RBD containing β1 and β3 (shown by *light shading*) that contains the highly conserved RNA 1 and 2 motifs interacts with RNA in a nonspecific general manner. This model should be examined by a direct structural analysis of the hnRNP D0 protein complexed with an RNA substrate, which is currently on its way.

Acknowledgments—We are grateful to Dr. A. Sarai and Dr. E. Muneyuki for critical comments and suggestions, Dr. G. Dreyfuss for monoclonal antibodies, S. McCulloth and Dr. M. Katahira for critical reading and comments on the manuscript, and Dr. S. Nishimura for continuous encouragement. Excellent technical and secretarial work by M. Komatsu is acknowledged.

REFERENCES

1. Dreyfuss, G., Matunis, M. J., Piñol-Roma, S., and Burd, C. G. (1993) *Annu. Rev. Biochem.* **62,** 289–321
2. Kenan, D. J., Query, C. C., and Keene, J. D. (1991) *Trends Biochem. Sci.* **16,** 214–220
3. Kiledjian, M., Burd, C. G., Görlach, M., Portman, D. S., and Dreyfuss, G. (1994) in *RNA-protein interactions* (Nagai, K., and Mattaj, I. W., eds) pp. 127–149, Oxford University Press, Oxford
4. Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. (1986) *Science* **231,** 1534–1539
5. Sierakowska, H., Szer, W., Furdon, P. J., and Kole, R. (1986) *Nucleic Acids Res.* **14,** 5241–5254
6. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992) *Mol. Cell. Biol.* **12,** 3165–3175
7. Bennett, M., Michaud, S., Kingston, J., and Reed, R. (1992) *Genes & Dev.* **6,** 1986–2000
8. Mayeda, A., and Krainer, A. R. (1992) *Cell* **68,** 365–375
9. Mayeda, A., Helfman, D. M., and Krainer, A. R. (1993) *Mol. Cell. Biol.* **13,** 2993–3001
10. Mayeda, A., Munroe, S. H., Cáceres, J. F., and Krainer, A. R. (1994) *EMBO J.* **13,** 5483–5495
11. Ishikawa, F., Matunis, M. J., Dreyfuss, G., and Cech, T. R. (1993) *Mol. Cell. Biol.* **13,** 4301–4310
12. Lahili, D. K., and Thomas, J. O. (1986) *Nucleic Acids Res.* **14,** 4077–4094
13. Sharp, Z. D., Smith, K. P., Cao, Z., and Helsel, S. (1990) *Biochim. Biophys. Acta* **1048,** 306–309
14. Tay, N., Chan, S.-H., and Ren, E.-C. (1992) *J. Virol.* **66,** 6841–6848
15. Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988) *J. Biol. Chem.* **263,** 1063–1071
16. Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., and Williams, K. R. (1991) *Biochemistry* **30,** 2968–2976
17. Zhang, W., Wagner, B. J., Ehrenman, K., Schaefe, A. W., DeMaria, C. T., Crater, D., DeHaven, K., Long, L., and Brewer, G. (1993) *Mol. Cell. Biol.* **13,** 7652–7665
18. Ehrenman, K., Long, L., Wagner, B. J., and Brewer, G. (1994) *Gene* (*Amst.*) **149,** 315–319
19. Khan, F. A., Jaiswal, A. K., and Szer, W. (1991) *FEBS Lett.* **290,** 159–161
20. Kamada, S., and Miwa, T. (1992) *Gene* (*Amst.*) **119,** 229–236
21. Smidt, M. P., Wijnholds, J., Snippe, L., van Keulen, G., and Greet, A. B. (1994) *Biochim. Biophys. Acta* **1219,** 115–120
22. Matunis, M. J., Matunis, E. L., and Dreyfuss, G. (1992) *J. Cell Biol.* **116,** 245–255
23. Kelly, R. L. (1993) *Genes & Dev.* **7,** 948–960
24. Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M., and Gronenborn, A. M. (1994) *Biochemistry* **33,** 2852–2858
25. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) *Nature* (*Lond.*) **348,** 515–520

26. Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., and Dreyfuss, G. (1992) *EMBO J.* **11,** 3289–3295
27. Kiledjian, M., and Dreyfuss, G. (1992) *EMBO J.* **11,** 2655–2664
28. Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., and Wilson, S. H. (1986) *J. Biol. Chem.* **261,** 3536–3543
29. Burd, C. G., Swanson, M. S., Görlach, M., and Dreyfuss, G. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 9788–9792
30. Biamonti, G., Ruggiu, M., Saccone, S., Valle, G. D., and Riva, S. (1994) *Nucleic Acids Res.* **22,** 1996–2002
31. Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G., and Riva, S. (1990) *EMBO J.* **9,** 1229–1235
32. Birney, E., Kumar, S., and Krainer, A. R. (1993) *Nucleic Acids Res.* **21,** 5803–5816
33. Matunis, E. L., Matunis, M. J., and Dreyfuss, G. (1992) *J. Cell Biol.* **116,** 257–269
34. Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B., and Rurns, A. L. (1990) *J. Biol. Chem.* **265,** 14922–14931
35. Shamoo, Y., Abdul-Manan, N., Patten, A. M., Crawford, J. K., Pellegrini, M. C., and Williams, K. R. (1994) *Biochemistry* **33,** 8272–8281
36. Casas-Finet, J. R., Smith, J. D. J., Kumar, A., Kim, J. G., Wilson, S. H., and Karpel, R. L. (1993) *J. Mol. Biol.* **229,** 873–889
37. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1994) *Nature* (*Lond.*) **372,** 432–438
38. Görlach, M., Burd, C. G., and Dreyfuss, G. (1994) *J Biol. Chem.* **269,** 23074–23078
39. Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M. T., Riva, S., and Morandi, C. (1988) *Nucleic Acids Res.* **16,** 3751–3770
40. Good, P. J., Rebbert, M. L., and Dawid, I. B. (1993) *Nucleic Acids Res.* **21,** 999-1006
41. Sillekens, P. T., Habets, W. J., Beijer, R. P., and van Venrooij, W. J. (1987) *EMBO J.* **6,** 3841–3848