A Specific Isoform of hnRNP D Interacts with DNA in the LR1 Heterodimer: Canonical RNA Binding Motifs in a Sequence-specific Duplex DNA Binding Protein

(Received for publication, June 16, 1998, and in revised form, July 7, 1998)

Laurie A. Dempsey†‡§, L. A. Hanakahi†‡§, and Nancy Maizels‡¶

From the Departments of Molecular Biophysics and Biochemistry and Genetics, Yale University, New Haven, Connecticut 06520-8114

The B cell-specific, sequence-specific duplex DNA-binding protein LR1 is a transcriptional activator and may also function in heavy chain class switch recombination. LR1 is composed of two polypeptides, a 106-kDa subunit that is nucleolin, and a 45-kDa subunit that we now show to be a specific isoform of hnRNP D. hnRNP D and nucleolin both contain canonical RNA binding domains (RBDs also called RRMs) and Arg-Gly-Gly (RGG) motifs. Although these motifs are not commonly associated with sequence-specific recognition of duplex DNA, nonetheless LR1 binds duplex DNA with high affinity ($K_D = 1.8$ nM) and clear sequence specificity. Two RBD-RGG proteins can therefore combine to produce a sequence-specific duplex DNA binding protein.

LR1 was first identified as an activity that binds duplex sites in the immunoglobulin heavy chain switch regions (1). LR1 also activates transcription at two promoters that function in transformed B cells, the c-myc P1 promoter (2) and the Epstein-Barr virus F promoter (3). LR1 DNA binding activity is induced in primary B cells activated for switch recombination and is abundant in pre-B and B cell lines, but is not found in resting primary B cells or B cell lines representing late stages of development, such as plasmacytomas or myelomas, or in other cell types, including T cells and fibroblasts (1). This spectrum of LR1 activity correlates with the developmental spectrum of cells that can support switch region-mediated recombination of transfected extrachromosomal switch substrates (4–8).

Highly purified preparations of LR1 consist of two polypeptides of 106 and 45 kDa (9). We have recently demonstrated that the 106-kDa polypeptide is nucleolin (10), a highly conserved and abundant protein that is concentrated within the nucleoli, where RNA biogenesis occurs. Here we report that the 45-kDa component of LR1 is a specific isoform of hnRNP D, which we identify as virtue of their interaction with hnRNA within the nucleus of mammalian cells (11–14). hnRNP D has been reported to bind in vitro to 3′ splice sites (12, 15, 16) and to the G-rich mammalian telomeric repeat as DNA and RNA (16). The murine and human hnRNP D polypeptides are remarkably conserved (97% identical and 99% similar), suggesting that this protein plays a critical role in mammalian cells (17).

As illustrated in Fig. 1, hnRNP D is actually a small family of polypeptides. These polypeptides derive from mRNAs that are related by alternative splicing of two coding exons, exon 2 and exon 7 (16, 17). We show that the isoform of hnRNP D that contacts DNA in the LR1 heterodimer contains sequences encoded by alternative exon 2.

Nucleolin and hnRNP D contain similar structural motifs: RNA binding domains (RBDs, also referred to as RNA recognition motifs or RRMs, reviewed in Refs. 13, 14, and 18) and reiterations of the sequence Arg-Gly-Gly (RGG), in which the arginine residues are dimethylated (19–21) (Fig. 1). These domains are typical of proteins that bind RNA, not duplex DNA. Nonetheless, LR1 binds with high affinity ($K_D = 1.8$ nM) and clear sequence specificity to duplex DNA. Our results therefore demonstrate that two RBD proteins can interact to form a factor that interacts tightly and specifically with duplex DNA.

EXPERIMENTAL PROCEDURES

DNA Labeling and DNA Binding Assays—The S1 duplex was formed by annealing top strand synthetic oligonucleotide, d(GATCCT-CCTGGGTCAAGGCTGAATAGACGCAGGGGA), with bottom strand oligonucleotide, d(GATCTCCCCTGCGTCTATTCAGCCTTGACCCAG-GAG), and labeled by filling with Klenow polymerase, as described previously (1, 10). Mutant duplexes were synthesized in which A, I, or 2-aminopurine was substituted at the underlined Gs in the top strand, and labeled with 32P. Mutant top strands containing inosine or 2-aminopurine substitutions were annealed to the bottom strand shown above, and mutant top strands containing substitutions of A were annealed to bottom strands carrying T at the complementary position. Binding to duplex DNA was carried out in 15-μl reactions containing 20 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.1% Nonidet P-40, 2.5% glycerol, 2% polyvinyl alcohol, 100 μg/ml bovine serum albumin, 4 fmol of 32P-labeled duplex DNA, and 0.2 pmol of purified LR1 (except where otherwise indicated) for 15 min at room temperature. Protein-DNA complexes were resolved by electrophoresis on 5% polyacrylamide gels in 90 mM Tris-borate, 1 mM EDTA, pH 8.3. Binding reactions with crude nuclear extract included 0.2 μg of poly(dI-dC) (Boehringer Mannheim) as nonspecific competitor; competitor concentration was decreased at successive purification steps, as described previously (9). When antibodies were included, they were pre-incubated with protein in a 10-μl volume on ice for 30 min before the addition of 10 μl of binding buffer containing labeled DNA.

Dissociation constants were determined by gel mobility shift assays carried out in the absence of competitor DNA and in the presence of highly purified LR1 at concentrations ranging from 0.05 to 2.5 nM.
Protein-DNA complex formation was quantitated by phosphoimager analysis of the dried gels, and $K_d$ values were calculated by plotting the fraction of free DNA at each protein concentration and extrapolating to 100% binding. Reported $K_d$ values are averages from at least three separate experiments. Apparent $K_d$ values were unaffected by changes in DNA concentration.

**LR1 Protein Purification**—LR1 was purified from 40 liters of PD31 murine pre-B cells (60 × 10^9 cells) cultured in RPMI 1640 containing 10% fetal bovine serum and 50 μM β-mercaptoethanol. During purification, LR1 binding to two diagnostic duplex sites was assayed, the wild-type S11 duplex and the G14A mutant duplex. Binding to G14A is reduced about 13-fold relative to S11 binding (Ref. 1 and Fig. 2). All purification steps were carried out at 4 °C, and extracts and fractions were stored at −70 °C. All buffer solutions contained 0.5 mM phenylmethylsulfonyl fluoride and 10 μg/mL leupeptin to inhibit proteases. Protein concentrations were determined by Bradford microassay (Bio-Rad).

Nuclear extract was prepared as described previously (1), except that cells were lysed at 4 °C in hypotonic buffer containing 10 mM HEPES, pH 7.9, 10 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, and 0.1 mM EGTA (22). Nuclei were pelleted, and extraction was carried out in buffer containing 20 mM HEPES, pH 7.9, 0.42 mM KCl, 1 mM DTT, 1 mM EDTA, 1 mM EGTA, 20% glycerol (3 ml/10^9 cells for extraction) by gentle rocking at 4 °C for 45 min. Pilot experiments showed that inclusion of chelating agents increased recovery of LR1 binding activity from 2- to 4-fold. Debris was pelleted by centrifugation at 12,000 rpm for 45 min (17,400 × g). Nuclear extract was dialyzed in buffer S (20 mM Tris, pH 7.5, 1 mM DTT, 1 mM EDTA, 0.2 mM EGTA, 10% glycerol, 0.01% Nonidet P-40) containing 0.1 mM KCl, then fractionated on a heparin Hi-Trap FPLC column (Amersham Pharmacia Biotech) pre-equilibrated in the same buffer. The column was eluted using a linear 0.1–1.0 M KCl gradient, and fractions containing LR1 DNA binding activity (0.22–0.30 M KCl) were pooled and diluted in buffer S to a final salt concentration of 0.15 M KCl and 0.1 M NaCl. Fractions that were active in DNA binding activity were pooled and fractionated by preparative SDS-polyacrylamide gel electrophoresis, and the 45-kDa band was submitted for sequence analysis at the Keck Biotechnology Resources Center, Yale University School of Medicine. Starting with 600 mg of nuclear extract (as determined by Bradford protein microassay), we obtained approximately 50 μg of purified protein (as determined by mass spectroscopy). The overall purification achieved was therefore estimated to be approximately 12,000-fold.

**Expression of Recombinant His<sub>6</sub>-hnRNP D and Maltose Binding Protein-HnRNP D Fusion Proteins**—His<sub>6</sub>-tagged recombinant hnRNP D was produced by cloning a murine *Hnrpd* DNA NcoI-NotI fragment from the pBluescript vector (17) into the pET30A(+) (Novagen) bacterial expression vector. This fusion deleted a 31-residue N-terminal region rich in alanine (9/31 residues) and glycine (6/31 residues), leaving the 2 RBD domains and C-terminal RGG motifs intact. His<sub>6</sub>-tagged hnRNP D was expressed and purified by nickel-chelate chromatography.
phy (Novagen). The yield of recombinant protein recovered was typically 2–5% total bacterial protein. Concentration of purified recombinant protein was determined by Bradford microassay (Bio-Rad).

The N terminus of nucleolin contains long acid regions interspersed with clusters of basic amino acids and cannot be expressed in E. coli. A truncated version of nucleolin lacking amino acids 1–283 but containing all 4 RBDs and the RGG repeats was expressed from the pMalNuc construct as a chimerà fused at its N terminus to E. coli maltose binding protein (MBP). The MBP-nucleolin chimera was then purified as described previously (10).

Peptide Synthesis, Antibody Production, and Immunoblotting—Anti-hnRNP D antibodies were raised against peptide CYGKVSRRGGHQN-SYKPY, which corresponds to the C terminus of hnRNP D, and peptide CSPRHSEAATAQRE and CGPSQNWNQYSYN, which correspond to sequences encoded by hnRNP D alternative exons 2 and 7, respectively (17). The peptides were synthesized at the W. H. Keck facility (Yale University School of Medicine) and coupled to maleimide-activated keyhole limpet hemocyanin carrier protein (Pierce) according to the manufacturer’s instructions; coupling efficiency was greater than 85% as estimated by measuring the free sulphydryl concentration with Ellman’s reagent (Pierce). The keyhole limpet hemocyanin conjugate was dialyzed against phosphate-buffered saline containing 0.9 M NaCl. Rabbits were immunized, boosted once 3 weeks later and at 2 week intervals thereafter with 100 µg/injection. The anti-hnRNP D exon 2 and exon 7 antibodies were affinity-purified on protein A-Sepharose (Amerham Pharmacia Biotech) and passed over poly(G)-agarose (Sigma) to remove non-specific DNA binding activities. Anti-C-terminal antibodies were affinity-purified on protein A-Sepharose (Sigma) to remove non-specific DNA binding activities. Anti-C-terminal antibodies were passed over poly(G)-agarose but not protein A-purified. Rabbit polyclonal antibodies raised against amino acids 284–709 of nucleolin have been previously described (10).

For immunoblots, 50 ng of purified recombinant hnRNP D (each isoform) was resolved on 12.5% SDS-polyacrylamide gel electrophoresis and electrotransferred to a nitrocellulose membrane. Subsequent steps were carried out at room temperature. Blots were blocked in 3% nonfat dry milk, 1% bovine serum albumin in TBST (10 mM Tris, pH 7.5, 150 mM NaCl, 0.05% Tween 20), washed with TBST 4 times, 15 min/wash, and probed with anti-hnRNP D antisera at 1:2000 dilution. Antibody was detected with Amersham ECL kits using a 1:2000 dilution of horseradish peroxidase-coupled anti-rabbit-Ig antibody in TBST.

RESULTS

Highly Purified LR1 Binds Duplex DNA with High Affinity ($K_d = 1.8$ nM)—LR1 was purified 12,000-fold from nuclear extract of PD31 pre-B cells by four chromatographic steps: heparin agarose, oligonucleotide capture, Mono Q, and Mono S. Using the highly purified LR1 preparation, we determined the specificity of the LR1/DNA interaction by measuring the dissociation constants of LR1 binding to a panel of DNA duplexes containing single base substitutions. Mutations at most positions in the LR1-S1 site, GGTCAAGGCTGAATA, do not affect LR1 binding, but mutation of either of the two underlined G residues to A diminishes binding significantly (1, 10). By substitution of inosine and 2-aminopurine for critical guanines, it is possible to distinguish major groove contacts (identical for guanine and inosine) and minor groove contacts (identical for guanine and 2-aminopurine). The underlined G residues were therefore replaced with inosine or 2-aminopurine, and dissociation constants of LR1 binding to labeled synthetic duplexes were analyzed by gel mobility shift. An example of a set of binding reactions used in this analysis is shown in Fig. 2A. Dissociation constants were calculated as described under “Experimental Procedures,” assuming that the stoichiometry of binding was 1:1 (LR1 heterodimer:DNA).

The results of these experiments showed that LR1 bound the 54-bp duplex site with $K_d = 1.8$ nM (Fig. 2B). Binding was not affected by substitution of inosine at G8 or 2-aminopurine at G14. Binding was diminished 4-fold by substitution of 2-aminopurine at G8 and 7.5-fold by substitution of inosine at G14. LR1 therefore contacts the major groove at G8, where it interacts with the O6 of guanine, and the minor groove at G14, where it interacts with the N2 of guanine. The minor groove interaction would also be available if a C-G pair replaced a G-C pair, consistent with the presence of C at this position in one LR1 site in the Epstein-Barr virus F promoter (3). As the major and minor groove contacts are separated by 6 base pairs, the critical DNA contacts for LR1 appear to lie on a single face of the DNA duplex (Fig. 2C).

The 45-kDa Component of LR1 Is hnRNP D—The 45-kDa polypeptide of LR1 was submitted to the W. H. Keck Biotechnology Resource Center for sequence analysis, and two tryptic peptides were sequenced. The molecular mass of the larger was 1494 daltons, and its sequence was XGHGQNSYKPY. The identification of this subunit of LR1 as hnRNP D is consistent with the size of hnRNP D, approximately 45 kDa (11, 12). The first position of the latter peptide (indicated as X) could not be determined unambiguously, but if dimethylated arginine were present at this position, the predicted molecular mass of the polypeptide, 1335 daltons, would be identical to the experimentally determined molecular mass. The hnRNP proteins contain dimethylarginine (21), and (m6D)GGHQNSYKPY is therefore predicted to be the N-terminal residue of this peptide.

To verify that hnRNP D is a component of LR1, we asked if anti-hnRNP D antibodies recognized purified LR1. The monoclonal anti-hnRNP D antibody 5B9 (generously provided by Dr. Gideon Dreyfuss, University of Pennsylvania) recognized the 42–45-kDa polypeptides in the highly purified LR1 preparation, as assayed by immunoblotting (data not shown). We raised antibodies against a synthetic peptide bearing the C-terminal sequence of hnRNP D, demonstrated that these antibodies were active in immunoblots (not shown), and then assayed the ability of these antibodies to recognize LR1 DNA binding activity. Gel mobility shift analysis showed that the anti-hnRNP D C-terminal immune serum recognized the LR1-DNA complex and retarded the mobility of the shifted band (Fig. 3). Pre-immune serum had no effect nor did pre-immune or immune serum alter the mobility of DNA in the absence of LR1.

A Specific Isoform of hnRNP D Contacts DNA in the LR1 Heterodimer—Three distinct classes of hnRNP D cDNA clones can be found in a murine pre-B cell library, and they result...
We have previously reported that treatment of LR1 with anti-nucleolin antibodies neither inhibited binding nor resulted in a supershift; instead, the antibodies produced a characteristic "subshifted" band (10). We initially interpreted the subshift as resulting from specific depletion of one component, nucleolin, from the binding complex, leaving the other component in contact with the DNA. We established that hnRNP D is indeed in the subshifted complex by treating purified LR1 with anti-nucleolin antibodies alone or with both anti-nucleolin antibodies and antibodies raised against exon 2 sequences. The anti-nucleolin antibodies produced the characteristic subshift, as previously reported (10), and the anti-exon 2 antibodies disrupted the subshift (Fig. 4B). These results provide further evidence that the isoform of hnRNP D in LR1 contains sequences encoded by alternative exon 2.

Analogous experiments were carried out to test the effect of the anti-exon 7 antibodies on LR1 duplex DNA binding activity. These antibodies had no effect on LR1 DNA binding activity, either alone or in the presence of the anti-nucleolin antibody (Fig. 4B).

Taken together, the results in Fig. 4 show that the 45-kDa component of the LR1 heterodimer is a specific isoform of hnRNP D, which contains sequences encoded by exon 2. These results also strongly suggest that the isoform of hnRNP D in LR1 does not contain sequences encoded by exon 7. Although it is in principle possible that exon 7 sequences may be present in the LR1 heterodimer but not accessible to or reactive with the antibody, additional experiments described below argue against this possibility.

hnRNP D Isoforms Containing Exon 7 Exert a Dominant Negative Effect on LR1 Duplex DNA Binding Activity—We tested the ability of each of the isoforms of hnRNP D to combine with recombinant MBP-nucleolin and reconstitute LR1 duplex DNA binding activity. We found that LR1 binding activity could not be reconstituted from recombinant protein expressed in E. coli (Fig. 5A). This may reflect a requirement for phosphorylation that has been documented in previous experiments (1), for other modifications, including dimethylarginylation (20), and/or for the presence of the N-terminal region of nucleolin, which cannot be expressed in E. coli and was therefore not included in the MBP-nucleolin expression construct.

Surprisingly, however, when we tested the effect of adding each of the three isoforms of recombinant hnRNP D to highly purified preparations of LR1, we found that either hnRNP D-M07 or hnRNP D-M27 diminished LR1 duplex binding activity (Fig. 5B). Essentially complete inhibition of binding was observed in reactions containing 200 nM recombinant hnRNP D-M07 or hnRNP D-M27 (approximately 10-fold molar excess relative to purified LR1). Control experiments (not shown) established that these recombinant protein preparations were all able to bind to the mammalian telomeric repeat sequence d(TTAGGG)$_n$ (16), ruling out trivial explanations for inhibition of binding by hnRNP D-M07 or M27. The observed inhibition of LR1 duplex binding activity is specific for hnRNP D isoforms containing the tyrosine-rich 49-amino acid domain encoded by exon 7.

The region encoded by hnRNP D exon 7 therefore exerts a dominant negative effect on LR1 duplex DNA binding activity. These results further establish that the 45-kDa subunit of LR1 is the specific isoform of hnRNP D, which contains exon 2 but not exon 7 sequences. They also raise the possibility that LR1 DNA binding activity in vivo may be controlled not only by cell type-specific modification of the polypeptide components of LR1 but also by regulation of the relative levels of the hnRNP D isoforms.

---

Fig. 4. The isoform of hnRNP D that contacts the DNA in the LR1 heterodimer contains sequences encoded by alternative exon 2. A, immunoblots of recombinant His$_6$-tagged hnRNP D-M20, hnRNP D-M27, and hnRNP D-M27 visualized with antibodies raised against synthetic peptides bearing sequences from hnRNP D alternative exon 2 (αExon2) or exon 7 (αExon7). Mobilities of molecular mass (MW) markers are indicated at left. B, gel mobility shift analysis of LR1 binding reactions incubated with anti-hnRNP D or anti-nucleolin antibodies. Tested were pre-immune sera (Pre) or antibodies raised against synthetic peptides carrying hnRNP D exon 2 (αExon2) or exon 7 (αExon7). Arrows indicate the positions of free duplex DNA, the LR1-DNA complex, and the hnRNP D-DNA complex (subshift).

from alternative splicing (16, 17). One class of cDNA contains exon 2, another contains exon 7, and the third class contains both alternative exons. These three different cDNA classes correspond to three alternatively processed mRNAs that could encode distinct protein isoforms (Fig. 1). To determine which isoform or isoforms of hnRNP D is present in LR1, we raised antibodies against synthetic peptides representing sequences from exon 2 and exon 7 of hnRNP D. We verified the specificity of these antibodies by immunoblotting all three isoforms of recombinant hnRNP D, hnRNP D-M20, hnRNP D-M27, and hnRNP D-M07. As shown in Fig. 4A, the anti-exon 2 antibodies recognize isoforms hnRNP D-M20 and hnRNP D-M27, which contain sequences encoded by exon 2, but they do not recognize hnRNP D-M07. Furthermore, the anti-exon 7 antibodies recognize isoforms hnRNP D-M20 and hnRNP D-M27, which contain sequences encoded by exon 7, but they do not recognize hnRNP D-M20. The antibodies are therefore specific.

The exon-specific antibodies were then tested for recognition of LR1 duplex DNA binding activity. Anti-exon 2 pre-immune serum did not affect the bound complex nor did the immune serum react with DNA in the absence of protein, but the anti-exon 2 antibodies completely inhibited LR1 DNA binding (Fig. 4B). The inhibition of binding observed with anti-exon 2 antibodies contrasted with the supershift evident when LR1 was treated with anti-hnRNP D C-terminal antibodies, suggesting that sequences encoded by exon 2 may be in or near the duplex DNA binding domain.
Following treatment with anti-nucleolin antibodies shows that the presence of nucleolin in the LR1 heterodimer is not essential for stable interaction with DNA. This raises the possibility that nucleolin may function as a chaperone to load hnRNP D onto the duplex binding site. We do not yet know whether hnRNP D can interact with duplex DNA on its own or requires nucleolin as a chaperone. Recombinant hnRNP D does not bind duplex DNA, nor could DNA binding activity be reconstituted by combining recombinant hnRNP D and recombinant nucleolin (Fig. 5A). However, the recombinant proteins we have assayed were expressed in bacteria and were therefore not posttranslationally modified. In particular, phosphorylation appears to be necessary for LR1 duplex DNA binding activity (1), and recombinant protein expressed in E. coli would lack this modification. Moreover, bacterial expression of full-length nucleolin has thus far proved impossible, and the longest nucleolin derivative assayed in reconstitution experiments contained only the four RBDs and the RGG-rich C terminus (residues 284–709) and lacked a considerable region of N-terminal sequence that may be important for DNA binding.

Nucleolin and hnRNP D are ubiquitous polypeptides, but LR1 duplex DNA binding activity is restricted to pre-B and B cells. About 1 μg of LR1 is recovered from about 10⁹ PD31 cells, suggesting that the abundance of this factor is comparable with that of many transcription factors, and LR1 DNA binding activity therefore represents only a small fraction of the total hnRNP D and nucleolin in a mammalian nucleus. These observations suggest that there must be cell type-specific mechanisms that regulate LR1 duplex DNA binding activity.

Our data identify at least two distinct mechanisms that may regulate cell type specificity of LR1 duplex binding activity: post translational modification and alternative processing of the hnRNP D transcript. Consistent with the observation that LR1 duplex DNA binding activity is dependent upon phosphorylation (1), activity could not be reconstituted by bacterially expressed recombinant hnRNP D and nucleolin. Consistent with the importance of alternative processing of the hnRNP D transcript, the isoform of hnRNP D in the LR1 heterodimer appears to contain sequences encoded by alternative exon 2 but not by alternative exon 7. However, alternative splicing alone does not appear to account for cell type specificity of binding activity. Isoforms containing sequences encoded by exon 7 (M07 and M27) exert a dominant negative effect on LR1 duplex DNA binding activity, so alternative processing may regulate DNA binding activity not only by determining the amount of exon 2-containing isoforms that are present but also by regulating the relative levels of the different hnRNP D isoforms. The dominant negative effect of the isoforms of hnRNP D that contain sequences encoded by exon 7 may also reflect tighter interactions of these isoforms with nucleolin.

LR1 has a very unusual composition for a sequence-specific duplex DNA-binding protein, as both nucleolin and hnRNP D contain RBDs and RGG motifs typically found in proteins that interact with RNA, not duplex DNA. We note that these domains in LR1 do provide a structural basis for possible function in recombination. Crystallographic analysis has shown that an RBD forms a platform upon which a nucleic acid can be bound in an open structure (25, 26). The RBD platforms in LR1 may therefore display a region of DNA so that it is available to interact with other nucleic acids or with enzymes involved in recombination. The RBD and RGG motifs also suggest that LR1 might interact with other nucleic acid substrates in addition to duplex DNA, and future experiments will explore this possibility.

Acknowledgments—We are grateful to our colleagues and friends for many interesting and useful discussions, particularly to W. P. Russ for his help in displaying the nucleic acid structural model.
REFERENCES