Identification and Characterization of Galectin-9, a Novel β-Galactoside-binding Mammalian Lectin*

Jun Wada and Yashpal S. Kanwar‡
From the Department of Pathology, Northwestern University Medical School, Chicago, Illinois 60611

A 36-kDa β-galactoside mammalian lectin protein, designated as galectin-9, was isolated from mouse embryonic kidney by using a degenerate primer polymerase chain reaction and cloning strategy. Its deduced amino acid sequence had the characteristic conserved sequence motif of galectins. Endogenous galectin-9, extracted from liver and thymus, as well as recombinant galectin-9 exhibited specific binding activity for the lactosyl group. It had two distinct N- and C-terminal carbohydrate-binding domains connected by a link peptide, with no homology to any other protein. Galectin-9 had an alternate splicing isoform, exclusively expressed in the small intestine with a 31-amino acid insertion between the N-terminal domain and link peptide. Sequence homology analysis revealed that the C-terminal carbohydrate-binding domain of mouse galectin-9 had extensive similarity to that of monomeric rat galectin-5. The presence of galectin-5 in the mouse could not be demonstrated by polymerase chain reaction or by Northern or Southern blot genomic DNA analyses. Sequence comparison of rat galectin-5 and rat galectin-9 cDNA did not reveal identical nucleotide sequences in the overlapping regions in both species. Southern blot genomic DNA analyses, using the galectin-9 specific probe derived from the N-terminal carbohydrate-binding domain, indicated the presence of a novel gene encoding galectin-9 in both mice and rats. In contrast to galectin-5, which is mainly expressed in erythrocytes, galectin-9 was found to be widely distributed, i.e. in liver, small intestine, thymus > kidney, spleen, lung, cardiac and skeletal muscle > reticulocyte, brain. Collectively, these data indicate that galectin-9 is a new member of the galectin gene family and has a unique intestinal isoform.

There is growing evidence that specific carbohydrate moieties and their putative binding proteins, i.e. lectins, play diverse roles in mammalian physiology and development and in various pathological states (1). The mammalian lectins are classified into four categories, C-type lectins (including selectins), P-type lectins, pentraxins, and galectins; the latter are referred to as S-type or S-Lac lectins (2, 3). Galectins are endowed with two essential biochemical properties: 1) characteristic amino acid homologous sequences; and 2) affinity for β-galactoside sugars, i.e. carbohydrate-binding domain. In addition, all the known galectins lack a signal peptide, have a cytoplasmic localization, and are secreted as soluble proteins by a nonclassical secretory pathway (4). Seven mammalian galectins, i.e. galectins-1 (5), -2 (6), -3 (7), -4 (8), -5 (9), -7 (10, 11), and -8 (12), have been cloned and characterized. Structural analyses of various galectins indicate the presence of homodimers of carbohydrate-binding domains in galectin-1 and galectin-2, a monomer of the carbohydrate-binding domain in galectin-5, and a single polypeptide chain with two carbohydrate-binding domains joined by a link peptide in galectin-4 and galectin-8. Galectin-3 has a carbohydrate-binding domain, a short N-terminal segment, consisting of PGAYPG(1)_1 repeats, and an intervening stretch of amino acids, enriched with proline, glycine, and tyrosine (2, 3). Although the overall structure of galectins varies, each carbohydrate-binding domain is highly conserved and is encoded by three major exons (13, 14). Expression analyses have revealed that certain galectins display a restricted distribution, e.g. galectin-2 in hepatoma, galectin-4 in the small intestine, galectin-5 in erythrocytes, and galectin-7 in keratinocytes. Galectins with a broad tissue distribution include galectin-1, expressed in cardiac, smooth, and skeletal muscles, neurons, thymus, kidney, and placenta; galectin-3, present in blood cells, intestine, kidney, and neurons; and galectin-8, expressed in liver, kidney, cardiac muscle, lung, and brain (2, 3, 5-12). Some of these are believed to be involved in cell-cell or cell-matrix interactions (2).

Such interactions between cell adhesion molecules and extracellular matrix glycoproteins, i.e. collagen, laminin, fibronectin, and proteoglycans, and their receptors (integrins) are highly relevant during embryonic development, including organogenesis of the kidney (15, 16). Conceivably, these interactions also involve the participation of other matrix- and plasmamembra-bound macromolecules, such as galectins-1, -3, and -8, that are abundantly expressed in the developing metanephric kidney. In view of these considerations, studies were initiated to search for other developmentally regulated novel galectins that may be pertinent to the biology of cell-cell and cell-matrix interactions. In this communication, we report the identification and characterization of a new galectin, i.e. galectin-9, including its isoform and relationship with galectin-5.

EXPERIMENTAL PROCEDURES

Degenerate Oligonucleotide-based Polymerase Chain Reaction (PCR) and Cloning and Nucleotide Sequencing of PCR Products Isolated from

‡ To whom requests for reprints and correspondence should be addressed: Dept. of Pathology, Northwestern University Medical School, 303 E. Chicago Ave., Chicago, IL 60611. Tel.: 312-503-0084; Fax: 312-503-0627.

* Supported by National Institutes of Health Grant DK28492. 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™/EBI Data Bank with accession number(s) U55060 (mouse galectin-9), U55061 (mouse galectin-9 intestinal isoform), U59462 (rat galectin-9), and U72741 (rat galectin-9 intestinal isoform).

† To whom requests for reprints and correspondence should be addressed: Dept. of Pathology, Northwestern University Medical School, 303 E. Chicago Ave., Chicago, IL 60611. Tel.: 312-503-0084; Fax: 312-503-0627.

The abbreviations used are: PCR, polymerase chain reaction; RACE, rapid amplification of 5′-cDNA ends; bp, base pair(s); kb, kilobase pair(s); RT, reverse transcriptase; G9, galectin-9; MEPBS buffer, phosphate-buffered saline with 4 mM β-mercaptoethanol, 2 mM
Mouse Embryonic Kidney—Total RNA was extracted from embryonic metanephroi of CD-1 mice at day 13 and day 17 of gestation by the guanidine isothiocyanate-CsCl centrifugation method (17), and poly(A)⁺ RNA was isolated by utilizing a FastTrack 2.0 kit (Invitrogen). First strand cDNA was synthesized by using Moloney murine leukemia virus reverse transcriptase (RT; RNase H−) (5 U/μl) and a primer (CLONTECH) (18, 19). Double-stranded cDNA was generated by using a mixture of Escherichia coli DNA polymerase I, RNase H, and E. coli DNA ligase (20, 21). Subsequently, an adaptor-ligated double-stranded cDNA was generated by a Marathon cDNA Amplification Kit (CLONTECH), which was used as template DNA for rapid amplification of cDNA ends (RACE) (19, 22). A degenerate antisense 18-mer oligonucleotide primer, corresponding to the β-galactoside-binding conserved sequence HNPFR, was designed with the following nucleotide sequence: 5'-(G/A/A/G/G/T/C/G/G-G/G/A/T/T/G/A/AA/G/A/G/TT/G-3'. 5'-RACE PCR was performed using template DNA in a 50-μl reaction mixture, containing 5 μl of 10× buffer (500 mM Tris-HCl [pH 9.2], 160 mM (NH₄)₂SO₄, 17.5 mM MgCl₂), 200 μM concentrations of each dNTP, 0.2 μM adaptor primer 1 (supplied in the kit), 1 μM degenerate antisense primer, and 1 μl of polymerase mix; the latter is a mixture of 14.3 μl of Taq/Power (Boehringer Mannheim) and 5.7 μl of TaqStart antibody (CLONTECH). The reaction was carried out for 35 thermal cycles, each consisting of 1.5 min at 95°C, 2 min at 37°C, 3 min at 68°C, and a final 10-min extension at 68°C. PCR products purified by agarose gel electrophoresis, 2 mM phenylmethanesulfonyl fluoride, and purified by dilutional secondary and tertiary screening. Phage plaques were sequenced by the dideoxy chain termination method using modified T7 DNA polymerase (Sequenase, Amersham). A library was prepared and screened as detailed previously (25, 26). In brief, total RNAs from various organs of adult mice and rats were used for the synthesis of cDNAs (26). For the synthesis of cDNA from fetal red blood cells, 13-day-old mouse embryos were bled into the culture medium. Pellets of fetal erythrocytes were prepared by centrifugation and utilized for isolation of mRNA by the Micro-Fast Track mRNA isolation kit (Invitrogen). In addition, total RNAs from spleen and reticulocytes of phenylhydrazine-induced anemic mice and rats (30) were isolated to prepare the cDNA. PCR analyses were performed on cDNAs of various organs by utilizing primers with the following nucleotide sequences: for mouse, 5'-GCAATGGTCCCTCACGGA TAG-TAG' (MG9AS) and 5'-GGTTCCCGAGAGCGGTCC-3' (MG9AS); and for rat, 5'-GCCGT GGTCTCTCAAGACAGC-3' (RG5E) and 5'-CCCTAG GCCGAGATCCCT-3' (RG5AS) (Fig. 3A). To further ensure the detection of expression of galectin(s) in mouse reticulocytes, RACE PCR was also performed using adaptor-ligated double-stranded cDNA and primer set adaptor primer 1 and MG9AS. The PCR products were gel purified, ligated into pCR®II cloning vector (Invitrogen), and sequenced as described above.

Expression of Galectin-9 by Northern Blot Analyses—Analyses were performed on total RNA, isolated from various adult mouse and rat tissues, as described previously (26). In brief, 30 μg of total RNA of each organ were geloxidated, subjected to 1% agarose gel electrophoresis, and transferred to a nylon membrane (Pall BioSupport). Under high stringency conditions, the membranes were hybridized with a [α-32P]dCTP-random-radiolabeled “mouse galectin-9 cDNA probe,” generated by PCR using MG9AS and MG9AS primers (see above, and see Figs. 3A and 4C). In addition, separate membrane filters were hybridized with a radiolabeled “mouse galectin-9-specific cDNA probe,” generated by PCR using MG9AS and GA/GA/GA/GA/GA/A/G/T/G-3', and antisense (5'-32P]-radiolabeled “mouse galectin-9-specific cDNA probe” under high stringency conditions.

Purification—Expression and Purification—Expression of recombinant protein with C-terminal e-myc-(His)₉ tag for mouse galectin-9 was carried out with pTrcHis2 vector (Invitrogen). As controls, mouse galectin-1 and -3 cDNAs, containing entire coding segments, were generated by RT-PCR using mouse newborn kidney cDNA. Primers were: 5'-GGAATTCATGGCAAGTGAGCAAGAGGAG-3' for galectin-1, 5'-GGGGAAGCTTTGTCTGCACGTGGTCCAG-3' for galectin-3, and 5'-GAATTCATTGAAGATGGGATCCTAG-3' (33). PCR products of galectins-1 and -3 were subcloned into pCR®II vector and sequenced. cDNA inserts of these 3 mouse galectins were reamplified by using the following primer sets: 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' and 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' for galectin-1, 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' and 5'-GAATTCATTGAAGATGGGATCCTAG-3' for galectin-3, and 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' and 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' for galectin-9. BamHI and HindIII restriction sites were also introduced, and underlined in each of the primer sequences. The generated PCR products were digested with BamHI and HindIII, gel purified, and ligated into pTrcHis2A plasmid to prepare galectin-1 (pTrcHis2G1), -3 (pTrcHis2G3), and -9 (pTrcHis2G9) constructs. The constructs were transfected into the TOP10 bacterial host (Invitrogen) and sequenced to ensure proper in-frame ligation and Taq polymerase fidelity. The bacteria were allowed to grow in 1 liter of LB media until an A₆₀₀nm of 0.6 was achieved. Expression of fusion proteins was monitored by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and further cultured for 5 h. To isolate recombinant proteins, bacterial pellets were prepared, and suspended in 100 μl of either MEBPES buffer (9, 12) or for galectin-3, 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' and 5'-GGGGGGGATCCATTGAGTTGCGTGGTC-3' (34). Purified recombinant proteins were suspended in a sample buffer (4% SDS, 150 mM Tris-HCl [pH 6.8], 20% glycerol, 0.1% bromphenol blue, 1% β-mercaptoethanol) and subjected to 12.5% SDS-polyacrylamide gel electrophoresis (PAGE) (36). Gel bands were visualized after staining with Coomassie Blue.

Immunoprecipitation—To determine the molecular size of endogenous mammalian galectin-9, immunoprecipitation experiments were performed. Polyclonal antibody was raised by immunizing rabbits with a synthetic peptide, KTQGNRPFAHPQPMQT. Its sequence matches between 148 and 162 amino acid residues of the link peptide of mouse galectin-9 (Fig. 1B). An additional lysine residue at the N-terminal was used for conjugation of the peptide to keyhole limpet hemocyanin. For immunoprecipitation experiments, newborn mouse liver and thymus were radiolabeled in vivo by an intraperitoneal injection of [35S]methionine (0.05 mCi/g body weight). After 24 h, the organs were harvested, EDTA, DTT, dithiothreitol; Tris-dithiothreitol buffer, 20 mM Tris (pH 7.4), 5 mM EDTA, 150 mM sodium chloride, 1 mM DTT, PAGE, polyacrylamide gel electrophoresis, 2 mM phenylmethylene sulfonyl fluoride, 1% Triton X-100; ELISA, enzyme-linked immunosorbent assay; PBS, phosphate-buffered saline.
homogenized in Tris-DTT buffer containing 10 mM benzamidine, 2 mM phenylmethanesulfonyl fluoride, 1% TritonX-100, 10 mM benzamidine, and 10 mM $\epsilon$-amino-$n$-caproic acid and sonicated. The homogenates were centrifuged at 10,000 $\times$ g for 30 min at 4°C, and the supernatants were applied to lactosyl-Sepharose columns and eluted with 0.2 M lactose in Tris-DTT buffer. Immunoprecipitation was performed by adding 10 $\mu$l of anti-mouse galectin-9 antibody to 0.5 $\mu$l of eluate, containing $\sim$0.5 $\times$ 10$^{-6}$ dpm. The eluate-antibody mixture was gently swirled in an orbital shaker for 18 h. Protein A-Sepharose 4B (Phar- macia) was added to the antibody-galectin-9 complex and mixed for 2 h, following which the pellets were prepared and washed four times with Tris-DTT buffer. The immunoprecipitated complexes were dissolved in a sample buffer and subjected to 12.5% SDS-PAGE under reducing conditions. The gels were fixed in 10% acetic acid, treated with 1 M salicylic acid, and dried, and autoradiograms were prepared. Preimmune serum was used as a control.

**Enzyme-linked Immunosorbent Assays (ELISA)—**To assess the specificity of the anti-galectin-9 antibody, ELISA assays were performed (37, 38). Wells of RIA/EIA titer plates (Costar) were coated with 50 $\mu$l of synthetic peptide solution (100 $\mu$g/ml) in 20 mM NaHCO$_3$, pH 9.0. The plates were allowed to dry overnight at 37°C. 100 $\mu$l of ice-cold methanol were added to each well the next day and allowed to evaporate for 2 h at 37°C. To reduce nonspecific binding of the antibody, 200 $\mu$l of bovine serum albumin solution (5 mg/ml), prepared in phosphate-buff ered saline (PBS), was added, and plates were left at 22°C for 1 h. The wells were washed twice with PBS, and 0.5 mg of antibody (IgG fraction) in 50 $\mu$l of bovine serum albumin-PBS (100 $\mu$g/ml) solution was added to the first well. Log dilutions of the antibody were made in

---

**FIG. 1. Deduced amino acid (aa)sequence and structural analyses of mouse galectin-9.** A, schematic drawing of mouse galectin-9 and its intestinal isoform. Two carbohydrate-binding domains are connected by a link peptide. The intestinal isoform insertion is located between the N-terminal domain and the link peptide. B, deduced amino acid sequence of mouse galectin-9 and its isoform. The intestinal isoform insertion sequence is boxed. The amino acid sequence of the synthetic peptide used for antibody production is double-underlined. C, hydrophobicity analyses show that galectin-9 lacks signal peptide and hydrophobic transmembrane segment.
### Galectin-9, a Novel β-Galactoside-binding Lectin

| hG3 | 1 MADNFSLHDA | LSGSGNPNPQ | GWPGAWGQNP | AGAGGYPGAS | YPGAYPGQPQ |
|-----|---------------|------------|------------|------------|------------|
| 51  | PGAYPGQQAPP   | GAYHGAPGAY | PGAPAPGCVY | GPPSGPGAYF | SSQGPSAPGA |
| 178 |               |            |            |            | MA         |
| 178 |               |            |            |            |            |
| 178 |               |            |            |            |            |

**Fig. 2.** Comparison of carbohydrate-binding domains of galectins. Amino acid sequences of human (h) galectin-1 (5), human galectin-2 (6), human galectin-3 (7), rat (r) galectin-3 (8), rat galectin-5 (9), human galectin-7 (10), and rat galectin-8 (12) are aligned by Pileup program (GCG sequence analysis package). Galectin-x; m, mouse. Residues with shared identity are boxed.
Galectin-9 Is a Novel Member of the Galectin Family—At the amino acid level, the C-terminal domain of galectin-9 shared an extensive homology (81.3%) with rat erythrocyte galectin-5 (9). Also at the nucleotide level, its 5'-untranslated segment, the 3' half of the coding region and the 3'-untranslated regions had a 80.3% homology with rat galectin-5 (Table I, Fig. 3A). To ensure that the cloned galectin-9 is not an isoform of galectin-5, a PCR cloning strategy was used. Primers were designed from 5'-and 3'-untranslated regions of galectin-9 cDNA (Fig. 3A, MG9SE and MG9AS). RT-PCR analyses of various mouse tissue mRNAs (heart, brain, liver, kidney, spleen, muscle, and thymus) yielded 1.0-kb products (Fig. 3B, panel a). In addition, a ~1.1-kbp product, exclusively expressed in the small intestine, was observed (Fig. 3B, panel a). By sequence analysis, the ~1.1-kbp product revealed a 31-amin acid insertion between the N-terminal domain and the link peptide (GenBank™ accession no. U55061) and was designated as a mouse galectin-9 intestinal isoform. Except for the 31-amino acid insertion, the intestinal isoform had a 100% homology with mouse galectin-9. Since rat galectin-9 was originally isolated from reticulocytes, RT-PCR was also carried out on mouse cDNAs, prepared from phenylhydrazine-induced reticulocytes and embryonic erythrocytes. Only ~1.0-kbp products were detected, and the unexpected 0.5-kbp product, i.e., mouse homologue of rat galectin-5, could not be amplified. RACE PCR was also performed using 47-bp adaptor-ligated mouse reticulocyte cDNA, and it also revealed only one ~1.1-kbp product; the 0.5-kbp band of galectin-5 was not detected.

Since RT-PCR and 5'-RACE PCR failed to document the presence of galectin-5 in mouse, we then selected the primer set RG5SE and RG5AS from rat galectin-5 cDNA (GenBank™ accession no. L36862), which corresponds to mouse MG9SE and MG9AS by a sequence alignment program (Fig. 3A). With these rat primers, RT-PCRs were performed on cDNAs from various adult rat tissues, including reticulocytes. Two bands, i.e., ~1.0 and ~0.5 kDa, were observed in several tissues, and
another ~1.1-kbp band was confined to the small intestine only (Fig. 3B, panel b). Analysis of the ~0.5-kbp product revealed that it had a nucleotide sequence identical with that of rat galectin-5 (30–472 bp, GenBank™ accession no. L36862), while the ~1.0-kbp product (GenBank™ accession no. U59462) had an 89.9% homology to mouse galectin-9. Thus, the ~1.0-kbp product was regarded as the rat homologue of mouse galectin-9. Sequence analysis of the ~1.1-kbp product (GenBank™ accession no. U72741) showed that it is an isoform of rat galectin-9 since it had a 32-amino acid insertion between the N-terminal carbohydrate domain and the link peptide, as in the mouse galectin-9 intestinal isoform. These two rat PCR products, the 0.5- and 1.0-kbp rat galectins-5 and -9 shared a 93.6% homology in the 443-bp overlapping region of C-terminal carbohydrate-binding domain (Fig. 3C), while the rat galectin-9 and its intestinal isoform had a 100% identity in the 974-bp overlapping region, spanning the N- and C-terminal carbohydrate-binding domains. These results indicate that the newly cloned galectin-9 is not an alternative splicing isoform of galectin-5, that it has a long intestinal isoform in both rats and mice, and that galectin-5 is not present in the mouse.

Expression of Galectins by Northern and Southern Blot Analyses—Using the galectin-9 (G9) cDNA probe (Fig. 4C), a ~2-kbp single transcript was observed in various mouse tissues (Fig. 4A). Smaller mRNA transcripts, corresponding to putative mouse galectin-5, were not detected. mRNA expression of galectin-9 in various mouse organs was as follows: liver, small intestine, thymus > kidney, spleen, lung, skeletal muscle, heart > reticulocyte, brain (Fig. 4A, upper panel). In rat tissues, the G9 cDNA probe hybridized with transcripts of ~2 and ~1.5 kb, corresponding to galectin-9 and galectin-5, respectively (Fig. 4B, upper panel). The 1.5-kbp mRNA transcripts of rat galectin-5 were abundantly expressed in reticulocytes and spleen. Although, mRNA expression of galectin-9 (~2 kb) in various rat tissues was similar to that in the mouse, its expression was relatively low in the thymus and substantially lower in kidney and skeletal muscle (Fig. 4B, upper panel). By using the G9-specific probe, only ~2-kbp transcripts were observed, and smaller transcripts (~1.5 kb) were not detected in various rat tissues (Fig. 4B, middle panel).

Southern blot analyses, using the G9 cDNA probe, revealed a single major band in various restriction enzyme digests of mouse genomic DNA (Fig. 5A, left panel). Identical results were obtained for mouse genomic DNA digests in blot hybridized with the G9-specific cDNA probe (Fig. 5A, right panel), supporting the presence of a gene encoding galectin-9 only. Rat genomic DNA digests revealed multiple bands when the G9-specific cDNA probe was used for Southern blot hybridization analyses (Fig. 5A, right panel); while hybridization with the G9 cDNA probe revealed a few additional bands in various restriction enzyme digests (Fig. 5B, left panel), suggesting the presence of galectin-9 and ~5 genes in rat genomic DNA.

Recombinant Galectin-9 Exhibits Lactose-binding Activity—Preliminary experiments revealed a time-dependent increasing isopropyl-1-thio-β-D-galactopyranoside-inducible expression of recombinant galectins, which was maximal at 5 h in E. coli lysates, prepared in the presence of DTT or β-mercaptoethanol. SDS-PAGE of the eluates of bacterial lysates from lactosyl-Sepharose column revealed ~39-kDa, ~37-kDa, and ~17-kDa bands of rat galectins-9, -3, and -1, respectively (Fig. 6A, lanes 3–8).

Immunoprecipitation and ELISA Assays—Homogenates of [35]Smethionine-labeled liver and thymus were applied to a lactosyl-Sepharose column, followed by immunoprecipitation of the eluted fractions with anti-galectin-9 antibody. SDS-PAGE autoradiograms of the immunoprecipitates revealed a ~36-
kDa band, indicating that endogenous galectin-9 also has lactose-binding activity and a comparable molecular weight (Fig. 5B, lanes 1 and 2). No discernible bands were observed when immunoprecipitation was performed with preimmune serum (Fig. 5B, lanes 3 and 4). Specificity of the antibody was also assessed by ELISA assay in which a fixed amount of antigen, i.e. synthetic peptide, and serial log dilutions of the antibody were used. With increasing dilutions of the antibody, a proportional decrease in A490 readings was observed (Fig. 7A). To confirm the specificity of the antibody, a competitive inhibition ELISA assay was performed. A fixed amount of diluted antibody (1:000) along with serial log dilutions of the competitive antigen, i.e. recombinant galectin-9, were added into the wells of the titer plate coated with the synthetic peptide. With increasing dilutions of the competitive antigen, a proportional increase in A490 readings was observed (Fig. 7B), documenting the specificity of anti-galectin-9 antibody.

**DISCUSSION**

In the present study, we have described a novel galectin, galectin-9, isolated from the embryonic kidney cDNA, using the degenerate oligonucleotide-based RACE-PCR cloning strategy. Although previous studies used traditional lactosyl-Sepharose column protein purification to identify galectins, the successful isolation of galectin-9 indicates that degenerate oligonucleotide-based RACE-PCR strategy is yet another useful method by which one can search for new galectins in various tissues. The biochemical characteristics of galectin-9 fulfill the criteria for its inclusion as a new member of the galectin family (2, 3). They include the following: 1) its deduced amino acid sequence which indicates two domains consisting of characteristic conserved sequence motifs that are implicated in binding to specific saccharides; and 2) the recombinant fusion protein (recombinant galectin-9) exhibits specific binding affinity for lactosyl groups. In addition, an endogenous protein, with an expected size of
Bluestaining. The molecular weights of the fusion proteins also include galectin-3 (37 kDa; lanes 7 and 8) and galectin-9 (39 kDa; lanes 1, 3, 5, and 7) are visualized after Coomassie Blue staining. The molecular weights of the fusion proteins also include ~3-kDa myc-(His)6, tag, B, homogenates of [35S]methionine-labeled thymus and liver were subjected to lactosyl-Sepharose chromatography, followed by immunoprecipitation of the eluates with anti-galectin-9 antibody and SDS-PAGE. A major ~36 kDa band is seen in both the liver and thymus in the SDS-PAGE autoradiogram. No band is seen in the homogenates treated with preimmune serum. Arrow, point of application of the sample.

~36 kDa, binds to lactosyl-Sepharose columns and can be immunoprecipitated by specific antibody, directed against a unique sequence of galectin-9 link peptide.

A feature common to galectin-9 and other galectins is that it lacks a classical signal sequence and a transmembrane hydrophobic segment. Thus, like other galectins that have a cytosolic distribution and are secreted as a soluble proteins by a non-classical secretory mechanism (4), galectin-9 could be covalent-ized with the aid of a carrier protein. Such a mechanism has been shown for other cytoplasmic proteins lacking a signal peptide, such as thymosin, interleukin-1, and fibroblast growth factor (39, 40).

Structural analyses of galectin-9 revealed some additional unique features. It consists of two distinct carbohydrate-binding domains connected by a link peptide. Among the known galectin family members, only galectin-4 (8), galectin-8 (12), and the Caenorhabditis elegans homologue (41) have similar dimeric structures. Other galectins, including galectin-1 (5) and galectin-2 (6), contain only one carbohydrate-binding domain but can function as homodimers to facilitate aggregation or agglutination of cell surface-bound glycoconjugates via non-covalent associations (3). Although galectins with two carbohydrate-binding domains may have similar agglutinating and aggregating potential, their properties could be influenced by the size and amino acid sequence of the link peptide affecting overall molecular conformation. Moreover, since the amino acid sequences of link peptides differ from one another, it raises the possibility that these may modify biological activities of various dimeric galectins in a given tissue. One of the unique features of galectin-9 is its alternate splicing isoform exclusively expressed in the small intestine. This isoform has a 31- and 32-amino acid insertion in mouse and rat, respectively. Since this insertion has no homology with the carbohydrate-binding domain sequences, it can be regarded as an extension of the link peptide. Such a long link peptide, with a stretch of 57–58 amino acids, may influence the macromolecular conformation of galectin-9 which may be necessary for certain yet to be defined functions related to the biology of intestinal epithelium. Such an extended version of the link peptide or the isoforms have not been reported in other dimeric galectins.

Analysis of the C-terminal domain of mouse galectin-9 revealed substantial sequence homology (81.3%) with rat galectin-5, while the N-terminal had a 23–45% amino acid sequence homology with other galectins. At the nucleotide level, the coding region of the C-terminal domain and the 5’- and 3’-untranslated regions had a 80.9% similarity to rat galectin-5 cDNA. Galectin-5 is a monomeric form of rat galectin, which is mainly expressed in the erythrocytes (9). Its genetic relationship with galectin-9 is not clear; i.e. are they alternate splicing isoforms, or are they derived from separate genes? Initially, we speculated that galectin-9 might be a novel isoform of galectin-5 because of its segmental sequence homology. Therefore, attempts were made to isolate putative mouse galectin-5 from various tissues by PCR cloning methods, using the primers derived from sequences of flanking the 5’- and 3’-untranslated regions of mouse galectin-9. The ~0.5 kbp PCR product, corresponding in size to galectin-5, could not be amplified in mouse cDNA, instead, we identified galectin-9 and an intestinal alternate splicing isoform of galectin-9. In contrast, in cDNAs of various rat tissues, three PCR products, corresponding to galectin-5, galectin-9, and the intestinal isoform of galectin-9, were amplified when primer sequences derived from the 5’- and 3’-untranslated regions of the reported rat galectin-5 cDNA were used. However, a comparison of the cDNA sequences of rat galectin-5 with those of galectin-9 still does not show identical nucleotide sequences in the overlapping regions; thus, galectin-9 is not an alternate splicing form of galectin-5. The existence of a long intestinal isoform in rat as well as in the mouse, with nucleotide sequences identical with that of galectin-9 in the respective species, further supports the notion that galectin-9 is indeed a novel member of the galectin gene family.
To further characterize the relationship between the galectin-5 and galectin-9, Northern blot and genomic Southern blot analyses were performed. By Northern blot analyses, galectin-9 mRNA transcripts were found in various tissues with a wide distribution in both rats and mice. However, galectin-5 mRNA transcripts were detected largely in rat reticulocytes and spleen. Galectin-5 mRNA transcripts were absent in mouse reticulocytes as well as in other mouse tissues. By using the galectin-9 (G9)-specific cDNA probe, genomic Southern blot analyses affirmed the existence of a unique gene encoding galectin-9 protein both in rat and mouse. In mouse genomic Southern blot analyses, a single band was detected in various restriction enzyme digests of the genomic DNA after hybridization with G9 specific or G9 cDNA probes; thus, the existence of the galectin-5 gene in the mouse is doubtful. Certainly, in our hands, attempts to elucidate the presence of mouse galectin-5 by RT-PCR, prime RACEPCR, Northern blot, or Southern blot genomic DNA analyses were unsuccessful.

By a recent homology search through the GenBank™, we have also found that Homo sapiens RCC313 mRNA (Z49107) (42) has a 70% homology with mouse galectin-9 and thus regard it as the human homologue of galectin-9. This cDNA was isolated by immunoscreening of a human Hodgkin’s lymphoma AZAP library with autologous patient serum and was found to be distributed in lymphoid tissues (42). Since, in both mice and rats, galectin-9 is expressed in lymphoid tissues, such as in the thymus and spleen, its potential role in the biology of the immune system would be anticipated, although this galectin is originally isolated from embryonic kidney. In addition, a 17.5-kDa galectin with a high amino acid sequence similarity to galectin-5 has been reported in the rat kidney (43). This galectin may be galectin-5 or a proteolytic fragment of galectin-9.

In summary, although the C-terminal carbohydrate domain of galectin-9 shares extensive homology with galectin-5, there are distinct differences between the two, such as different tissue distributions and carbohydrate-binding domains. Finally, like certain other galectins (2), galectin-9 seems to be developmentally regulated in various embryonic tissues, including the kidney, the organogenesis of which is heavily influenced by cell-cell and cell-matrix interactions (15, 16). Since galectins are believed to be involved in cell-matrix interactions (44–49) and are developmentally regulated, it would be of great interest to investigate their relevance in various embryological processes regulating cell growth and differentiation (32, 50).

Acknowledgments—We are grateful to Dr. Samuel H. Barondes and his colleagues for valuable suggestions during the course of this investigation and for suggesting the inclusion of galectin-9 as a new member of the galectin gene family. We are also thankful to Drs. D. G. ScarpeUi, E. I. Wallner, and K. Alveraz for carefully editing the manuscript.

REFERENCES

1. Varki, A. (1993) Glycobiology 3, 97–130
2. Barondes, S. H., Cavenago, V., Cooper, D. N. W., Cummings, R. D., Dickamer, H., Li, M., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K., Wallner, E. I., and Alveraz, K. Alvaraz for carefully editing the manuscript.

To further characterize the relationship between the galectin-5 and galectin-9, Northern blot and genomic Southern blot analyses were performed. By Northern blot analyses, galectin-9 mRNA transcripts were found in various tissues with a wide distribution in both rats and mice. However, galectin-5 mRNA transcripts were detected largely in rat reticulocytes and spleen. Galectin-5 mRNA transcripts were absent in mouse reticulocytes as well as in other mouse tissues. By using the galectin-9 (G9)-specific cDNA probe, genomic Southern blot analyses affirmed the existence of a unique gene encoding galectin-9 protein both in rat and mouse. In mouse genomic Southern blot analyses, a single band was detected in various restriction enzyme digests of the genomic DNA after hybridization with G9 specific or G9 cDNA probes; thus, the existence of the galectin-5 gene in the mouse is doubtful. Certainly, in our hands, attempts to elucidate the presence of mouse galectin-5 by RT-PCR, PRIME RACEPCR, Northern blot, or Southern blot genomic DNA analyses were unsuccessful.

By a recent homology search through the GenBank™, we have also found that Homo sapiens RCC313 mRNA (Z49107) (42) has a 70% homology with mouse galectin-9 and thus regard it as the human homologue of galectin-9. This cDNA was isolated by immunoscreening of a human Hodgkin’s lymphoma AZAP library with autologous patient serum and was found to be distributed in lymphoid tissues (42). Since, in both mice and rats, galectin-9 is expressed in lymphoid tissues, such as in the thymus and spleen, its potential role in the biology of the immune system would be anticipated, although this galectin is originally isolated from embryonic kidney. In addition, a 17.5-kDa galectin with a high amino acid sequence similarity to galectin-5 has been reported in the rat kidney (43). This galectin may be galectin-5 or a proteolytic fragment of galectin-9.

In summary, although the C-terminal carbohydrate domain of galectin-9 shares extensive homology with galectin-5, there are distinct differences between the two, such as different tissue distributions and carbohydrate-binding domains. Finally, like certain other galectins (2), galectin-9 seems to be developmentally regulated in various embryonic tissues, including the kidney, the organogenesis of which is heavily influenced by cell-cell and cell-matrix interactions (15, 16). Since galectins are believed to be involved in cell-matrix interactions (44–49) and are developmentally regulated, it would be of great interest to investigate their relevance in various embryological processes regulating cell growth and differentiation (32, 50).

Acknowledgments—We are grateful to Dr. Samuel H. Barondes and his colleagues for valuable suggestions during the course of this investigation and for suggesting the inclusion of galectin-9 as a new member of the galectin gene family. We are also thankful to Drs. D. G. ScarpeUi, E. I. Wallner, and K. Alveraz for carefully editing the manuscript.

REFERENCES

1. Varki, A. (1993) Glycobiology 3, 97–130
2. Barondes, S. H., Castronovo, V., Cooper, D. N. W., Cummings, R. D., Dickamer, H., Li, M., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K., Wallner, E. I., and Alveraz, K. Alvaraz for carefully editing the manuscript.

To further characterize the relationship between the galectin-5 and galectin-9, Northern blot and genomic Southern blot analyses were performed. By Northern blot analyses, galectin-9 mRNA transcripts were found in various tissues with a wide distribution in both rats and mice. However, galectin-5 mRNA transcripts were detected largely in rat reticulocytes and spleen. Galectin-5 mRNA transcripts were absent in mouse reticulocytes as well as in other mouse tissues. By using the galectin-9 (G9)-specific cDNA probe, genomic Southern blot analyses affirmed the existence of a unique gene encoding galectin-9 protein both in rat and mouse. In mouse genomic Southern blot analyses, a single band was detected in various restriction enzyme digests of the genomic DNA after hybridization with G9 specific or G9 cDNA probes; thus, the existence of the galectin-5 gene in the mouse is doubtful. Certainly, in our hands, attempts to elucidate the presence of mouse galectin-5 by RT-PCR, prime RACEPCR, Northern blot, or Southern blot genomic DNA analyses were unsuccessful.

By a recent homology search through the GenBank™, we have also found that Homo sapiens RCC313 mRNA (Z49107) (42) has a 70% homology with mouse galectin-9 and thus regard it as the human homologue of galectin-9. This cDNA was isolated by immunoscreening of a human Hodgkin’s lymphoma AZAP library with autologous patient serum and was found to be distributed in lymphoid tissues (42). Since, in both mice and rats, galectin-9 is expressed in lymphoid tissues, such as in the thymus and spleen, its potential role in the biology of the immune system would be anticipated, although this galectin is originally isolated from embryonic kidney. In addition, a 17.5-kDa galectin with a high amino acid sequence similarity to galectin-5 has been reported in the rat kidney (43). This galectin may be galectin-5 or a proteolytic fragment of galectin-9.

In summary, although the C-terminal carbohydrate domain of galectin-9 shares extensive homology with galectin-5, there are distinct differences between the two, such as different tissue distributions and carbohydrate-binding domains. Finally, like certain other galectins (2), galectin-9 seems to be developmentally regulated in various embryonic tissues, including the kidney, the organogenesis of which is heavily influenced by cell-cell and cell-matrix interactions (15, 16). Since galectins are believed to be involved in cell-matrix interactions (44–49) and are developmentally regulated, it would be of great interest to investigate their relevance in various embryological processes regulating cell growth and differentiation (32, 50).

Acknowledgments—We are grateful to Dr. Samuel H. Barondes and his colleagues for valuable suggestions during the course of this investigation and for suggesting the inclusion of galectin-9 as a new member of the galectin gene family. We are also thankful to Drs. D. G. ScarpeUi, E. I. Wallner, and K. Alveraz for carefully editing the manuscript.

REFERENCES

1. Varki, A. (1993) Glycobiology 3, 97–130
2. Barondes, S. H., Castronovo, V., Cooper, D. N. W., Cummings, R. D., Dickamer, H., Li, M., Gitt, M. A., Hirabayashi, J., Hughes, C., Kasai, K., Wallner, E. I., and Alveraz, K. Alvaraz for carefully editing the manuscript.