The Primary Structure and Carbohydrate Specificity of a β-Galactosyl-binding Lectin from Toad (Bufo arenarum Hensel) Ovary Reveal Closer Similarities to the Mammalian Galectin-1 than to the Galectin from the Clawed Frog Xenopus laevis*

(Received for publication, June 18, 1996, and in revised form, September 10, 1996)

Hafiz Ahmed‡, Jan Pohl*, Nilda E. Fink‡¶, Fred Strobel, and Gerardo R. Vasta**

From the ‡Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, Maryland 21202 and ¶Microchemical Facility, Winship Cancer Center and Department of Chemistry, Emory University, Atlanta, Georgia 30322

The detailed characterization of a galectin from the toad (Bufo arenarum Hensel) ovary in its primary structure, carbohydrate specificity, and overall biochemical properties has provided novel information pertaining to structural and evolutionary aspects of the galectin family. The lectin consists of identical single-chain polypeptide subunits composed of 134 amino acids (calculated mass, 14,797 daltons), and its N-terminal residue, alanine, is N-acetylated. When compared to the sequences of known galectins, the B. arenarum galectin exhibited the highest identity (48% for the whole molecule and 77% for the carbohydrate recognition domain (CRD)) with the bovine spleen galectin-1, but surprisingly less identity (38% for the whole molecule and 47% for the CRD) with a galectin from Xenopus laevis skin (Marschal, P., Herrmann, J., Leffler, H., Barondes, S. H., and Cooper, D. N. W. (1992) J. Biol. Chem. 267, 12942–12949). Unlike the X. laevis galectin, the binding activity of the B. arenarum galectin for N-acetyllactosamine, the human blood group A tetrasaccharide and Galβ1,3GalNAc relative to lactose, was in agreement with that observed for the galectin-1 subgroup and those galectins having “conserved” (type I) CRDs (Ahmed, H., and Vasta, G. R. (1994) Glycobiology 4, 545–549). Moreover, the toad galectin shares three of the six cysteine residues that are conserved in all mammalian galectins-1, but not in the galectins from X. laevis, fish, and invertebrates described so far. Based on the homologies of the B. arenarum galectin with the bovine spleen galectin-1 and X. laevis skin galectin, it should be concluded that within the galectin family the correlation between conservation of primary structure and phylogenetic distances among the source species may not be a direct one as proposed elsewhere (Hirabayashi, J., and Kasai, K. (1993) Glycobiology 3, 297–304). Furthermore, galectins with conserved (type I) CRDs, represented by the B. arenarum ovary galectin, and those with “variable” (type II) CRDs, represented by the X. laevis 16-kDa galectin, clearly constitute distinct subgroups in the extant amphibian taxa and may have diverged early in the evolution of chordate lineages.

Galectins (Barondes et al., 1994a), comprise S-type β-galactosyl binding lectins, present in both homeotherm and poikilotherm vertebrates and invertebrates, that require a reducing environment but do not require divalent cations for their binding activity (Hirabayashi and Kasai, 1993; Barondes et al., 1994b). The primary structures of a considerable number of galectins are currently available (Hirabayashi and Kasai, 1993; Gitt et al., 1995; Hadari et al., 1995; Madsen et al., 1995; Hirabayashi et al., 1996). Furthermore, the three-dimensional structures of a limited number of galectins (galectins-1 (Bourne et al., 1994; Liao et al., 1994) and galectin-2 (Lobsanov et al., 1993)) have been reported recently, providing definitive information about their folding patterns, the amino acid residues that interact with the ligand and determine the architecture of the binding site, and the nature of the interactions that are established. Among the amino acid residues that are substantially conserved among various galectins, His41, Asn46, Arg48, Asn61, Trp68, Glu71, and Arg73 (residue numbers are those of bovine spleen (Liao et al., 1994)) are recognized as critical for sugar binding (Lobsanov et al., 1993; Bourne et al., 1994; Liao et al., 1994) whereas Ser29, Phe30, Asn33, His44, Asn46, Arg48, Asn61, and Arg111 interact with each other to provide the architecture of the CRD (Liao et al., 1994).

Based on selected features such as their primary structure and subunit architecture, the galectin family has been subdivided in several subgroups (“proto,” “tandem,” and “chimera” (Hirabayashi and Kasai 1993); galectin-1–8 (Barondes et al., 1994a)). Although all members of the galectin family bind lactose/N-acetyllactosamine, a limited diversity exists in the carbohydrate specificity (Leffler and Barondes, 1986; Abbott et al., 1988; Ahmed et al., 1990; Marschal et al., 1992; Oda et al., 1993; Ahmed and Vasta, 1994). Based on the differences in

*This work was supported by Lucille P. Markey Trust Fund Grant 95-31 and National Science Foundation Grant MCB-94-06649 (to G. R. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby advertised in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡Present address: Cátedra de Hematología, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calles 1 y 115, Argentina.

¶To whom correspondence should be addressed: Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E. Pratt St., Baltimore, MD 21202. Tel.: 410-234-8826; Fax: 410-234-8896; E-mail: vasta@umbi.umd.edu.

1 The abbreviations used are: CRD, carbohydrate recognition domain; ASF, asialofetuin; HRP, horse radish peroxidase; ABTS, diaminonarosa 2,2’-azinobis(3-ethylbenzothiazoline-6-sulfonate); PBS, phosphate-buffered saline; ME, 2-mercaptoethanol; PBSI:1.0, PBS diluted 10-fold with water; ac, acetyl; CAM, carboxamidomethyl; CAM-M-G, reduced and carboxamidomethylated galectin performed on a solid phase (mild conditions); CAM-G, reduced and carboxamidomethylated galectin performed in solution (drastic conditions); CM, carboxymethyl; ESI, electrospray ionization mass spectrometry; MALDI, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; PTC, phenylthiocarbamyl; PTH, phenylthiohydantoin; RP, reversed phase; HPLC, high performance liquid chromatography; BSA, bovine serum albumin; CHO, Chinese hamster ovary.
specificity and the conservation of amino acid residues that interact with the carbohydrate ligands, we classified galectins into two types: “conserved” (type I) and “variable” (type II) (Ahmed and Vasta, 1994). This classification may reflect not only common features of their carbohydrate specificity, but possibly evolutionary aspects of their recognition functions. Most of the lectins under galectin-1 group have “conserved” CRDs and exhibit a very similar carbohydrate specificity (Leffler and Barondes, 1986; Abbott et al., 1988; Ahmed et al., 1990, 1994). The CRDs of lectins grouped as galectin-2, -3, -4, and carbohydrate specificity of a galectin from B. arenarum (Marschal et al., 1992; Oda et al., 1991; and Xenopus laevis (Marschal et al., 1992), subclass Urodela: Ambystoma mexicanum (Allen et al., 1992), a number of distinct lectin activities have been identified in various adult tissues, including skin, muscle, and gonad. Among these, members of the galectin family have been isolated and characterized (Fink et al., 1987; Shet and Madaiah, 1989; Rana catesbeiana (Ozeki et al., 1991), and Xenopus laevis (Marschal et al., 1992). However, a detailed biochemical and structural characterization, including the complete primary structure and carbohydrate specificity, has been accomplished only for the X. laevis galectin (Marschal et al., 1992). Yet the position of this lectin remains to be settled in the most recent classification (Barondes et al., 1994a). Because of subunit size and architecture, carbohydrate specificity, and primary structure of the CRD, this galectin has been classified as a “proto” type member (Hirabayashi and Kasai, 1993) carrying a “variable” (type II) CRD (Ahmed and Vasta, 1994).

In contrast with X. laevis, where galectins are mainly confined to adult skin (Marschal et al., 1992) and have not been detected in embryos (Marschal et al., 1994), in the toad B. arenarum galectins are expressed in oocytes and further post-fertilization stages such as single cell, two-cell, blastula, gastrula, and somite development (Elola et al., 1987). Differences in the location and developmental expression of galectins in the two amphibian species raise the possibility that, although present in phylogenetically closely related taxa, they might mediate mechanisms associated with substantially different biological functions, such as host defense in X. laevis (Marschal et al., 1992) or developmental processes in B. arenarum (Elola et al., 1987). In mouse embryogenesis, galectin-1 and -3 are detected in the trophectoderm of the blastocyst as early as just before its implantation in the uterine wall, and they are believed to play a role in attachment by binding lacto-N-fucopentaose I (Colnot et al., 1996). Following gastrulation, galectin-1 is first expressed in somite myotomes, whereas galectin-3 is confined to the notochord. However, unlike the mammalian galectin-1 or galectin-3, the activity of the B. arenarum galectin can be detected in every stage prior to blastula, suggesting further differences in its putative function(s) proposed for mammalian galectins, such as embryo implantation and development. Therefore, the toad B. arenarum may constitute a suitable model for the elucidation of the biological function(s) of galectins in the embryogenesis of poikilotherm vertebrates.

As a first step of a systematic initiative aimed at the investigation of galectin structure-function relationships, we report herein the biochemical characterization, primary structure, and carbohydrate specificity of a galectin from B. arenarum ovary and examine its primary structure similarities with those reported elsewhere from homeothermic and poikilotherm vertebrates. Surprisingly, our results suggest that the toad ovary galectin is closer, both in primary structure and carbohydrate specificity, to the bovine galectin-1 than to the galectin from a related amphibian species, X. laevis, indicating that, within this lectin family, structural and functional divergence may have occurred early in vertebrate evolution.

EXPERIMENTAL PROCEDURES

Reagents—The protein assay reagent was from Bio-Rad. Ammonium PAGPlate and gel permeation chromatography molecular weight standards and ribonuclease A were purchased from Pharmacia Biotech Inc. The peroxidase substrate diammonium 2,2′-azinobis(3-ethylbenzothiazoline-6-sulfonate) (ABTS) was from Kirkegaard & Perry Laboratories. Sequencing grade reagents and solvents for protein sequencing, amino acid analysis, and HPLC were from Applied Biosystems (Division of the Perkin-Elmer Corp.). Sequencing grade endoprotease Asp-N, endoprotease Glu-C, and tryspin were from Boehringer-Manheim, and lysyl endopeptidase was from Wako Bioproducts. Horseradish peroxidase (HRP) was from Sigma. All other reagents were of the highest grade commercially available.

Toad Ovaries—Toad (B. arenarum Hense) ovaries weighing approximately 50 g each were dissected from human chorionic gonadotrophin-stimulated females as reported elsewhere (Fink et al., 1987) and stored at ~80 °C.

Purification of Toad Ovary Galectin—The galectin was purified through an improved protocol as reported elsewhere (Ahmed et al., 1996). Briefly, toad ovaries (500 g) were homogenized with cold (4 °C) phosphate-buffered saline (diluted 1:10), 0.01 M 2-mercaptoethanol, 0.1 M lactose (PBS(1:10)/ME/lactose) containing 0.1 mM phenylmethylsulfonyl fluoride (2 ml buffer/g wet tissue). After centrifugation, the clear supernatant was mixed with DEAE-Sepharose (10 ml of supernatant/1 ml resin) prequiequilibrated with PBS(1:10)/ME. After gentle mixing for 1 h at 4 °C, the slurry was transferred to a fritted glass funnel, and the resin was washed with 10 bed volumes of cold PBS(1:10)/ME to remove lactose and unbound protein. The bound proteins were eluted with 500 ml of PBS/M/ME, 0.002 M EDTA, 0.5 M NaCl, and the eluate was adsorbed on a column of lactosyl-Sepharose prequiequilibrated with PBS/ME, 0.002 M EDTA, 0.5 M NaCl. The column was thoroughly washed with equilibrating buffer followed by 5 bed volumes of PBS(1:10)/ME and the bound protein was eluted with 0.1 M lactose in PBS(1:10)/ME. The fractions containing protein were pooled, and aliquots were absorbed on DEAE-Sepharose columns (4 ml bed volume), overlaid with 50% glycerol in eluting buffer, and stored at ~20 °C until use.

Agglutination Tests—Agglutination tests were carried out in BSA-coated 96-well Terasaki plates (Robbins Scientific, Mountain View, CA) as reported earlier (Vasta et al., 1986). An equal volume of a 0.5% suspension of guartalardehyde-fixed protein-treated rabbit red blood cells was added to 5 μl of 2-fold dilutions of galectin in PBS/ME (pH 7.5). The plates were gently vortex-mixed and incubated at room temperature for 1 h. Agglutination was read under a microscope and scored from 0 (negative) to 4. The reciprocal of the highest dilution of galectin showing an agglutination score of 1 was recorded as the titer. The specific activity of the lectin was defined as the titer/mg of protein/ml. Controls were carried out by adding PBS/ME instead of lectin.

Protein Determination—Protein concentrations were determined on 96-well flat bottom plates with the Bio-Rad protein assay following a modification of the manufacturer’s protocols, using BSA as a standard. To 100 μl of 5–40 μg/ml protein standard solutions and samples, 100 μl of Coosamissie Blue dye reagent prediluted 2.5-fold with water were added. After 5 min, the reactions were read in a Molecular Devices Plate Reader at 595 nm, and the data were analyzed through the Softmax program.

Polyacrylamide Gel Electrophoresis—Analytical polyacrylamide slab gel electrophoresis in the presence of sodium dodecyl sulfate (2%) was carried out on 15% (w/v) acrylamide gels under reducing conditions as reported elsewhere (Laemmli and Favre, 1973).

Gel Permeation Chromatography—Gel permeation chromatography of the galectin was performed on a Pharmacia Superose 6 (1 × 30 cm) column equilibrated with PBS/ME, 0.25 M NaCl, 0.01 M lactose (pH 7.5) in a high performance liquid chromatography (HPLC) system that consisted of a Beckman-116 pump, a Beckman programmable detector module-166 (280 nm), and a Panasonic KX-1080 recorder, at a flow rate of 0.4 ml/min. Gel permeation chromatography of the horseradish peroxidase (HRP)-conjugated galectin was carried out with PBS (azide-
Biochemical Characterization of B. arenarum Galectin

Characterization of the Carbohydrate Specificity and Preparation of the Galectin-HRP Conjugate—The purified toad ovary galectin was carboxamidomethylated with iodoacetamide on a solid phase under mild conditions in the presence of excess ligand, yielding carboxamidomethylated galectin (CAMb-G) (200 pmol) with pepsin (10 pmol) was done in 0.01 M Tris-HCl (pH 8.5), 1 M guanidine hydrochloride (E/S = 1:50, w/w, at 30°C, 10–20 h). Cleavage of CAM-G with endoproteinase Glu-C was done in 0.1 M ammonium hydroxide carbonate (pH 8.5), 1 M urea (E/S = 1:100, w/w, at 30°C, 20 h). Cleavage of the Ac–1(–21) CAM fragment of CAM-G (200 pmol) with pepsin (10 pmol) was done in 0.01% aqueous trifluoroacetic acid (50 μl) at 25°C for 150 min. The digests of CAM-G were acidified to pH 2–3 with 10% trifluoroacetic acid, and the fragments were purified by HPLC.

HPLC Purification of Peptides—The peptides were separated on a microbore RP-HPLC system consisting of Applied Biosystems model 140A pumps and model 1000S diode-array detector (2.3-μl flow cell, 0.0255-inch inside diameter tubing). Fractionation of the peptides was performed on a C18 silica column (0.1×25 mm, 300 Å pore size, Microtech Scientific, Saratoga, CA), or on an Aquapore ODS-300 C-18 silica column (0.1×25 cm, d<sub>r</sub> = 7 μm, 300 Å pore size, Applied Biosystems) equilibrated at room temperature in 0.1% aqueous trifluoroacetic acid, and eluted at flow rates of 50–80 μl/min using linear gradients of acetonitrile/water/trifluoroacetic acid (80:20:0.1, v/v) The column effluent was monitored at 215 nm, the UV absorption spectra of the absorbing material were determined, and the eluate was manually collected and stored at −20°C prior to further analysis. Rechromatography of some fractions in the second RP-HPLC elution solvent system consisting of 2-propanol/acetonitrile/water/trifluoroacetic acid (70:20:10:0.1, v/v) was required prior to sequencing.

Sequence and Amino Acid Analysis—Automated Edman degradation of the peptides (Hewick et al., 1981) was performed on an Applied Biosystems model 477A/120A sequencing system. The PTH-amino acids were separated and identified as described (Pohl, 1994). The O-carboxamidomethyl ester derivative of PTH-Glu, PTH-Glu(CAM), which was identified during sequencing of several CAM-G peptides, co-eluted with PTH-Thr. About 20% of PTH-Glu(CAM) was converted into PTH-Glu during conversion of its PTC derivative. Automated amino acid compositional analysis of the samples was done in an Applied Biosystems model 420A/130A derivatization system equipped with an on-line hydrolysis unit assembly. The standard manufacturer’s protocol was used to carry out both acid hydrolysis and derivatization with phenylisothiocyanate. The PTC-amino acids were identified and quantitated using the Applied Biosystems model 420A/130A derivatizer/PTC-amino acid analyzer. The deamidated amino acid compositions of the samples were done in an Applied Biosystems model 420A/130A derivatizer/PTC-amino acid analyzer equipped with an on-line hydrolysis unit assembly. The standard manufacturer’s protocol was used to carry out both acid hydrolysis and derivatization with phenylisothiocyanate. The PTC-amino acids were identified and quantitated using the Applied Biosystems model 420A/130A derivatizer/PTC-amino acid analyzer equipped with an on-line hydrolysis unit assembly.
argon. The N → O acyl shift on the serine and threonine residues (Fontana and Gross, 1986) was induced by incubating the trifluoroacetic acid solution at 54 °C for 8 h or 12 h. The peptides were directly loaded onto the sequencer glass fiber disc (precelyed with 3 mg of polybrene) and were subjected to Edman degradation.

**Peptide Synthesis**—Ac-(1–13), a 13-residue N4-acetylated model peptide, was synthesized by solid-phase synthesis on the Rainin Instruments model Symphony/Multiplex peptide synthesizer, using the standard synthesis protocol utilizing the Fmoc (N-(9-fluorenyl)alkoxy carbonyl)/tetrabutyl protection strategy. The peptide was purified by preparative RP-HPLC; its mass and amino acid composition were confirmed.

**Mass Spectrometry Analysis**—The MALDI mass spectrometry analyses of fragments of CAM3-G were performed on a Kratos Instruments model COMPACT MALDI III mass spectrometer or on a Bruker Instruments Protein TOF mass spectrometer. The HPLC fractions (0.3 μl) were spotted on a target site of a 20-sample slide, followed by addition of 0.3 μl of matrix (saturated a-cyano-4-hydroxycinnamic acid, Aldrich) dissolved in ethanol/water (1:1, v/v). The sample matrix was allowed to dry at room temperature for 5 min, and each sample was desorbed with 50 laser shots, each giving a spectrum. The shots were averaged to give the final spectrum. The instrument was operated in the linear mode and was calibrated using the external standard peptides. All ES1 mass spectrometry analysis and tandem mass spectrometry analysis was performed at the Emory University Mass Spectrometry Center on a JEOL SX102/SX102B five sector mass spectrometer (configuration BEE-BEE) using a JEOL Generation 3 ESI source. All peptides were dissolved in either 5.51 or 5.15:1 water/methanol/acetic acid and introduced by flowing the solution into the ESI source at a flow rate of 1 μl/min at concentrations of 1–20 pmol/μl. The full scan spectra were 50–150 scans signal averaged at a resolving power of 1000 or 3000. Tandem mass spectrometry experiments were performed by mass selecting the ion of interest with MS1(B1E1) and examining the metastable products using link B/E scans (B2E2). The narrow range tandem mass spectrometry experiment was 1353 spectra signal averaged. The full range tandem mass spectrometry experiments were over 273 scans signal averaged over the m/z range of 550–700. These conditions provided the best signal to noise ratio for the ions of interest.

**RESULTS**

**Galectin Purification and Homogeneity**—The improved protocol optimized for a mammalian galectin-1 (Ahmed et al., 1996) was equally satisfactory for purification of the toad ovary galectin. The use of DEAE-Sepharose prior to affinity chromatography allowed the partial isolation of the galectin and the removal of lactose present in the extraction buffer. The high salt eluate from DEAE-Sepharose was immediately loaded on the lactoysl-Sepharose column, the resin was washed to the base line, and the bound protein eluted with lactose. The yield of the purified protein was 8–10 mg/kg of wet ovaries and had a specific activity of 2–5 x 10^4 mg⁻¹ ml⁻¹. The purified galectin showed a single polypeptide corresponding to approximately 14.5 kDa on SDS-polyacrylamide gel electrophoresis under reducing conditions (Fig. 1C). However, gel permeation chromatography under non-denaturing conditions suggested that the galectin subunits may undergo a concentration-dependent dimerization (Fig. 1A). At the loading concentration of 1.4 mg/ml, the dimer:monomer ratio was 48:1, whereas at a concentration of 15 μg/ml, the equilibrium shifted toward the monomeric species, with a dimer:monomer ratio of 3:1. Molecular mass estimates from gel permeation chromatography were 14.5 and 32 kDa for the monomer and dimer, respectively (Fig. 1B). On isoelectric focusing, the galectin exhibited nine isoforms with pI values ranging from 4.49 to 4.92 (Fig. 2) with prevalent components at pI 4.73 (44%), 4.78 (19%), and 4.8 (17%).

**Stability**—The toad ovary galectin retained full agglutinating activity when maintained at temperatures up to 37 °C, but a gradual decrease of activity occurred between 42 and 70 °C. The activity was completely abolished when the protein was heated at 85 °C for 30 min (Fig. 3). The stability of the lectin binding activity in nonreducing environments was examined under two different conditions: adsorbed on affinity matrices and in solution, with and without its ligand. The lectin, adsorbed on lactoysl-Sepharose or asialofetuin-Sepharose in a buffer containing no reducing agent, retained full activity after 8 days at room temperature. The lectin also retained full activity in the absence of a reducing agent if maintained in solution in the presence of excess ligand (0.1 M). The activity of the lectin, however, was reduced almost 40-fold in solution, in the absence of both reducing agent and ligand, and could not be restored by the addition of the reducing agent (data not shown).

**Preparation of Toad Ovary Galectin-HRP Conjugate and Optimization of the Solid Phase Galectin Binding Assay**—The purification of galectin-HRP conjugate was monitored by assessing both its binding and enzyme activities at each step. After the conjugation procedure, approximately 60% of the total protein was recovered between lactoysl-Sepharose. Gel permeation chromatography on Superose 6 allowed the separation of the galectin-HRP conjugate from unreacted galectin (Fig. 4). The peak A (Fig. 4) corre-

![Fig. 1. Native and subunit structure of toad ovary galectin.](image-url)

**Fig. 1. Native and subunit structure of toad ovary galectin.** A. gel permeation chromatography of toad ovary galectin on Superose 6 (1 x 30 cm) showing concentration-dependent monomer-dimer equilibrium in PBS/ME, 0.25 mM NaCl, 0.01 mM lactose, pH 7.5: a. 15 μg/ml; b. 140 μg/ml; c. 1.4 mg/ml. Molecular mass standards were 1, BSA (66 kDa); 2, ovalbumin (43 kDa); 3, carbonic anhydrase (29 kDa); 4, ribonuclease A (13.7 kDa). B, estimation of molecular weight of the monomer and dimer. C, estimation of subunit molecular weight of the lectin (indicated by the arrow) from SDS-polyacrylamide gel electrophoresis under reducing condition (0.5% 2-mercaptoethanol). The sample (2 μg) was loaded on 15% polyacrylamide gel using a discontinuous buffer system and stained with Coomassie Blue. Standards (Sigma) from higher to lower molecular weight were: BSA, ovalbumin, glyceraldehyde-3-phosphate dehydrogenase, carbonic anhydrase, trypsinogen, soybean trypsin inhibitor, myoglobin (1-153), α-lactalbumin, myoglobin 1+III.
sponding to approximately 71 kDa, represents the galectin-HRP conjugate, most likely resulting from equimolar crosslinking of galectin (29.6 kDa) and HRP (approximately 44 kDa), and exhibits both peroxidase and hemagglutinating activity. The fractions containing the 71-kDa species were pooled and used for binding assays. About 20% of the affinity-purified protein was recovered in the active galectin-HRP conjugate pool. To optimize the solid phase assay, the wells of the microtiter plates were coated with varying concentrations (0.015–16 mg/well) of ASF, followed by fixing and extensive washing, and finally, the addition of varying concentrations (5–40 ng/well) of galectin-HRP conjugate. Fig. 5a shows the binding profiles for each conjugate concentration tested: increased binding was observed with increasing amounts of conjugate added for all concentrations of the coating ASF. Except for the highest conjugate concentration tested (40 ng/well), maximum binding was achieved at approximately 0.5 μg of ASF/well, with absorbance values reaching a plateau beyond that concentration. Fig. 5b shows the binding of variable amounts of conjugate to ASF-coated wells (0.5 μg/100 μl/well) as a function of time. For all conjugate concentrations tested, the binding was approximately linear during incubation times up to 21 min. From these preliminary studies, the optimal amount of ASF for coating the plates was established as 0.5 μg/100 μl/well, and the amount of galectin-HRP conjugate to be added for the binding and binding inhibition experiments, as 20 ng/100 μl/well. The optimal substrate incubation time was determined to be 20 min. Approximately 2 ng of the galectin-HRP conjugate was bound to ASF under the optimal conditions. No further blocking of the ASF-coated wells was required, but the addition of Tween 20 to binding and washing buffers substantially reduced any background absorbance (results not shown). Under the optimal conditions established, increasing concentrations of lactose inhibited the binding of galectin-HRP conjugate to ASF in an approximately sigmoid profile reproducible on several experiments (Fig. 5c). The lactose concentration required for 50% inhibition of the galectin-HRP conjugate binding to ASF varied from 60 to 90 μM. The binding activity of toad ovary galectin to ASF was optimum at pH 7.5–8.5 (Fig. 5d). The activity declined drastically below pH 7.0. All further experiments were carried out at pH 7.5. As optimized, this assay for determining galectin’s carbohydrate-binding specificity is simple, sensitive, and rapid, since it was completed in less than 3 h after the plates were coated with ASF.

**Carbohydrate Specificity**—The carbohydrate-binding specificity of toad ovary galectin was determined by analyzing the binding of the galectin-HRP conjugate to ASF in the presence of several saccharides through the solid phase assay, under the optimal conditions established. Preliminary studies applying this method to selected carbohydrate ligands yielded a relative inhibitory activity profile identical to that obtained through hemagglutination-inhibition assays with the unmodified toad ovary galectin (Fink et al., 1987), indicating that the specificity of the lectin was not modified through the conjugation procedure. This validation of the solid phase method developed allowed comparisons of the relative inhibitory activities of carbohydrates tested with those estimated through other well-established methods that use unmodified (Ahmed et al., 1990) or radiolabeled (Leffler and Barondes, 1986) galectins. For each test saccharide, a complete inhibition curve was determined, and the molar concentrations that inhibited the binding of the lectin conjugate to ASF by 50% (I50) were calculated and nor-
malized with respect to the lactose included in each plate as a standard (Table I). The inhibitory potencies of several test oligosaccharides and methyl glycosides relative to the \( I_{50} \) value of galactose, the nonreducing terminal monosaccharide common to all, are summarized in Table II. The binding of the toad ovary galectin with Galβ1,3(4)GlcNeuAc and thiodigalactoside were about 4-fold higher than that with lactose. The binding of the galectin to the human blood group A tetrasaccharide Fucα1,2GalNAcβ1,3Galβ1,4Glc was about 9-fold weaker than to lactose. The binding to Galβ1,3GalNAc was negligible. Therefore, the overall binding inhibition pattern of the toad ovary galectin suggests a specificity very similar to those observed in mammalian 14-kDa lectins exhibiting conserved (type I) carbohydrate-recognition domains (Ahmed and Vasta, 1994), but rather different from that of the clawed frog (type I) carbohydrate-recognition domains (Ahmed and Vasta, 1992).

**Experimental Procedures.**

Inhibition of toad ovary galectin binding to asialofetuin by a series of oligosaccharides expressed as \( I_{50} \) of each compound relative to lactose \( I_{50} \).

**TABLE I**

Inhibition of binding of toad ovary galectin-HRP conjugate to asialofetuin was determined by solid phase assay as described under “Experimental Procedures.” In the binding curve (m), \( 100 \) ng of the native lectin or CAMb-G did not yield an unambiguous sequence by Edman degradation (see below), suggesting that these are the expected peptides. All expected peptides were purified by RP-HPLC and that did not yield an unambiguous sequence on Edman degradation (see below), suggesting that these are the expected tryptic and lysyl endopeptidase cleavage sites were obtained by sequencing the Asp-N and Glu-C digests of CAMb-G (Fig. 6). The core of the amino acid sequence was thus elucidated; however, the structure of the blocked N terminus remained to be determined.

**TABLE II**

| Compounds | Relative inhibitory activity |
|-----------|-----------------------------|
| Galβ3     | 1.0                         |
| Galα-OMe  | 6.0                         |
| Galβ-OMe  | 1.2                         |
| GalNAc    | <1.2                        |
| MeO6-Gal  | <1.2                        |
| Galα1,6Glc| 1.5                         |
| Galβ1,6Glcβ1,2Fruf | <2.0 |
| Galα1,6Galα1,6Glcβ1,2Fruf | 1.00 |
| Galβ1,4Glc| 200.0                       |

* All sugars are in \( \alpha \) form.

**Primary Structure Analyses**—Edman degradation of 100–200 pmol of the native lectin or CAMβ-G did not yield an N-terminal sequence, indicating that the protein N terminus is not accessible to sequencing, i.e., “it is blocked.” CAMβ-G was then cleaved with trypsin and lysyl endopeptidase, and the resulting peptides were purified by RP-HPLC. All expected fragments from both digests were recovered and completely sequenced (Fig. 6); however, it was noted that cleavage of several tryptic and/or lysyl endopeptidase sensitive bonds (e.g., Lys183-Pro14, Lys289–Gly290, Lys464–Glu465, and Arg74–Glu79) proceeded slowly and led to substantial material losses during HPLC purification. For this reason, the protein was also cleaved using a mixture of trypsin and lysyl endopeptidase. These fragments were recovered in high yield on RP-HPLC, were all sequenced, and had their masses confirmed by MALDI. The overlaps for the tryptic and lysyl endopeptidase cleavage sites were obtained by sequencing the Asp-N and Glu-C peptides of CAMβ-G (Fig. 6). The core of the amino acid sequence was thus elucidated; however, the structure of the blocked N terminus remained to be determined.

**The N-terminal Structure**—The tryptic and lysyl endopeptidase digests of CAMβ-G contained peptides, Ac(1–21) and Ac(1–13), respectively, that were eluted as distinct entities on RP-HPLC and that did not yield an unambiguous sequence on Edman degradation (see below), suggesting that these are the blocked N-terminal peptides of CAMβ-G.2 Cleavage of Ac(1–21) with pepsin afforded five fragments: three yielded a unique sequence, VAVTNL, VTNL, and NLKPGHCVEIK. The other two fragments were N-terminally blocked; their MALDI mass spectrum indicated that the N-terminal sequences were VAVTNL and VTNL, respectively. Using a combination of N-terminal sequencing and Edman degradation, the N-terminal sequence was unambiguously determined; however, the structure of the noncleaved N terminus remained to be determined.

**Methods**

The N-blocked peptides were also identified in the endoproteinase Asp-N and Glu-C digests of CAMβ-G.
concluded that a acid treatment induced the acyl N-terminal residues identified in the acid-treated material, it was Ac-(1–21) (see Fig. 6). Since Ser and Thr were the only N-terminal sequences of the trifluoroacetic acid treated Ac-(1–13), and HCVEIK in Ac-(1–21), was identified (Table III). Two sequenceable fragments were found in each fraction, and the sequencing analysis was unsuccessful. The peptic peptide of Ac-(1–21), NLKPGHCVEIK (observed average molecular mass: 1295 daltons, calculated: 1295 daltons), identifies Ac-(1–21) as the N-terminal fragment of the protein (Fig. 6).

Sequencing of both Ac-(1–21) and Ac-(1–13) yielded a low PTH signal in the first several cycles. The PTH signal for several amino acids was, however, onset and leveled off in the later cycles as is demonstrated for Ala, Gly, and Val in Table III. It was reasoned that the delayed onset of the sequencing signal was caused by an incremental removal of the N-protective group, or by a partial cleavage of the selected peptide bonds. Such modifications would likely occur during the acid cleavage step of Edman degradation (Fontana and Gross, 1986; Weller et al., 1990). To test this possibility, the peptides were treated with trifluoroacetic acid at 54°C for 8 or 12 h prior to sequencing. Indeed, Edman degradation of the acid-treated peptides proceeded efficiently from the first cycle: two sequenceable fragments were found in each fraction, and the C-terminal sequence of the longer fragment, TNLNLK in Ac-(1–13), and HCVEIK in Ac-(1–21), was identified (Table III). These C-terminal sequences align with the peptic fragments of Ac-(1–21) (see Fig. 6). Since Ser and Thr were the only N-terminal residues identified in the acid-treated material, it was concluded that acid treatment induced the acyl N → O shift on these two residues. Clearly, the N → O shift on the Thr residue opened the TNLNLK sequence. By inference, therefore, the N → O shift on the Ser residue opened the SAGAVTNLKL sequence. Hence, the partial sequence of Ac-(1–21) was identified as SAGAVTNLKLPGHCVEIK (X being the blocking group).

The identity of the blocking group was determined by amino acid analysis. The difference between the amino acid composition of Ac-(1–13), Asx3, Ala3, Ser1, Gly1, Thr1, Val2, Leu2, Lys1 (Table IV), and its partial sequence determined by Edman degradation, SAGAVTNLKL, is one alanine residue. The difference between the determined monoisotopic molecular mass of the peptide (1298.7 daltons) and its monoisotopic molecular mass based on its amino acid composition (calculated: 1256.7 daltons) is 42.0 daltons, suggesting that X is N-acetyl-alanyl. To further determine that the N-terminal sequence of Ac-(1–13) was indeed C2H3O-Ala-Ser-, tandem mass spectrometry was performed. The peptide should yield a distinct \( y_{1w} \) (nomenclature of Roepstorff and Fohlman (1984)) fragment of 1186.68 daltons. The doubly charged species (\( m/z = 650 \)) was chosen for the tandem mass spectrometry. Experiments with synthetic Ac-(1–13) showed that the proposed sequence produced a doubly charged \( y_{1w} \). Therefore, a narrow range metastable mass spectrum of the isolated peptide was obtained. Indeed, the \( y_{1w} \)-ion (\( m/z = 593.8 \)) that was produced corresponds to the C2H3O-Ala-Ser- sequence (data not shown). Attempts to obtain full scan range tandem mass spectra to prove by mass spectrometry the entire sequence were unsatisfactory because of the amount of the sample available. However, the observed peaks in the full scan range tandem mass spectra were consistent with the proposed sequence (not shown).

Molecular Mass and Amino Acid Composition of Galectin—The ESI mass spectra of the full-length protein are in complete agreement with the determined sequence. The ESI of the native protein found ions that correspond to average masses of 14,791, 14,867, 14,943, and 15,020 daltons. These correspond to the pure native protein and to its mono-, di-, and tri-disulfide-linked adducts with 2-mercaptoethanol (calculated: 14,797, 14,873, 14,949, 15,015).4 CAMa-G, carboxamidomethylated at its four cysteine residues, gave an ion that corresponds to an average molecular mass of 15019 daltons (calculated: 15,025 daltons). A minor species of CAMa-G (average molecular mass: 15,078) that was also identified likely corresponds to CAMa-G esterified at Glu-37 (calculated: 15,082) by iodoacetamide.5 The amino acid composition of CAMa-G (Table IV) is in excellent agreement with the determined sequence.

Comparison to Other Galectins—The amino acid sequence of the toad ovary 14.8-kDa lactose-binding lectin was compared to those of galectins from bovine (Abbott et al., 1989; Ahmed et al., 1996) human (Hirabayashi et al., 1989), rat (Clerc et al., 1988), mouse (Wells and Malluci, 1991), chicken (Hirabayashi et al., 1987; Sakakura et al., 1990), frogs (Ozeki et al., 1991; Marshal et al., 1992), teleost fishes (Paroutaud et al., 1987; Muramoto and Kamiya, 1992), nematode (Hirabayashi et al., 1992, 1996), and a sponge (Pfeifer et al., 1993) (Fig. 7). The toad ovary lectin is clearly a member of the soluble β-galactoside-binding lectin family, and it contains all relevant amino acids that are present in the galectin-1 subgroup. Unlike galectin-1, the toad ovary galectin contains only 3 of the 6 conserved cysteine residues that are critical for oxidative inactivation. The percent identities with galectins-1 from homeotherm vertebrates such as human, bovine, rat, and mouse were 42–48%, being maximum with those from bovine spleen/spleen. These galectins share 87–95% of their amino acid sequences. The percent identities with galectins from the different taxa compared were as follows: 45% with two galectins from another poykilotherm vertebrates (chicken); 28–38% with galectins from poykilotherm vertebrates (X. laevis (38%), electric eel and conger eel (28%)); and 14–23% with galectins from invertebrates (nematode C. elegans (21–23%) and sponge G. cydonium (14%)). The short peptide sequence available for Rana oocyte lectin was 48% identical to the toad ovary galectin sequence. Within the CRD, the percent of identities with all galectins compared ranged between 31 and 77%, being maximum with bovine spleen lectin (77%). The percent identities with CRDs of galectins from homeotherm vertebrates were higher (63–77%) than with those from poykilotherm vertebrates (33–50%) and invertebrates (31–43%). The percent identity between the CRD of the toad ovary galectin and that of another

3 As 2-mercaptoethanol (0.01 M) was present in all buffers during purification of the lectin, the adducts with 2-mercaptoethanol were likely formed by air oxidation during dialysis at pH 8.5.

4 The esterified peptides were identified in all proteolytic digests.
TABLE III
Edman degradation of Ac-(1–13) and Ac-(1–21)

The N-terminal peptides of CAMb-G, Ac-(1–21) and Ac-(1–13), were subjected to Edman degradation yielding no PTH-amino acid signal above the background level signal in the first cycle, as demonstrated for PTH-Ala, -Gly, and -Val (data for Ac-(1–13) are not shown). Following exposure of Ac-(1–21) and Ac-(1–13) to neat TFA, two PTH-amino acids were identified in each sequencing cycle for both peptides, beginning with PTH-Ser and -Thr in the first cycle.

| Cycle | Ac-(1–21) (250 pmol), untreated | Ac-(1–13) (170 pmol), trifluoroacetic acid-treated<sup>a</sup> | Ac-(1–21) (130 pmol), trifluoroacetic acid-treated<sup>a</sup> |
|-------|-------------------------------|-------------------------------------------------|-------------------------------------------------|
|       | Ala | Gly | Val | Ala | Gly | Val | Ala | Gly | Val |
| 1     | 0.6 | 2.9 | 0.5 | Ser (6.6)<sup>b</sup> | Thr (1.2)<sup>c</sup> | Ser (2.3)<sup>b</sup> | Thr (2.1)<sup>c</sup> |
| 2     | 1.4 | 2.3 | 0.1 | Ala (45) | Asn (7.4) | Ala (30) | Asn (4.3) |
| 3     | 12.0 | 3.3 | 0.2 | Gly (46) | Leu (29) | Gly (32) | Leu (12) |
| 4     | 8.0 | 17.0 | 11.0 | Asn (9.3) | Leu (26) | Asn (4.5) | Leu (12) |
| 5     | 7.4 | 13.0 | 11.0 | Ala (46) | Ala (25) | Ala (25) | Leu (12) |
| 6     | 19.0 | 11.0 | 10.0 | Val (64) | Lys (0.8) | Val (29) | Lys (3.1) |
| 7     | 12.0 | 8.6 | 16.0 | Thr (9.3)<sup>c</sup> | Ala (45) | Thr (7.5)<sup>c</sup> | Pro (6.1) |
| 8     | 10.0 | 9.0 | 14.0 | Asn (10) | Asn (9.3) | Asn (9.3) | Gly (5.1) |
| 9     | 8.8 | 8.0 | 11.0 | Leu (26) | Leu (26) | Leu (22) | His (0.2) |
| 10    | 8.8 | 8.2 | 11.0 | Asn (7.9) | Asn (9.0) | Asn (9.0) | Cys (CIM) (2.1) |
| 11    | 7.5 | 7.0 | 10.0 | Leu (19) | Leu (18) | Leu (18) | Val (3.5) |
| 12    | 6.6 | 6.8 | 9.0 | Lys (0.5) | Lys (4.5) | Lys (4.5) | Glu (2.7) |
| 13    | 5.5 | 6.1 | 8.3 | Pro (8.4) | Pro (8.4) | Pro (8.4) | Ile (4.8) |
| 14    | 4.9 | 6.3 | 7.0 | Gly (17) | Gly (17) | Gly (17) | Lys (1.7) |
| 15    | 5.1 | 6.6 | 7.2 | His (1.0) | His (1.0) | His (1.0) | Lys (1.0) |
| 16    |      |      |      | Cys (CIM) (5.0) |      |      |      |
| 17    |      |      |      | Val (12) |      |      |      |
| 18    |      |      |      | Glu (4.5) |      |      |      |
| 19    |      |      |      | Ile (8.1) |      |      |      |
| 20    |      |      |      | Lys (3.5) |      |      |      |

<sup>a</sup> The N → O shift was induced on Ser and Thr. The cleavage step of Edman degradation is affected by heating the sample in liquid trifluoroacetic acid at 54 °C for 10 min. The trifluoroacetic acid-treated peptides were heated at 54 °C for 8 h (Ac-(1–13)) or 12 h (Ac-(1–21)) prior to sequencing.

<sup>b</sup> The PTH derivatives of dehydro-Asp and its adducts with dithiothreitol were present but not quantitated.

<sup>c</sup> The PTH derivatives of dehydroalanine and its adducts with dithiothreitol were present but not quantitated.

TABLE IV
Amino acid composition of CAMb-G and Ac-(1–13)

The experimental values are the averages of three independent determinations; the numbers in parentheses are derived from the CAMb-G sequence.

| Residue<sup>a</sup> | CAMb-G (260 pmol) | Ac-(1–13) (350 pmol) |
|---------------------|------------------|---------------------|
| Asx                 | 14.3 (14)        | 2.18 (2)            |
| Cys(Cm)             | 4.10 (4)         | NF<sup>b</sup> (0)  |
| Glx                 | 16.0 (16)        | 0.09 (0)            |
| Ser                 | 10.4 (11)        | 0.97 (1)            |
| Gly                 | 10.3 (10)        | 1.07 (1)            |
| His                 | 2.83 (3)         | NF (0)              |
| Arg                 | 3.01 (3)         | NF (0)              |
| Thr                 | 4.25 (4)         | 0.96 (1)            |
| Ala                 | 9.08 (9)         | 2.92 (3)            |
| Pro                 | 7.23 (7)         | 0.22 (0)            |
| Tyr                 | 1.05 (1)         | NF (0)              |
| Met                 | 0.87 (1)         | NF (0)              |
| Val                 | 10.1 (11)        | 2.20 (2)            |
| Ile                 | 6.23 (6)         | NF (0)              |
| Leu                 | 10.0 (10)        | 2.00 (2)            |
| Phe                 | 11.9 (12)        | NF (0)              |
| Lys                 | 8.74 (9)         | 1.01 (1)            |
| Trp                 | ND<sup>c</sup> (1) | ND<sup>c</sup> (0)  |

<sup>a</sup> Asx = Asp + Asn; Glx = Glu + Gln.

<sup>b</sup> NF, not found.

<sup>c</sup> ND, not determined.

The UV spectrum of HPLC fractions containing Ac-(1–13) had no absorption between 250 nm and 290 nm indicating that the peptide does not contain aromatic amino acid residues.

amphibian, the clawed frog *X. laevis* was 48%. The percent similarities of the toad ovary galectin were 82 and 79% with bovine spleen galectin-1 and *X. laevis* galectin, respectively. We compared the predicted secondary structures of toad ovary galectin, bovine spleen galectin, and *X. laevis* galectin, calculated according to the method of Chou and Fasman (1978) (Fig. 8). The three-dimensional structure of the bovine spleen galectin-1 (Liao et al., 1994) showed a correlation with its predicted secondary structure in 9 of 11 β-strands. Although the overall secondary structures of all three galectins (Fig. 8, a–c) were similar, within the CRDs (Fig. 8d), the toad and bovine galectins showed almost identical profiles, whereas significant differences between those and the *X. laevis* galectin were observed at positions corresponding to residues 52, 53, 63–70, and 73 (residue numbers are those of bovine spleen galectin). The hydrophathy profiles of the above three galectins (Fig. 9) also indicated that the CRDs of toad ovary and bovine spleen galectins similar (Fig. 9, a and b), but both considerably different from that of *X. laevis* galectin (Fig. 9c).

**DISCUSSION**

The concentration-dependent monomer-dimer equilibrium observed for the purified toad ovary galectin is noteworthy and similar to that reported for the galectin-1 from CHO cells (Cho and Cummings, 1995). This noncovalent dimerization also resembles observations for other lectins that yield higher order aggregates as the protein concentration is increased (Vasta et al., 1986). Furthermore, experimentally cross-linked galectins are more effective than dimeric galectins in mediating cell adhesion to plastic surfaces (Ahmed et al., 1992). Because most galectins have one binding site per subunit, it becomes clear that properties, such as agglutination, glycan precipitation, or the cross-linking between cells and extracellular matrix, would be mediated by the dimer rather than the monomeric form. Although the monomer has carbohydrate-binding capability which is independent of dimer formation, at very low concentrations the galectin may exist mostly as a monomer and as such, remain "inactive" in mediating any biological events requiring cross-linking of carbohydrate moieties. The dimer:monomer ratio (48:1) at the highest concentration (1.4 mg/ml, equivalent to 47 μM) of toad ovary galectin was comparable to that observed for CHO galectin-1 at 80 μM (Cho and Cummings, 1995). However, the concentration range of dimer:monomer equilibrium for the toad ovary galectin was broader than that of CHO galectin-1. At 800 nM CHO galectin-1 exists mostly...
as monomer (80%) (Cho and Cummings, 1995), whereas most (75%) of the toad galectin was in dimer form at 15 mg/ml (500 nM). The concentration at which most of the toad ovary galectin was present as a monomer was not established in the present study, but it would probably be below 20 ng/ml, because the lectin still agglutinated rabbit erythrocytes at that concentration (specific activity 2–5 \times 10^4 mg^{-1} ml^{-1}).

As reported for most highly purified galectin preparations from diverse animal sources, the presence of reproducible isoform profiles was observed in multiple preparations of the toad ovary galectin. The range of pIs of the nine lectin bands was higher than the one reported earlier (Fink et al., 1987), and the pI values of the three major components (pI 4.73–4.80) were close to the theoretical value (pI 4.92) calculated from the primary structure. Peptide sequencing of the purified protein failed to yield alternate residues at any position that could provide a structural basis for the presence of isoforms, and therefore, it is unlikely that point mutations are the mechanism responsible. For galectin-3, the unequal phosphorylation of serine residues was shown to be the cause of isolectin formation (Huflejt et al., 1993), and although it is possible that such derivatization of amino acids is the source of the heterodispersity observed in most galectins, this has not been demonstrated yet for galectin-1.

The analysis of the optimal binding activity and stability of the toad ovary galectin under a variety of experimental conditions revealsthat this protein can remain fully active for a long period of time in the absence of reducing agents, only if the cysteines are carboxamidomethylated or if the intact protein is stored in the presence of soluble or solid phase-bound ligand, in this case lactosyl- and ASF-Sepharose. Similar results were obtained with the CHO galectin-1 when bound to laminin-Sepharose (Cho and Cummings, 1995). For the electric eel galectin, lactose protects the lectin against inactivation as assessed by prevention of the loss of fluorescence by oxidation (Levi and Teichberg, 1981). It has been proposed that the
also similar to the bovine spleen galectin and other mammalian galectins, but surprisingly different from that of the clawed frog *X. laevis*, another amphibian species. On the basis of differences in carbohydrate-binding patterns and the conservation of critical amino acid residues in the CRD, we categorized galectins into two groups, type I (conserved) and type II (variable) (Ahmed and Vasta, 1994). The relative inhibitory efficiencies of four key oligosaccharide structures, lactose, N-acetyllactosamine, Gal$_1$3GalNAc, and the human blood group A-tetrasaccharide (Fig. 10a–d) can provide the preliminary information required for the assignment of a galectin to either group. The hydroxyls (Fig. 10, *boldface letters*) at C-4' and C-6' of galactose residue are critical for both conserved and variable CRDs. Neither epimerization (for OH at C-4') nor substitutions (for both OH) are possible without considerable changes in binding affinity. The 3-OH (*shadowed background*) of Glc/GlcNAc (in Gal$_1$4Glc or Gal$_1$4GlcNAc) or the 4-OH of GlcNAc (in Gal$_1$3GlcNAc) cannot be epimerized or substituted for the conserved CRDs, but for variable CRDs the epimerization is allowed, since Gal$_1$3GlcNAc (Fig. 10c) is an equally potent inhibitor as Gal$_1$4Glc in RI36-I (Oda et al., 1993; Ahmed and Vasta, 1994). The substitution of 2-OH (*outlined letters*) of Glc residue by NHAc in Gal$_1$4GlcNAc (Fig. 10b) promoted binding 5–10-fold compared to Gal$_1$4Glc for conserved CRDs, but for variable CRDs, various degrees of binding (negligible to 11-fold better) were observed (Ahmed and Vasta, 1994). The substitution of OH (*italic letters*) at C-2' or C-3' individually or both, did not affect the binding dramatically. The 2–3-fold weaker binding in the case of conserved CRD was observed probably because of steric hindrance between the bulky substituents and the interacting amino acids of the protein. Interestingly, for some galectins having variable CRDs, the substitutions at C-2' and C-3' (Fig. 10d) promoted binding 10–32-fold (Leffler and Barondes, 1986; Sparrow et al., 1987; Oda et al., 1993), probably due to deletions in the CRD domain that yield a structure that can accommodate the bulky substitutions (Ahmed and Vasta, 1994). In summary, the analyses of sugar
binding specificity suggest that the relative binding affinity of the conserved (type I) CRDs for the oligosaccharides in question would be in the following decreasing order: N-acetyllactosamine > lactose > A-tetrasaccharide > Gal β1,3GalNAc. Both the galectin-1 from bovine spleen (Ahmed et al., 1994, 1996) and the toad ovary galectin (this study) showed the above order of specificity.

The results of peptide sequencing described above (Fig. 6) established that the 14.8-kDa lactose-binding lectin from the B. arenarum ovary is a member of the S-lac lectin family and very close to the galectin-1 group (Barondes et al., 1994a). Within the galectin-1 group, it most closely resembled the galectin-1 from bovine heart or spleen (48% overall identity) (Fig. 7), and contrary to what might be expected from consideration of the phylogenetic distances, the overall identity was 38% with the 16-kDa galectin from another amphibian species X. laevis. Interestingly, the carbohydrate specificities of the two amphibian and mammalian galectins reflected their structural relationships. Studies on three-dimensional structures of galectin-1 from bovine spleen and heart (Liao et al., 1994; Bourne et al., 1994) and galectin-2 from human (Lobsanov et al., 1993) suggest that the CRD comprises the amino acid residues 44–73 (residue numbers are those of bovine spleen (Liao et al., 1994)). Within the CRD, the relevant amino acids responsible for sugar binding (His44, Asn46, Arg48, His52, Asp54, Asn61, Trp68, Glu71, and Arg73) are identical among the sequences of toad ovary galectin and the galactos-1 from bovine spleen/heart, human lung and rat lung. The X. laevis galectin CRD also contains all the above amino acids except for positions 52 (Ser instead of His) and 73 (Lys instead of Arg).

Although all three galectins from toad ovary, bovine spleen and X. laevis apparently showed an overall similarity in secondary structure (Fig. 8a–c) as calculated according to the method of Chou and Fasman (1978), it is clear that the profile for the toad ovary galectin is almost identical to that of bovine spleen, and both are less similar to the X. laevis galectin. Within the CRDs, the β-sheet content profiles (Fig. 8d) for toad ovary (closed symbol) and bovine spleen (open symbol) are considerably different from that of X. laevis (cross symbol) in residues 52, 53, 63–70, and 73 (residue numbers are based on the bovine spleen galectin amino acid sequence). Among those, Arg73 is directly involved in the binding of bovine spleen galectin-1 to N-acetyllactosamine hydroxyl on C-3 of GlcNAc, Trp68 is involved in hydrophobic interactions with the Gal pyranose ring, whereas the main chain carbonyl of His52 and side chain of Arg73 participate in water-mediated interactions with the N atom of the N-acetyl group (Liao et al., 1994). X. laevis lacks His52 and Arg73 in its CRD and also shows a decrease of binding toward N-acetyllactosamine as compared to that for lactose. The hydrophathy profile of the toad ovary galectin CRD also closely resembled that of the bovine spleen galectin CRD, but both differ from the X. laevis profile (Fig. 9). Therefore, it is likely that the differences in relevant amino acid residues (Ser instead of His52 and Lys instead of Arg73) in the CRD of X. laevis constitute the structural basis for the differences in carbohydrate specificity observed between this galectin and that from toad ovary.

In the present study we have determined the primary structure of the toad ovary galectin that reveals the presence of all relevant amino acids identical to mammalian galectin-1 and exhibits a type I (conserved) CRD, also found in other mammalian examples such as bovine spleen galectin-1. As can be predicted from the conserved CRD, the toad ovary galectin exhibits carbohydrate-binding specificities very similar to that described in bovine spleen galectin-1 (Ahmed et al., 1996). Therefore, our results suggest that galectins carrying conserved (type I) CRDs, such as the B. arenarum ovary galectin and those with variable (type II) CRDs and represented by the X. laevis 16-kDa galectin, are clearly distinct subgroups in the extant amphibian taxa and may have diverged early in the evolution of chordate lineages. Whether a 14-kDa galectin similar to the toad ovary galectin and to those included in the galectin-1 subgroup, such as the bovine galectin-1, is present in X. laevis ovary remains to be demonstrated. A preliminary assessment on X. laevis eggs and embryos by conventional methodology (Nishihara et al., 1986; Milos et al., 1990) and by sensitive immunological and RNase protection assays (Marschal et al., 1994) reported elsewhere, failed to reveal the presence of any galectin activity. Based on the homologies among the three galectins calculated either for their full-length amino acid sequences (48% between toad ovary and bovine spleen and 38% between toad ovary and X. laevis) or for their CRDs alone (77% between toad ovary and bovine spleen and 47% between toad ovary and X. laevis) it should be concluded that the correlation of their phylogenetic distances among the source species may not be as direct for the galectin family members as proposed elsewhere (Hirabayashi and Kasai, 1993; Kasai and Hirabayashi, 1996). A likely explanation for this observation may be that mutation rates would not be constant within the galectin family and the galectin-1 subgroup may have evolved under low mutation rates due to functional constraints. If this is the case, the galectin-1 would be a considerably homogeneous category within this lectin family and the high conservation of residues in the type I CRD that interact directly with the ligand, relative to those in the Type II CRD (Ahmed and Vasta, 1994) would buttress this idea.

Acknowledgments—We thank M. C. Sullards for the ESI mass spectrometry analysis of the whole protein, Dr. A. Woods for help with
MALDI mass analysis of several peptides, and Drs. M. S. Quesenberry and F. Hubalek for critically reading the manuscript.

REFERENCES

Abbott, W. M., Hounsell, E. F., and Feizi, T. (1988) Biochem. J. 252, 283–287
Abbott, W. M., Mellor, A., Edwards, Y., and Feizi, T. (1989) Biochem. J. 259, 283–290
Ahmed, H., and Vasta, G. R. (1994) Glycobiology 4, 545–549
Ahmed, H., Allen, H. J., Sharma, A., and Matta, K. L. (1990) Biochemistry 29, 5315–5319
Ahmed, H., Sharma, A., DiCioccio, R. A., and Allen, H. J. (1992) J. Mol. Recogn. 1, 1–8
Ahmed, H., Fink, N. E., and Vasta, G. R. (1994) Ann. N. Y. Acad. Sci. 712, 315–317
Ahmed, H., Fink, N. E., Pohl, J., and Vasta, G. R. (1996) J. Biochem. (Tokyo) 120, 1007–1019
Allen, H. J., Ahmed, H., and Sharma, A. (1992) Comp. Biochem. Physiol. 103B, 313–315
Barondes, S. H., Castronovo, V., Cooper, D. N. W., Cummings, R. D., Drickamer, K., Feizi, T., and Camilliau, C. (1994) Nat. Struct. Biol. 1, 863–870
Barondes, S. H., Cooper, D. N. W., Gitt, M. A., and Leffler, H. (1994b) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–147
Clerch, L. B., Whitney, P., Haas, M., Iliev, K., Miller, T., Warner, R., and Massaro, D. (1988) Biochemistry 27, 692–699
Colnot, C., Ripoche, M.-A., Scaerou, D., Fowlis, D., and Poirier, F. (1996) Biochem. Soc. Trans. 24, 141–146
Eloila, M. T., Fink, N. E., and Herkovits, J. (1987) Braz. J. Med. Biol. Res. 20, 749–753
Fink, N. E., Caron, M., Joubert, R., Eloila, M. T., Bladier, D., and Herkovits, J. (1987) FEBS Lett. 223, 330–334
Fontana, A., and Gross, E. (1986) in Practical Protein Chemistry. A Handbook (Darbre, A. ed) pp 67–120, J. Wiley & Sons, Chichester, UK
Gitt, M. A., Wiser, M. F., Leffler, H., Herrmann, J., Xia, Y.-R., Massa, S. M., Cooper, D. N. W., Luisa, A. J., and Barondes, S. H. (1995) J. Biol. Chem. 270, 5092–5098
Hadari, Y. R., Paz, K., Dekel, R., Mestrovic, T., Accili, D., and Zick, Y. (1995) J. Biol. Chem. 270, 3447–3453
Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) J. Biol. Chem. 256, 7990–7997
Hirabayashi, J., and Kasai, K. (1993) Glycobiology 3, 297–304
Hirabayashi, J., Kawasaki, H., Suzuki, K., and Kasai, K. (1987) J. Biochem. (Tokyo) 101, 17–24
Hirabayashi, J., Ayaki, H., Soma, G., and Kasai, K. (1989) Biochem. Biophys. Acta 1008, 85–91
Hirabayashi, J., Satoh, M., and Kasai, K. (1992) J. Biol. Chem. 267, 15485–15490
Hirabayashi, J., Ubukata, T., and Kasai, K. (1996) J. Biol. Chem. 271, 2497–2505
Huflejt, M. E., Turck, C. W., Lindstedt, R., Barondes, S. H., and Leffler, H. (1993) J. Biol. Chem. 268, 26712–26718
Kasai, K., and Hirabayashi, J. (1990) J. Biochem. 119, 1–8
Kyte, J., and Dodillte, R. F. (1982) J. Mol. Biol. 157, 105–132
Laemmli, U. K., and Favre, M. (1973) J. Mol. Biol. 80, 575–599
Leffler, H., and Barondes, S. H. (1986) J. Biol. Chem. 261, 10119–10126
Levi, G., and Teichberg, V. I. (1981) J. Biol. Chem. 256, 5735–5740
Liao, D., Kapadia, G., Ahmed, H., Vasta, G. R., and Herzberg, O. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 1428–1432
Lohsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H., and Rini, J. M. (1993) J. Biol. Chem. 268, 27034–27038
Madesen, P., Rasmussen, H. H., Flint, T., Gromov, P., Krusel, T. A., Honore, B., Vorum, H., and Celis, J. E. (1995) J. Biol. Chem. 270, 5823–5829
Marschall, P., Herrmann, J., Leffler, H., Barondes, S. H., and Cooper, D. N. W. (1992) J. Biol. Chem. 267, 12942–12949
Marschall, P., Cannond, V., Barondes, S. H., and Cooper, D. N. W. (1994) Glycobiology 4, 297–305
Miles, N. C., Ma, Y., Varma, P. V., Bering, M. P., Mohamed, Z., Pilaraski, L. M., and Frunchak, Y. N. (1990) Anat. Embryol. 182, 319–327
Muramoto, K., and Kamiya, H. (1992) Biochim. Biophys. Acta 1116, 129–136
Nishihara, T., Wyrick, R. E., Working, P. K., Chen, Y., and Hedrick, J. L. (1986) Biochemistry 25, 6013–6020
Oda, Y., Herrmann, J., Gitt, M. A., Turck, C. W., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) J. Biol. Chem. 268, 5929–5939
Onuki, Y., Matsui, T., Nitta, K., Kawauchi, H., Takayyanagi, Y., and Titani, K. (1991) Biochem. Biophys. Res. Commun. 178, 407–413
Paroutaud, P., Levi, G., Teichberg, V. I., and Stromberg, A. D. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 6345–6348
Pfeifer, K., Hassemann, M., Gamulin, V., Bretting, H., Fahrenholz, F., and Muller, W. E. G. (1993) Glycobiology 3, 179–184
Pohl, J. (1994) Methods Mol. Biol. 36, 167–129
Roberson, M. E., and Armstrong, P. B. (1986) Proc. Natl. Acad. Sci. U. S. A. 77, 3460–3463
Roepstorff, P., and Fohlman, J. (1984) Biomed. Mass Spectrom. 11, 601
Sakakibara, F., Kawauchi, H., Takayyanagi, G., and Ise, H. (1979) Cancer Res. 39, 1347–1352
Sakakibara, Y., Hirabayashi, J., Oda, Y., Ohyama, Y., and Kasai, K. (1990) J. Biol. Chem. 265, 21573–21579
Shet, M. S., and Madaiah, M. (1989) Biochim. Biophys. Acta 991, 465–469
Sparrow, C. P., Leffler, H., and Barondes, S. H. (1987) J. Biol. Chem. 262, 7383–7390
Tracey, B. M., Feizi, T., Abbott, W. M., Carruthers, R. A., Green, B. N., and Lawson, A. M. (1992) J. Biol. Chem. 267, 10342–10347
Vasta, G. R., and Marchalonis, J. J. (1986) J. Biol. Chem. 261, 9182–9186
Vasta, G. R., Hunt, J. C., Marchalonis, J. J., and Fish, W. W. (1986) J. Biol. Chem. 261, 9174–9181
Weller, D., Panneerselvam, C., and Horecker, B. L. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 1947–1949
Wells, V., and Malluci, L. (1991) Cell 64, 91–97