PROTEINS: Structure, Function, and Genetics 40:378–388 (2000)

Soluble β-Galactosyl-Binding Lectin (Galectin) From Toad Ovary: Crystallographic Studies of Two Protein-Sugar Complexes

Mario A. Bianchet,[1] Hafiz Ahmed,[2] Gerardo R. Vasta,[2] and L. Mario Amzel[1]*
[1]*Department of Biophysics and Biophysical Chemistry, Johns Hopkins Medical School, Baltimore, Maryland*
[2]*Center of Marine Biotechnology, University of Maryland Biotechnology Institute, Baltimore, Maryland*

ABSTRACT **Galectin-1, S-type β-galactosyl-binding lectins present in vertebrate and invertebrate species, are dimeric proteins that participate in cellular adhesion, activation, growth regulation, and apoptosis. Two high-resolution crystal structures of *B. arenarum* galectin-1 in complex with two related carbohydrates, LacNAc and TDG, show that the topologically equivalent hydroxyl groups in the two disaccharides exhibit identical patterns of interaction with the protein. Groups that are not equivalent between the two sugars present in the second moiety of the disaccharide, interact differently with the protein, but use the same number and quality of interactions.**

The structures show additional protein-carbohydrate interactions not present in previously reported lectin-lactose complexes. These contacts provide an explanation for the enhanced affinity of galectin-1 for TDG and LacNAc relative to lactose. Galectins are in dimer-monomer equilibrium at physiological protein concentrations, suggesting that this equilibrium may be involved in organ-specific regulation of activity. Comparison of *B. arenarum* with other galectin-1 structures shows that among different galectins there are significant changes in accessible surface area buried upon dimer formation, providing a rationale for the variations observed in the free-energies of dimerization.

The structure of the *B. arenarum* galectin-1 has a large cleft with a strong negative potential that connects the two binding sites at the surface of the protein. Such a striking characteristic suggests that this cleft is probably involved in interactions of the galectin with other intra or extra-cellular proteins. Proteins 2000;40:378–388. © 2000 Wiley-Liss, Inc.

Key words: galectins; X-ray diffraction; thiodigalactose; N-acetyllactosamine; lectin

INTRODUCTION

Galectins-1,[1] S-type β-galactosyl binding lectins, are present in invertebrates as well as in both homeotherm and poikilotherm vertebrates.[2,3] It has been proposed that they participate in cell adhesion, cell activation, growth regulation, cancer metastasis, and apoptosis.[3,4] Although their biological roles are not fully understood, it is known that their expression is developmentally regulated. On the basis of selected features such as their primary structure and subunit architecture, the galectin family has been subdivided in several groups (Galectin-1 through -10). Furthermore, based on conservation of amino acid residues that interact with the carbohydrate ligands (CRD[†]), galectins have been classified in two subgroups: "conserved" (Type I) and "variable" (Type II).[5] This classification not only reflects common features of the carbohydrate specificity but may also reveal evolutionary aspects of their recognition functions. All members of the galectin-1 group characterized so far have conserved (Type I) CRDs, whereas those from the 2, 3, 4, 5, 7, 8, 9, and -10 categories have variable CRDs.[5]

A number of galectins from amphibian species (*Bufo arenarum*, *Rana catesbeiana, X. laevis*, and *Ambystoma mexicanum* subclasses Anura and Urodela) have been isolated and characterized from tissues ranging from embryos and tadpole larvae to adult tissues (including skin, muscle, and gonad).[6–9] Among these, taxa galectins exhibit considerable structural diversity, suggesting that structural and functional divergence in the galectin family may have occurred early in vertebrate evolution. Intriguingly, the primary structure of a toad (B. *arenarum* Hensel) ovary galectin is closely related to the galectin-1 subgroup found in homeotherms: it has 48% sequence identity and it contains three of the six cysteines found in bovine galectin-1. In fact, none of the invertebrate and poikilotherm vertebrate galectins described so far except that of B. *arenarum* share more than one of the galectin-1 conserved cysteines. Moreover, the fine carbohydrate speci-

Abbreviations: LacNAc, N-acetyllactosamine, Gal- β-1,4-GlcNAc; GlcNAc, N-acetylglucose; TDG, thiodigalactose, Gal- β-1-S- β-1-Gal); La, lactose; Gal, galactose; CRD: carbohydrate recognition domain;, DTT, DL-Dithiothreitol; ASA, accessible (by the solvent) surface area.

Grant sponsor: National Institute of General Medical Sciences; Grant number: 1P01GM51362; Grant sponsor: Lucille P. Markey Trust Fund; Grant number 95-31; Grant sponsor: National Science Foundation; Grant number: MCB-94-06649.

*Correspondence to: Dr. L. Mario Amzel, Department of Biophysics and Biophysical Chemistry, Johns Hopkins Medical School, Baltimore, MD 21205. E-mail: mario@neruda.med.jhmi.edu

Received 31 August 1999; Accepted 23 March 2000

[†]The term CRD is wide-spread in this field, although the stretch of aminoacids that includes those interacting directly with the carbohydrate ligand does not constitute a true domain. Certainly "Carbohydrate-binding site" is more appropriate for the galectin family and may be used in the future.

TABLE I. Data Collection Statistics of *Bufo arenarum* Hensel Lectin

Ligand (5 mM)	TDG[a]	LacNAc[b]
Space group	$P2_1$	$P4_12_12$
Lattice	Monoclinic	Tetragonal
Cell (Å)	$a = 51.8\ b = 51.0$ $c = 56.3$	$a = b = 54.3$ $c = 186.5$
β (°)	97.2	90.0
Mosaicity (°)	0.25	0.23
Resolution (Å)	2.16	2.23
Measured refl.	32134	32134
Unique refl. (used)	14540	13192(11873)
Completeness (%)	85	91
R_{sym} (%)[c]	8.6	7.5
Crystal habit	Plates	Rhombohedrons
R value/R_{free}	0.194/0.243	0.183/0.245
Water molecules	45	30
RMSD bonds (Å)	0.011	0.010
RMSD angles (°)	1.75	2.5
RMSD impropers (°)	1.41	1.44

[a]Thiodigalactoside (Gal-β-1-S-β-1-Gal).
[b]N-acetyllactosamine (Gal-β-1,4-GlcNAc).
[c]$R_{sym} = \Sigma h\ \Sigma f_j|If_j - \langle I_h\rangle|/\Sigma_h\ \Sigma f_j|I_{hj}|$, where h represents a unique reflection and j means symmetry equivalent indices. I is the observed intensity and $\langle I\rangle$ is the mean value of I.

ficity and primary structure of *B. arenarum* galectin[10] suggest that it carries a conserved (Type I) CRD.[5] In contrast, a galectin from another amphibian, the clawed frog *X. laevis,* exhibits a variable CRD and a carbohydrate specificity[8] that is different from that of *B. arenarum*.[5] Therefore, the *B. arenarum* galectin appears to be structurally and functionally closer to mammalian members of the galectin-1 group than to the galectin from *X. laevis*, another anuran species.[10]

Galectins preferentially bind N-acetyllactosamine (LacNAc: Gal- β-1,4-GlcNAc) and glycans containing this terminal disaccharide.[11,12] In addition to LacNAc, most galectins also bind thiodigalactoside (TDG: Gal- β-1-S- β-1-Gal), a synthetic disaccharide. The association constants of galectins-1 for LacNAc and TDG are in all the range 1 to 5×10^4 M^{-1} (temperatures of 15 to 28°C),[13] but the different galectin categories show variations in their relative affinities for the two saccharides.[14] While most galectins-1 have higher affinity for TDG than for lactose (Lac), other galectins (e.g., *X. laevis*; galectin-4 from rat intestine, domain I) show the opposite pattern: they have lower affinity for TDG than for Lac.[8,14] The three-dimensional structures of galectins (recently reviewed by Rini[15]) co-crystallized with Lac, a biantennary oligosaccharide, and LacNAc[16–18] have been used to try to rationalize the observed patterns of saccharide affinities. Unfortunately, since no experimental structural information has been available until now for TDG complexes, it was not possible to ascertain whether topologically equivalent hydroxyls of TDG actually interact with the same amino acids of the protein involved in LacNAc binding and whether there are additional interactions involving the sulfur atom.

Understanding all these characteristics would strongly benefit from access to structural information. In this paper, we report the crystal structures of the complexes of the *B. arenarum* galectin with two disacharides, LacNAc and TDG. The complex with LacNAc, a tetragonal crystal, was refined with data to 2.23 Å resolution, while the TDG complex, a monoclinic crystal, was refined with data to 2.16 Å resolution (Table I). The work presented here, in addition to reporting the first structure of an amphibian galectin, provides the first structure of a galectin-TDG complex, allowing to compare and contrast the binding of the synthetic disaccharide TDG with that of the natural ligand LacNAc.

METHODS

Crystallization and Data Collection

Isolation and purification of amphibian 14.8 kDa galectin was carried out using methods described previously.[10] Crystals were obtained in the presence of 5 mM concentrations of either of two ligands, LacNAc and TDG, using vapor diffusion (Ahmed and Vasta, unpublished results). Drops of equal amounts of 10–12 mg/ml protein and reservoir solution were equilibrated against 1 ml of $(NH_4)_2SO_4$ at 56% saturation in 100 mM Tris-acetate buffer, pH 6.6, 1% 2-methyl-2,4-pentanediol (hexylene glycol), 1 mM DTT. In the presence of TDG, the crystals are monoclinic of the space group $P2_1$, while in the presence of LacNAc, a tetragonal form, with space group $P4_12_12$, is obtained (Table I).

Diffraction data were collected at room temperature using an R-Axis II Image Plate System and processed with the R-Axis 3.4 software (Table I).

Structure Determination and Refinement

Initial models were obtained by molecular replacement using the package AMoRe.[19] A monomer of bovine spleen galectin[18] (48% sequence identity; code 1SLT) was used as the search molecule for the first structure (TDG). The orientation and position of each monomer was found independently. Rotation function searches with a single monomer showed two independent peaks 4σ above the next highest peak. The final correlation of the translation search for the two monomers (dimer), after rigid body fitting, was 55% with an initial R value of 47%. The galectin-LacNAc complex searched with the TDG model gave two peaks 3σ above the next higher peak in the rotation function, and a translation search found the position of the molecules with values of the correlation function of 57% and an initial R value of 45%. The space group ambiguity in the LacNAc complex was solved by comparing the correlation values obtained for the two possible space groups ($P4_12_12$ and $P4_32_12$). The initial molecular models for the ligands was built with the program Chemnote and the coordinates were optimized by energy minimization with CHARMM version 2.3 (both included in the molecular modeling program QUANTA; Molecular Simulation, Inc., Palo Alto, CA).

The disaccharide models (positions and conformations) were built by inspection of difference Fourier maps calculated during the initial stages of refinement. The lectin-ligand structural models were refined with the program

XPLOR.[20] Conventional conjugate gradients refinement was followed by a slow cooling simulated annealing protocol with an initial temperature of 3,000 °K, followed by additional cycles of conventional conjugate gradients refinement, constrained B factor refinement, and manual rebuilding using the program O.[21] The atomic coordinates of the LacNAc complex (code 1GAN) and of the TDG complex (code 1A78) were deposited in the Protein Data Bank. Final statistics of the crystallographic refinement are included in Table I. Figures were drawn with the programs SETOR,[22] MOLSCRIPT,[23] and RASTER3D.[24] Differences in the solvent-accessible surface area (ASA) were calculated with the algorithm of Lee and Richards,[25] by using the programs AREAIMOL, DIFFAREA, and SURFACE in the CCP4 suite programs.[26] The probe radius was 1.4 Å and the slice was 0.25 Å.

RESULTS AND DISCUSSION

Description of the Structures

General description

The *B arenarum* galectin has a β-sheet fold with jelly roll topology found in legume lectins[27,28] and in serum amyloid P component,[29] a mammalian pentraxin: a dimeric 22-strand β-sheet sandwich with two sugar binding sites per dimer (Fig. 1).[16–18] As in other galectins, the two monomers are related by a two-fold axis of symmetry perpendicular to the β-sheet plane with both the N and C termini at the dimer interface.[16,18] The sugar binding sites are located in the region of each monomer that is furthest away from the dimer interface (Fig. 1). In both complexes, the refined electron density maps shows well-defined density for all the polypeptide chains and disaccharides (Fig. 2).

The structures of the *B. arenarum* galectin in the two complexes reported here are highly similar: the root-mean-square deviations (RMSD) between Cα atoms positions between monomers is 0.44 Å and between dimers 0.55 Å. In both crystal forms there is one dimer in the asymmetric unit. Thus, the two monomers in a dimer (hereafter monomers A and B) have different crystal contacts such that one (and only one) of the two sugars in a dimer participates in crystal contacts. For consistency, we have labeled as "monomer A" the one binding the sugar that participates in crystal contacts. These distinct crystal contacts produce small differences in the conformations of the two sugars bound to the same dimer (Table II), as well as differences in dimerization angles between the two complexes.

Dimer structure

The galectin dimer is shaped like a dumbbell but with a concave and a convex side. The concave side shows a 57 Å long S-shaped cleft that contains the sugar binding pockets at the ends. Intriguingly, in the *B. arenarum* galectin, this cleft has a large negative potential (Fig. 3A). Residues Ser-2, Thr-8, Asp-27, Asp-51, Asp-55, Asp-67, Glu-72, Glu-124, and Ser-122 contribute to the net negative charge. By contrast, the charge of the same cleft in the bovine spleen galectin-1 is less pronounced (Fig. 3B). However, since the location of surfaces of interaction between proteins tend to be conserved, one might speculate that in other galectins-1 this cleft is also involved in the physiological interactions. The differences in surface properties between the galectins of *B. arenarum* and bovine spleen may reflect different modes of interaction with their putative protein partners.

TABLE II. Carbohydrate Conformation in Each Monomer[†]

Complex	TDG (A/B)	LacNAc (A/B)
C1′-X-C(1 or 4)	98.9/95.5	118.2/113.6
Φ	−59.75/−57.79	−66.9/−67.8
Ψ	133.0/135.1	132.4/132.6

[†]Definition of glycosidic bond torsion angles: Φ, O′-C1′-S-C1 (TDG) or O′-C1′-O4-C4 (LacNAc); Ψ, C1′-S1-C1-C2 (TDG) or C1′-O4-C4-C3 (LacNAc).

Dimer interface

The extended β-sandwich running across the dimer is formed by joining the external strands of both monomer sheets through four additional H-bonds per side. These bonds are formed between the N-terminal strands of the two monomers (on the concave side) and the two C-terminal strands (on the convex side). A two fold axis of symmetry, perpendicular to the plane of the β-sheets, relates amino acids from different monomers that form the interface. On the N-terminal side of the dimer interface, the main chain oxygen and nitrogen atoms of Ala-6, in one monomer, form H-bonds with the corresponding main chain atoms of Thr-8 of the other monomer. Ile-131 satisfies all of its possible main-chain H-bonds with its two fold partner. An H-bond between the nitrogen atom of Thr-133 and the carbonyl oxygen of Lys-129 completes the two fold symmetrical hydrogen bond network of the C-terminal portion of the interface. In its central part, the interface typically buries two hydrophobic residues (Val-7 and Ile-131 in *B. arenarum*, Ile-7, and Ile-131 in bovine galectin-1, and Val-7 and Phe-128 in the human galectin-2). Table III shows the contribution of each residue to the apolar and polar solvent-accessible surface area (ASA) buried (ΔASA) in dimer formation. Val-5, Val-7, and Phe-128 are buried on the inner side of the β-sandwich and make the largest contributions to the apolar ΔASA upon dimer formation (Table III), indicating that they are involved in dimer stabilization. For example, in the case of bovine galectin-1 the single mutation of Val-5 to Asp impairs dimerization.[30,31] In *B. arenarum*, the N and C-termini contribute similarly to apolar ΔASA (302 and 288 Å^2/monomer, respectively), but the N-termini have a larger contribution to the polar ΔASA (302 Å^2 vs. 226 Å^2/monomer). Including the residues not in the termini, the total apolar and polar ΔASAs of dimerization in *B.*

Fig. 1. Dimer of *B. arenarum* lectin-thiodigalactoside complex. Bound sugar and residues participating in the binding are shown.

Fig. 2. Electron density map (2Fo-Fc) around the bound carbohydrate in the monomer A of the dimer. **A:** TDG complex. **B:** LacNAc complex.

Fig. 3. Surface charge of the *B. arenarum* galectin dimer. **A:** *B. arenarum.* **B:** bovine galectin-1. The colors represent electric potential at the surface (blue: positive; red:negative).

Figure 1.

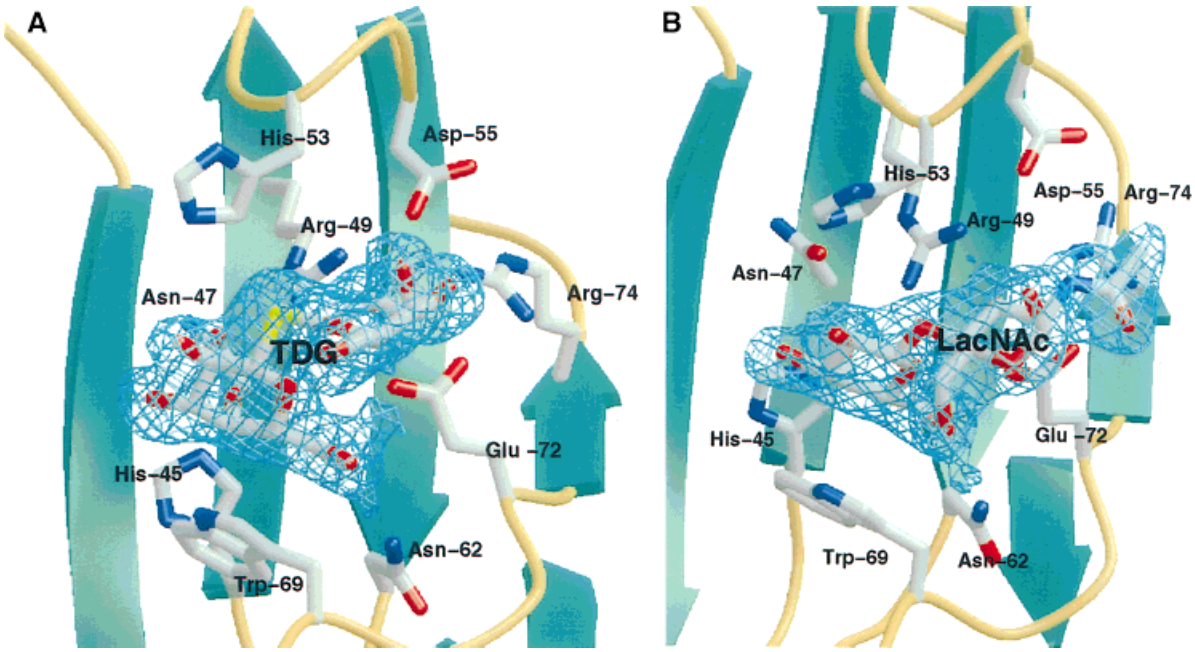

Figure 2.

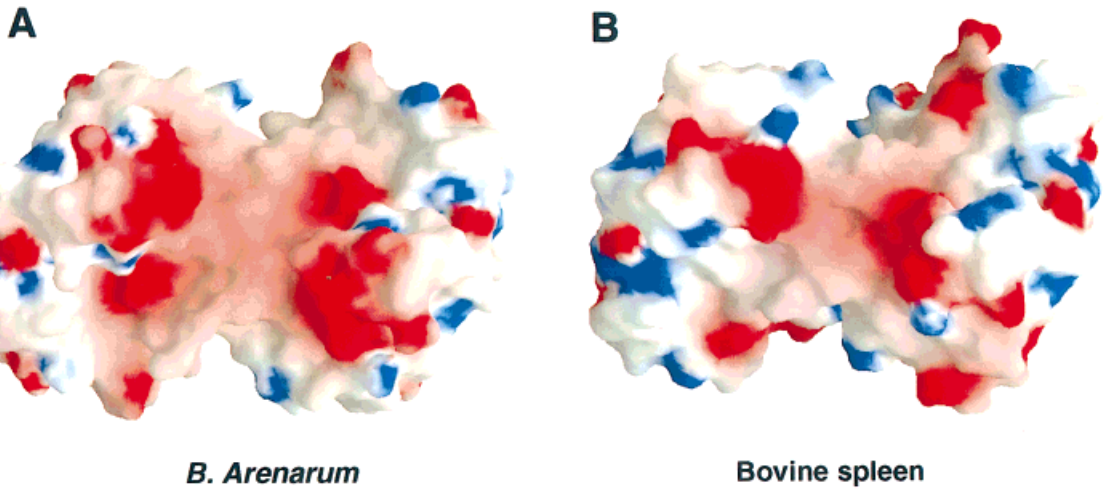

Figure 3.

arenarum are 643 Å^2/dimer and 449 Å^2/dimer, respectively. Bovine galectin-1 buries less apolar and more polar areas upon dimerization (611 and 504 Å^2/dimer, respectively). Human galectin-2 has ΔASAs similar to those of *B. arenarum* galectin (641 Å^2 of apolar area and 464 Å^2 of polar per dimer). These differences in the buried areas are probably responsible for the variations in the association constant for dimer formation among different galectins.

Sugar binding

Tables IV and V summarize the sugar-protein contacts (distances < 3.5 Å). The first hexose is the same in the two disaccharides (Gal 1 of LacNAc and TDG). All contacts involving this residue are conserved in the two complexes (Figs. 4, 5 and Table IV). In each of the complexes, the hydroxyl group at C4′ of Gal 1 interacts with His-45, Arg-49, and Asn-47, whereas the OH on C6′ interacts with Asn-62 and Glu-72. Two residues with planar side chains, Trp-69 and His-53, provide contacts that help align the sugar. Trp-69 participates in a stacking interaction with galactose ring carbons C3′, C4′, and C5′ on one face of the galactose ring, and restricts the orientation of the OH on C4′ to the axial form. His-53 provides a limited number of van der Waals contacts with C2′ and O2′ on the other side of the galactose ring and with the bridging atom (O4 in LacNAC or S1 in TDG). Arg-49, Asp-55, Glu-72, and Arg-74 interact with each other through a network of hydrogen bonds that holds them in position (Fig. 4, 5) to bind the second sugar of the disaccharide.[16–18] Stabilization of this sugar is primarily achieved by the interaction of an equatorial hydroxyl (C2-OH in the second galactose moiety of the TDG or C3-OH in the GlcNAc moiety of the LacNAc) with Arg-49 and Glu-72. GlcNAc is further stabilized by a water molecule that bridges the nitrogen of the NAc group with O_δ of Asp-55, the main chain carbonyl of His-53 and Nη of Arg-74. An equivalent water molecule is found in TDG (undisturbed monomer B) mediating the interactions of the axial C3-OH with Nη of Arg-74, Oη of Asp-55, and the main-chain carbonyl of His-53. This water molecule is present in all lectin-disaccharide complexes studied to date.[15,16] The guanidinium group of Arg-74 also plays a role in binding the second sugar in the TDG complex by making a hydrogen bond with the C3-OH**.** The S-atom in TDG makes contacts with His-53 and Arg-49 (Table V) similar to those made by the bridging oxygen in the LacNAc complex.

In summary, the first Gal moiety, shared by both disaccharide ligands, interacts with the protein in a conserved way. The second moiety (GlcNAc in LacNAc and another Gal in TDG) utilizes different contacts in the TDG and in the LacNAc complex, although the same number and quality of H-bonds are present in both cases (two direct and one water mediated to the protein). The apolar and polar accessible surface areas buried per dimer upon binding of ligand in the LacNAc complex are 232 Å^2 and 255 Å^2, respectively. Ligand binding in the TDG complex results in less apolar (180 Å^2) and almost identical polar (256 Å^2) buried ASA. One might expect that the galectin would have lower affinity for TDG because this complex buries less apolar area. However, TDG and LacNAc containing carbohydrates bind the *B. arenarum* lectin with similar affinity.[10] The reason for this is probably that although LacNAc buries 52 Å^2 more of apolar ASA, the rotation of the acetyl group is frozen upon binding, resulting in a compensating loss of conformational entropy.

Cysteine residues: oxidation states

Galectins usually require addition of thiols or other reducing agents to retain in vitro carbohydrate-binding activity (this characteristic suggested the family's earlier name of S-lectins). These reagents can have many effects, including preventing covalent oligomerization of the protein. Cysteine residues might be considered good candidates to be part of an oxidation/reduction-dependent binding activity. Alignment of the sequences of known lectins shows that all galectins-1 characterized in homeotherm vertebrates share six cysteine residues. Of these Cys-16, Cys-61, and Cys-89 are also present in *B. arenarum* galectin. No cysteines are conserved in any other galectin from poikilotherm vertebrates and invertebrates described so far. In the *B. arenarum* galectin, Cys-61 is located in a highly conserved portion of the CRD structure, the inner part of the β-sandwich, well buried in the core of the protein. In contrast, Cys-16 and Cys-89, located on the surface of the molecule on the face opposite to the CRD, are relatively exposed, surrounded by long side chain hydrophilic residues such as Lys-100, Glu-19, and Glu-91. It is noteworthy that in the crystal structures presented here, cysteines 16 and 89 appear to be in various degrees of oxidation. In monomer A of the lectin-TDG complex an additional horseshoe-shaped density bridging Cys-16 and Cys-89 was interpreted as a covalently bound DTT molecule. Additional electron density found close to the putative sulfur atoms was assumed to show as sulfur bound oxygen atoms that were not included in the refinement. In the LacNAc complex, electron density was found bridging cysteines in two consecutive β-strands. This density was modeled as an alternative conformation making a disulfide bridge between Cys-16 and Cys-89. To our knowledge, this is the first observation of mixed oxidation states in galectins.

In bovine galectin-1, intramolecular disulfide formation between Cys-2 and Cys-130, Cys-16 and Cys-88 (corresponding to the pair Cys-16, Cys-89 in *B arenarum*), and Cys-42 and Cys-60 (Cys-61 in *B. arenarum*)[32] has been shown to be associated with the oxidative inactivation of carbohydrate-binding. In the bovine galectin-LacNAc crystal, the disulfide bond between Cys-42 and Cys-60 was found to be reduced and buried, while the thiol groups of Cys-16, Cys-88, and Cys-130 were oxidized and that of Cys-2 was disordered.[18] None of these cysteines is in direct contact with the ligand. Similarly, in the complexes of the *B. arenarum* galectin reported here, none of the ligands is in contact with any of the three cysteines (Cys-16, Cys-61, and Cys-89). Since Cys-16 and Cys-89 are found in various states of oxidation in the crystals, while the galectin binding site is fully occupied (average B factor 30.3 Å^2), the oxidative states of Cys-16 and Cys-89 do not appear to have an effect on the activity of the protein. Thus, Cys-61 (aligned to Cys-60 of bovine spleen galectin-1) is a likely

TABLE III. Contribution to the Solvent-Accessible Surface Area (ASA) and ASA Buried (ΔASA) of Each Residue of the Dimer Interface in *Bufo arenarum* Galectin-1

	ASA monomer		ASA dimer		ΔASA buried	
	Apolar	Polar	Apolar	Polar	Apolar	Polar
Ala-3	69.4	25.1	58.7	5.0	10.7	20.1
Gly-4	4.5	10.2	3.0	6.2	1.5	4.0
Val-5	49.7	1.6	6.9	1.6	42.8	0.0
Ala-6	32.0	37.8	19.3	0.0	12.8	37.8
Val-7	42.7	0.0	4.1	0.0	38.6	0.0
Thr-8	49.6	42.8	39.6	4.2	10.0	38.6
Asn-9	28.3	86.4	6.3	51.5	21.9	64.4
Leu-10	10.2	13.8	0.0	2.7	10.2	11.1
Phe-128	65.1	22.4	9.6	16.9	42.7	5.5
Lys-129	59.8	67.1	37.6	29.4	22.2	37.7
Ser-130	25.6	4.3	1.2	3.8	24.3	0.5
Ile-131	30.3	25.0	4.1	0.0	26.2	25.0
Thr-132	46.0	14.9	17.1	14.9	28.8	0.0
Thr-133	10.9	57.9	5.9	13.2	5.0	44.7
Total[a]	7105	5876	6462	5428	643.0	448

[a]Total ASA buried upon dimerization from two monomers to one dimer. Solvation waters were excluded in the calculations.

TABLE IV. Carbohydrate-Protein Short Contacts (<3.5 Å) of the Non-Reducing Gal Moiety (Common to Both Disaccharides)[†]

Sugar atom	Protein atom	Distances for each monomer (Å) in TDG (A/B)	LacNAc (A/B)
C2′-OH	$N_{\varepsilon 2}$ His-53	3.34/3.56	3.73/3.33
	$C_{\xi 1}$ His-53		3.89/3.34
	$O_{\delta 1}$ Asn-57′[a]	3.19/—	
C3′-OH	O W1	2.77/2.70	2.97/2.68
C4′	$N_{\varepsilon 2}$ His-45	3.44/3.30	3.26/3.33
C4′-OH	$N_{\varepsilon 2}$ His-45	2.97/2.77	2.69/2.83
	$O_{\delta 1}$ Asn-47	3.15/3.23	3.36/2.91
	$N_{\eta 2}$ Arg-49	2.82/2.81	2.81/2.73
	O W1	3.05/3.10	3.12/3.10
C6′	$O_{\varepsilon 2}$ Glu-72	3.18/	3.48/3.60
C6′-OH	$N_{\delta 2}$ Asn-62	2.80/2.86	2.72/2.77
	C_{β} Asn-62	3.63/3.61	3.29/3.49
	C_{γ} Asn-62	3.67/3.70	4.56/4.46
	$O_{\varepsilon 2}$ Glu-72	2.63/2.88	2.88/2.89
	C_{δ} Glu-72	3.40/3.49	3.48/3.39
	C_{γ} Glu-72	3.18/3.28	3.14/3.27
O′	$N_{\eta 2}$ Arg-49	3.04/3.12	2.91/2.99
	$O_{\varepsilon 2}$ Glu-72	3.26/3.49	3.18/3.25

[†]Longer distances are included for comparison.
[a]Superscript prime in the second column indicates amino acids from symmetry related molecules.

TABLE V. Carbohydrate-Protein Short Contacts (<3.5 Å) of the Second Moiety (Gal in TDG and GlcNAc in LacNAc)[†]

Sugar atom	Protein atom[b]	Distances for each monomer (Å) in TDG (A/B)	LacNAc (A/B)
S1/O4[a]	$N_{\eta 2}$ Arg-49	3.34/3.55	3.54/3.58
	$C_{\delta 2}$ His-53	3.33/3.50	
C1/4-OH	$O_{\delta 1}$ Asn-95′		2.81/—
C2/3	$O_{\varepsilon 2}$ Glu-72	3.32/3.27	
C2/3-OH	$N_{\eta 1}$ Arg-49	2.80/2.82	2.86/2.91
	$N_{\eta 2}$ Arg-49	3.25/3.33	—/3.27
	$O_{\varepsilon 1}$ Glu-72	3.40/3.43	
	$O_{\varepsilon 2}$ Glu-72	2.56/2.51	2.47/2.40
	$N_{\eta 2}$ Arg-74	3.23/3.32	3.47/3.57
	C_{ξ} Arg-49		3.48/3.52
C3/2	$O_{\varepsilon 2}$ Glu-72	3.39/3.38	
C3-OH/N_2	O W2	—/2.69	2.88/2.92
	$N_{\eta 2}$ Arg-74	3.26/3.34	3.56/3.37
C4/1-OH	NH Gln-82′	2.92/—	
	O W3	—/2.69	
C8[c]	O_{γ} Ser-108′		3.13/—

[†]The topologically related atoms are in the same row. The atoms or positions correspond to the TDG and LacNAc, respectively. Longer distances are included for comparison.
[a]Bridging atom.
[b]Superscript prime indicates aminoacids from symmetry related molecules.
[c]Atom of the LacNAc's N-acetyl moiety.

candidate to be associated with the oxidative inactivation of the galectin, although the mechanism for this effect remains unclear because no proximal cysteine (equivalent to Cys-42 of bovine spleen galectin-1) is present as a possible disulfide bond partner. The galectin from the electric eel has no cysteines and its oxidative inactivation has been attributed to the oxidation of a tryptophan residue to an oxindole.[32] It is interesting that the 14 kDa β-galactoside-binding lectin from eggs of *R. catesbeiana* does not require a reducing agent for preservation of its binding activity; a structural basis for this observation has not been proposed.[7] An answer to the question of whether oxidation of the thiol group of Cys-61 and/or the indole of Trp-69 are responsible for the progressive loss of carbohydrate-binding activity of *B. arenarum* galectin under non-reducing conditions[10] and whether disulfide formation between Cys-42 and Cys-60 has this effect in mammalian

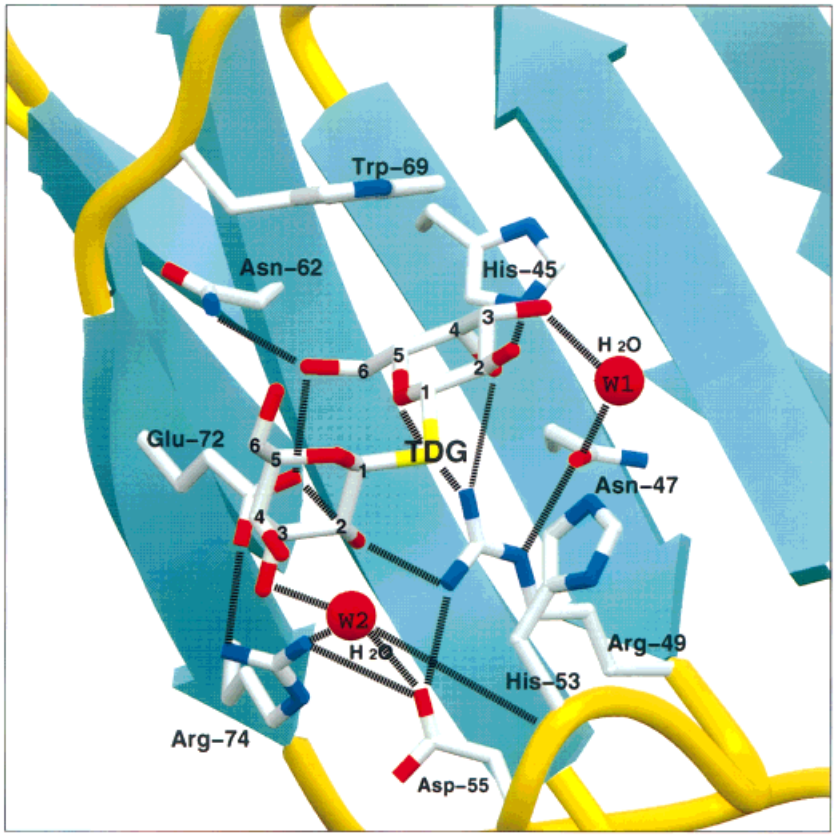

Fig. 4. Thiogalactosamine bound to *B. arenarum* lectin (monomer B). The relevant CRD residues and the hydrogen bonds (dashed lines) to the sugar are shown. Carbon atoms are in white, oxygen in red; sulfur in yellow; nitrogen in blue.

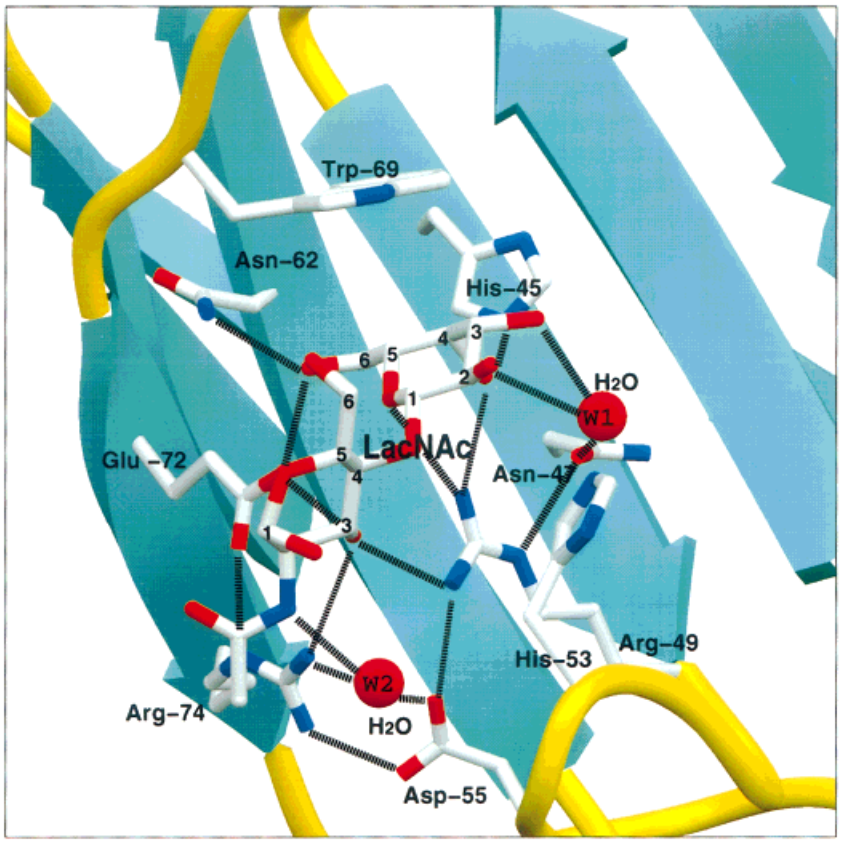

Fig. 5. N-acetyllactosamine bound to *B. arenarum* lectin (monomer B). The relevant CRD residues and the hydrogen bonds (dashed lines) to the sugar are shown. The atoms are colored as in Figure 4.

```
                                                                        |-----
Toad            ASAGVAVTNLNLKPGHCVEIKGSIPPDCKGFAVNLGEDASNFLLHFNARF
                |   ||  ||||||| |    |    | | |  ||| |  |  |||| ||
Bovine H/S      ACGLVA_SNLNLKPGECLRVRGEVAADAKSFLLNLGKDDNNLCLHFNPRF

                ---C R D---------------|
Toad            DLHGDVNKIVCNSKEADAWGSEQREEVFPFQQGAEVMVCFEY_QTDKIII
                  |||||  |||||| | ||| ||||  |||| |  | ||    |||   |
Bovine H/S   51 NAHGDVNTIVCNSKDAGAWGAEQRESAFPFQPGSVVEVCISFNQTD_LTI

Toad            KFSSGDQFSFPVRKVLPSIPFLSLEGLQFK__SITTE
                |  |  | || |  |  |   ||  |  ||      |
Bovine H/S  101 KLPDGYEFKFPNRLNLEAINYLSAGGD_FKIKCVAFE
```

Fig. 6. Sequence comparison between toad ovary and bovine (Spleen) galectins-1.

```
                                        45 4749 53 55     62     69 72 74
                                         | | |   | |      |      |  | |
Conserved CRD   Toad ovary              HFNARFDLHGDVNKIVCNSKEADAWGSEQR
                Human14-1               HFNPRFNAHGDANTIVCNSKDGGAWGTEQR
                Bovine H/S              HFNPRFNAHGDVNTIVCNSKDAGAWGAEQR
                Rat14                   HFNPRFNAHGDANTIVCNSKDDGTWGTEHR
                Mouse14                 HFNPRFNAHGDANTIVCNTKEDGTWGTEQR

                Chick14                 HFNPRFDAHGDVNLIVCNSKKMEEWGTEQR
                Chick16                 HFNPRFDCHGDVNTVVCNSKEDGTWGEEDR

Variable CRD    Human14-2               HFNPRFS---E-STIVCNSLDGSNWGAEQR
                Rat int36-I             HFNPRFDGW-D--KVVFNTMQSGQWGKEEK

                X. laevis               HCNPRFEYSVDKNTIVLNSKQNDVWDIEKK
```

Fig. 7. CRDs of several galectins. The numbering corresponds to that of the toad ovary sequence.[10] Human14-1;[40] Bovine H/S: (heart or spleen);[32] Rat14;[34] Mouse14;[35] Chick14;[36] Chick16;[37] Human14-2;[41] Rat int36-I;[14] *X. laevis*.[8]

galectins-1 must await the determination of the structures of the inactive forms.

Packing Interactions

The common characteristic of the two crystals presented here is the participation of one of the two sugars of every dimer in crystal contacts. In all galectin crystal forms determined to date, bound disaccharides participate in crystal contacts, suggesting that these contacts may be instrumental for crystal formation in galectins: in the *B. arenarum* lectin-LacNAc complex, the sugar makes an H-bond with Glu-95 of a symmetry-related molecule (Table V). This interaction locks the axial conformation for the C1-OH in the bound sugar. In contrast, the C1-OH of the monomer B, which does not participate in crystal contacts, shows a less well-defined density that indicates the presence of both anomers. The second Gal moiety in monomer A of the lectin-TDG complex participates in packing interactions through two of its solvent exposed hydroxyl groups: C2-OH interacts with Asn-57, and C4-OH makes an H-bond with the main chain NH of Gln-82 of a symmetry-related molecule. Also, C4-OH participates in packing through a water mediated interaction.

Comparison to Other Animal Galectins and Their CRDs

The primary sequences[10] of the toad ovary galectin and the bovine spleen galectin-1 show 48% identity (Fig. 6). Alignment of their three dimensional structures shows that for 132 Cα atoms, the RMSD between main chain of bovine spleen galectin-116 and the TDG and LacNAc complexes are 0.86, 0.92, and 0.87 Å, respectively. Superposition of the alpha carbons of the monomers of the toad ovary galectin with the other previously reported animal lectin structures (codes 1HLC, 1SLA, and 1SLT) shows RMSD of 0.95, 0.76, and 0.80 Å, respectively, extending the similarity of toad ovary to all animal galectins.

The amino acid residues corresponding to the CRD are strongly conserved among several species (Fig. 7): His-45, Asn-47, Arg-49, His-53, Asp-55, Asn-62, Trp-69, Glu-72, and Arg-74, which participate in the binding of ligands, are invariant among galectins carrying "conserved" CRDs,[5] such as those from toad ovary, human (type 1), bovine spleen,[33] rat14,[34] mouse14,[35] chick14,[36] and chick16.[37] These residues line the disaccharide binding grove and their spatial arrangement is strongly conserved as well: the relevant residues occupying highly similar positions in the galectin structures described so far (Fig. 7).

The fine carbohydrate-binding specificities of all galectins-1 tested so far have been found to be in the order LacNAc ≈ TDG > Lac (Table VI). On the other hand, other galectins such as those of rat intestine 36-I,[14] *X. laevis*, and human galectin-2, which carry "variable" CRDs, have different specificity. In these galectins, His-45, Asn-47, Arg-49, Asn-62, Trp-69, and Glu-72 are also invariant, but His-53, Asp-55, and Arg-74 are either substituted by

TABLE VI. Carbohydrate-Binding Specificities of Galectins With Conserved and Variable Domains

	Relative activity[a]					
Sugars	Toad ovary	Human 14-1	Bovine spleen	Rat14	Rat intestine 36-I	*X. laevis*
Galβ1,4Glc	1	1	1	1	1	1
Galβ1,4GlcNAc	3.8	7.8	5.5	5	0.2	0.5
Galβ1Sβ1Gal	3.9	3.5	5.9	4.0	0.8	0.2

[a]Data were taken from previous reports as follows: toad ovary;[10] human 14-I;[42] bovine spleen;[43] rat14;[11] rat intestine 36-I;[14] *X. laevis.*[8]

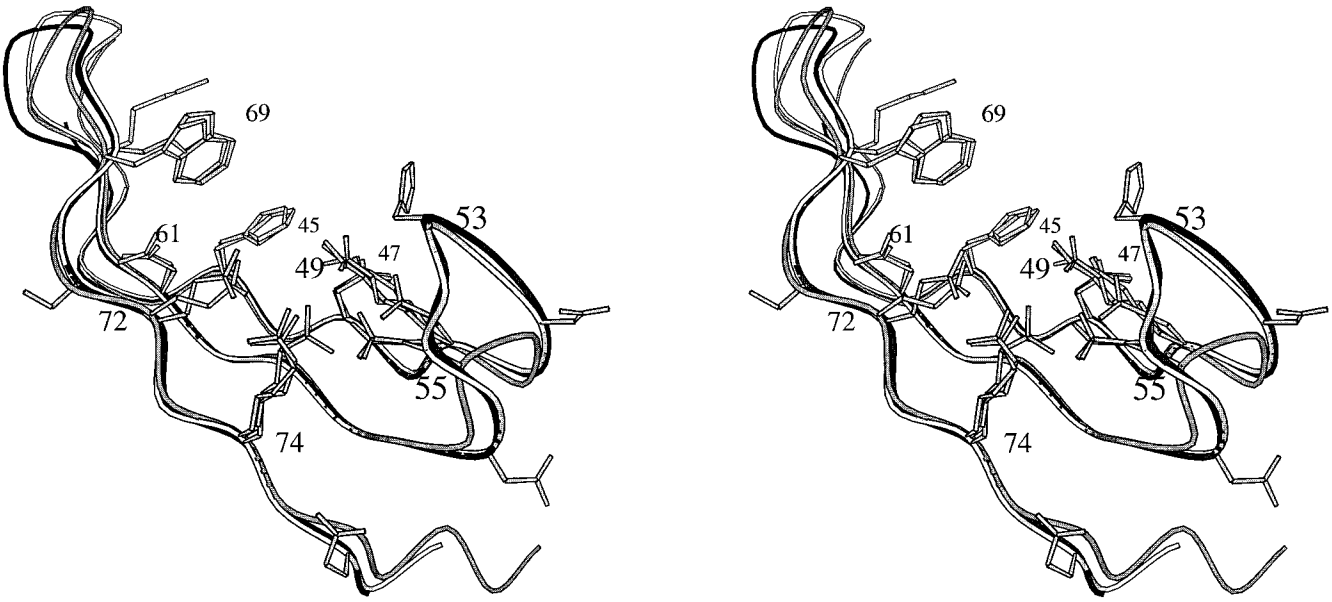

Fig. 8. Stereo view showing the spatial superposition of the structure of the CRDs of *Bufo arenarum* (black) with those of animal lectins of known structures (bovine galectin-1, light gray; human galectin-2, medium gray).

different amino acids or His-53, Asp-55, or both are deleted. Binding of LacNAc to rat intestine 36-I is approximately five-fold less inhibitory than Lac, whereas TDG exhibits inhibitory activity similar to Lac.[14] For the *X. laevis* galectin, LacNAc and TDG are two and five-fold less effective than Lac.[8] In human galectin-2, for which the three-dimensional structure has been reported,[16] sequence deletions (Fig. 7) seem to shorten the His-53 spanning loop but Glu-52 (human galectin-2 numbering), another acidic group, is positioned to play the role of Asp-55 in the hydrogen bonding network of the binding pocket that binds the equatorial OH (Fig. 8). In *X. laevis* His-53 and Arg-74 are replaced by Ser and Lys, and in rat intestine 36-I, these two amino acids are replaced by Trp and Lys. These proteins have, in addition, a deletion of three amino acids between residues 54 and 57 in the CRD. In the structure of the *B. arenarum* galectin reported here, Arg-74 is involved in a direct interaction with C3-OH of GlcNAc (in LacNAc) or C2-OH of the second Gal (in TDG) and in a water-mediated interaction with the nitrogen of the NAc group of GlcNAc (in LacNAc) or C3-OH of the second Gal (in TDG). The main chain carbonyl of His-53 interacts through a water molecule with the N2 of GlcNAc (in LacNAc) or the C3-OH atom of TDG. The interactions of this structural water (Water 2, Fig. IV and V) with Arg-74 and His-53 could provide a direct explanation for the enhanced binding of LacNAc and TDG by lectins with a conserved CRD (type I). Replacement of His-53 and Arg-74 by different amino acids in *X. laevis* and rat intestine 36-I (type II lectins) results in reduced affinity towards LacNAc and TDG than to Lac (Table VI). These observations shed additional light on the evolution of galectins. As mentioned above, *B. arenarum* galectin is more similar to type I mammalian galectins than to the galectin from *X. laevis* (type II). The residues present at positions 53 and 74 seem to correlate with the fine specificity of the lectins: His and Arg at these positions results in high affinity for LacNAc and TDG as is the case of *B. arenarum* galectin. This seems to indicate that both the early evolutionary divergence of *B. arenarum* and *X. laevis* and the conservation of sequence between *B. arenarum* and mammalian galectins reflect a functional differentiation based on the fine specificity of the lectins.

SUMMARY AND CONCLUSIONS

Galectins, previously known as S-type or S-lac lectins, are thought to participate in cell-cell and cell-extracellular matrix interactions that mediate important biological

processes such as embryogenesis, myogenesis, host defense, inflammation, apoptosis, and tumor metastasis.[3] However, the exact function of a particular galectin is probably related to its tissue distribution and ligand availability. For example, the galectin from the toad *B. arenarum*, which is expressed in oocytes and post-fertilization stages,[38] may be involved in developmental processes. In contrast, the galectins from another amphibian, the clawed frog *Xenopus laevis*, which is thought to participate in host defense mechanisms, is mainly confined to adult skin and has not been detected in embryo.[8] Also, although the bovine galectin-1 and the galectin from *B arenarum* are highly similar in their binding and structural properties, they show important functional differences. In mammals, galectin-1, which can be detected in the trophectoderm of the blastocyst just before its implantation in the uterine wall, is believed to play a role in embryo implantation by binding lacto-N-fucopentaose I.[39] In contrast, the *B. arenarum* galectin can be detected in every stage prior to blastula.[38] Therefore, despite their striking similarities in carbohydrate specificity, the localization of *B. arenarum* galectin in the developing embryo suggests that its function(s) may be different from those proposed for mammalian galectins. These characteristics of the *B. arenarum* galectin make it a suitable model for the elucidation of the biological function(s) of galectins in the embryogenesis of poikilotherm vertebrates.

It has been suggested that control of the steady-state concentration of galectins in individual organs and tissues affects the monomer-dimer equilibrium and provides an additional mechanism for regulating galectin activity. The structures of *B. arenarum* galectin show that positions 5, 6, and 128 have the largest contributions to the apolar surface area buried upon dimer formation, and therefore, their identities determine the dimerization affinity.

Although all galectins bind lactose, there are subtle differences in carbohydrate specificity among galectin family members. The correlation of these differences in carbohydrate specificity with the amino acid sequences of the carbohydrate recognition domains (CRDs) suggested the classification of galectins into two types: "conserved" (Type I) and "variable" (Type II).[5] This classification may reflect not only common features of their carbohydrate specificity but possibly evolutionary aspects of their recognition functions. Of the nine residues involved in direct carbohydrate contact, residues 53, 55, and 74 (*B. arenarum* numbering) carry the determinants of fine specificity. *B. arenarum* and other type I galectins have histidine, aspartate, and arginine at these positions. Type II galectins, which have different fine specificity, show substitutions as well as deletions at the same positions. In the structure of *B. arenarum* reported here, the N2 of GlcNAc or the C3 OH of TDG interact through a water molecule with His-53 and Arg-74. These interactions provide the rationale for the enhanced specificity of type I galectins for GlcNAc and TDG compared to those of type II. This finding suggests that the divergence of the lectins of *B. arenarum* (type I) and *X. laevis* (type II), as well as the conservation between *B. arenarum* and mammalian galectins, reflect a functional differentiation based on their fine carbohydrate specificities. In addition, consideration of their specificities, as well as analysis of sequence and structural information, suggests that the structural and functional divergence of the two amphibian galectins may have occurred early in vertebrate evolution.

Analysis of the structures shows that in the *B. arenarum* galectin, the large superficial concavity that connects the two carbohydrate binding sites, has a large net negative charge. This striking property suggests that the cleft may be functionally important, probably for interaction with other cellular components. The distinctive feature of the cleft in the *B. arenarum* galectin (a strong negative potential) may have identified a region that is also important in other galectins.

ACKNOWLEDGMENTS

We thank Dr. S. Yuhasz for critical reviewing of the manuscript, and the Biomedical Supercomputing Center of the Frederick Cancer Research and Development Center (Frederick, MD). Grant 1P01GM51362 was provided from the National Institute of General Medical Sciences to L.M.A., and Grant 95-31 from the Lucille P. Markey Trust Fund and Grant MCB-94-06649 from the National Science Foundation to G.R.V.

REFERENCES

1. Barondes SH et al. Galectins: a family of animal β-galactoside-binding lectins. Cell 1994;76:597–598.
2. Hirabayashi J, Kasai K. The family of metazoan metal-independent β-galactoside-binding lectins: structure, function and molecular evolution. Glycobiology 1993;3:297–304.
3. Barondes SH, Cooper DNW, Gitt MA, Leffler H. Galectins: structure and function of a large family of animal lectins. J Biol Chem 1994;269:20807–20810.
4. Perillo NL, Pace KE, Seilerhamer JJ, Baum LG. Apoptosis of T-cells mediated by galectin1 Nature 1995;378:736–739.
5. Ahmed H, Vasta GR. Galectins: conservation of functionally and structurally relevant amino acid residues defines two types of carbohydrate recognition domains Glycobiology 1994;4:545 –549.
6. Fink de Cabutti NE, Caron M, Joubert R, Elola MT, Bladier D, Herkovitz J. Purification and some characteristics of a beta-galactoside binding soluble lectin from amphibian ovary. FEBS Lett 1987;223:330–334.
7. Ozeki Y, Matsui T, Nitta K, Kawauchi H, Takayanagi Y, Titani K.. Purification and characterization of beta-galactoside binding lectin from frog (*Rana catesbiana*) eggs. Biochem Biophys Res Commun 1991;178:407–413.
8. Marschal P, Hermann J, Leffler H, Barondes SH, Cooper DNW. Sequence and specificity of a soluble lactose-binding lectin from *Xenopus laevis* skin. J Biol Chem 1992;267:12942–12949.
9. Allen HJ, Ahmed H, Sharma A. Isolation of lactose-binding lectins from axolotl (*Ambystoma mexicanum*). Comp Biochem Physiol 1992;103B:313–315.
10. Ahmed H, Pohl J, Fink NE, Strobel F, Vasta GR. The primary structure and fine 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 clawed frog *Xenopus laevis*. J Biol Chem 1996;271: 33083–33094.
11. Leffler H, Barondes SH. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian β-galactosides. J Biol Chem 1986;261:10119–10126.
12. Ahmed H, Allen HJ, Sharma A & Matta KL. Human splenic galaptin: carbohydrate-binding specificity and characterization of the combining site. Biochemistry 1990;29:5315–5319.
13. Ramkumar R, Surolia A, Podder SK. Energetics of carbohydrate binding by a 14 kDa S-type mammalian lectin. Biochem J 1995;308:237–241

14. Oda Y, Herrmann J, Gitt MA, Turck CW, Burlingame AL, Barondes SH, Leffler H. Soluble lactose-binding lectin from rat intestine with two different carbohydrate-binding domains in the same peptide chain. J Biol Chem 1993;268:5929–5939.
15. Rini JM. Lectin structure. Annu Rev Biophy Biomol Struct, 1995;24:551–577.
16. Lobsanov YD, Gitt MA, Leffler H, Barondes SH, Rini JM. X-ray crystal structure of the human dimeric S-lac Lectin. L-14-II, in complex with lactose at 2.9-Å resolution. J Biol Chem 1993;268: 27034–27038.
17. Bourne Y, Bolgiano B, Liao DI, Strecker G, Cantau P, Herzberg O, Feizi T, Cambillau C. Crosslinking of mammalian lectin (galectin1) by complex biantennary saccharides. Nature Struct Biol 1994;1:863–870.
18. Liao DI, Kapadia G, Ahmed H, Vasta GR, Herzberg O. Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. Proc Natl Acad Sci USA 1994;91:1428–1432.
19. Navaza J. AMoRe: an automated package for molecular replacement. Acta Cryst 1994;A50:157–163.
20. Bruenger AT. X-PLOR: a System for crystallography and NMR, Version 3.1 Manual. New Haven: Yale University Press; 1992.
21. Jones TA, Zou JY, Cowan SW, Kjelgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst 1991;A47:110–119.
22. Evans SV. SETOR: Hardware lighted three-dimensional solid model representation of macromolecules. J Mol Graphics 1993;11: 134–138.
23. Kraulis P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J App Crystallogr 1991;24: 946–950.
24. Merritt EA, Murphy MEP. Raster3D version 2.0: a program for photorealistic molecular graphics. Acta Cryst D50:869–873.
25. Lee B, Richards FM. The interpretation of protein structures: estimation of static stability. J Mol Biol 1971;55:374–400.
26. The CCP4 suite: Programs for Protein Crystallography. Acta Crystallogr 1994;D50:760–763.
27. Becker JW, Reeke GN, Wang JL, Cunningham BA, Edelman GM. The covalent and three-dimensional structure of concanavalin A. J Biol Chem 1975;250:1513–1524.
28. Reeke GN, Becker JW, Edelman GM. The covalent and three-dimensional structure of concanavalin A. J Biol Chem 1975;250: 1524–1547.
29. Emsley J et al. Structure of pentameric human serum amyloid P component. Nature 1994;367:338–345.
30. Cho M, Cummings RD. Characterization of monomeric forms of Galectin-1 generated by site-directed mutagenesis. Biochemistry 1996;35:12801–13088.
31. Cho M, Cummings RD. Galectin-1: oligomeric structure and interactions with polylactosamine. Trends Glycobiol GlycoTech 1997;9:47–66.
32. Tracey BM, Feizi T, Abbott WM, Carruthers RA, Green BN, Lawson AM. Subunit molecular mass assignment of 14, 654 Da to the soluble β-galactoside-binding lectin from bovine heart muscle and demonstration of intramolecular disulfide bonding associated with oxidative inactivation. J Biol Chem 1992;267: 10342–10347.
33. Abbott WM, Mellor A, Edwards Y, Feizi T. Soluble bovine galactoside-binding lectin cDNA reveals the complete amino acid sequence and an antigenic relationship with the major encephalitogenic domain of myelin basic protein. Biochem J 1989;259:283–290.
34. Clerch LB, Whitney P, Hass M, Brew K, Miller T, Werner R, Massaro D. Sequence of a full-length cDNA for rat lung β-galactoside-binding protein: primary and secondary structure of the lectin. Biochemistry 1988;27:692–699.
35. Wells V, Malluci L. Identification of an autocrine negative growth factor: mouse β-galactoside-binding protein is a cytostatic factor and cell growth regulator. Cell, 1991;64:91–97.
36. Hirabayashi J, Kawasaki H, Suzuki K, Kasai K. Complete amino acid sequence of 14 kDa β-galactoside-binding lectin of chick embryo. J Biochem (Tokyo) 1987;101:775–787.
37. Sakakura Y, Hirabayashi J, Oda Y, Ohyama Y, Kasai K. Structure of chicken 16 kDa β-galactoside-binding lectin: complete amino acid sequence, cloning of cDNA, and production. J Biol Chem 1990;265:21573–21579.
38. Elola MT, Fink NE, Herkovits J. A developmentally regulated lectin in *Bufo arenarum* embryos. Brazilian J Med Biol Res 1987;20:749–753.
39. Colnot C, Ripoche M A, Scaerou D, Fowlist D, Poirier F. Galectins in mouse embryogenesis. Biochem Soc Trans 1996;24:141–146.
40. Hirabayashi J, Ayaki H, Soma G, Kasai K. Cloning and nucleotide sequence of a full length cDNA for human 14 kDa β-galactoside-binding lectin. Biochim Biophys Acta 1989;1008:85–91.
41. Gitt MA, Massa SM, Leffler H, Barondes SH. Isolation and expression of a gene encoding L-14 II, a new human soluble lactose-binding lectin. J Biol Chem 1992;267:10601–10606.
42. Sparrow CP, Leffler H, Barondes SH. Multiple soluble β-galctoside-binding lectin from human lung. J Biol Chem 1987;262:7383–7390.
43. Ahmed H, Fink NE, Pohl J, Vasta GR. Galectin-1 from bovine spleen: biochemical characterization, carbohydrate specificity and tissue-specific isoform profiles. J Biochem 1996;120:1007–1019.