Crystal Structure of the Galectin-9 N-terminal Carbohydrate Recognition Domain from *Mus musculus* Reveals the Basic Mechanism of Carbohydrate Recognition*§

Masamichi Nagae, Nozomu Nishi, Takeomi Murata, Taichi Usui, Takanori Nakamura, Soichi Wakatsuki, and Ryuichi Kato

From the Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization, Tsukuba, Ibaraki 305-0801, Department of Endocrinology, Faculty of Medicine, Kagawa University, 1750-1 Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0793, and the Department of Applied Biological Chemistry, Faculty of Agriculture, Shizuoka University, 836 Ohyama, Suruga-ku, Shizuoka 422-8529, Japan

The galectins are a family of β-galactoside-binding animal lectins with a conserved carbohydrate recognition domain (CRD). They have a high affinity for small β-galactosides, but binding specificity for complex glycoconjugates varies considerably within the family. The ligand recognition is essential for their proper function, and the structures of several galectins have suggested their mechanism of carbohydrate binding. Galectin-9 has two tandem CRDs with a short linker, and we report the crystal structures of mouse galectin-9 N-terminal CRD (NCRD) in the absence and the presence of four ligand complexes. All structures form the same dimer, which is quite different from the canonical 2-fold symmetric dimer seen for galectin-1 and -2. The β-galactoside recognition mechanism in the galectin-9 NCRD is highly conserved among other galectins. In the apo form structure, water molecules mimic the ligand hydrogen-bond network. The galectin-9 NCRD can bind both N-acetyllactosamine (Galβ1–4GlcNAc) and T-antigen (Galβ1–3GalNAc) with the proper location of Arg-64. Moreover, the structure of the N-acetyllactosamine dimer (Galβ1–4GlcNAcβ1–3Galβ1–4GlcNAc) complex shows a unique binding mode of galectin-9. Finally, surface plasmon resonance assay showed that the galectin-9 NCRD forms a homophilic dimer not only in the crystal but also in solution.

Lectins recognize and bind carbohydrates covalently linked to proteins and lipids on the cell surface and within the extracellular matrix, and they mediate many cellular functions ranging from cell adhesion to pathogen recognition. The galectins are a family of animal lectins that share a conserved carbohydrate recognition domain (CRD)² of about 130 amino acids. The galectin CRDs all bind small β-galactosides, but the overall binding affinity for more complex glycoconjugates varies substantially. To date, 14 members of the mammalian galectin family have been identified (1). Hirabayashi and Kasai (2) proposed designating galectin subfamilies as proto-, chimera-, and tandem-repeat types based on their domain organization. The prototype galectins (galectin-1, -2, -5, -7, -10, -11, -13, and -14) consist of a single CRD with a short N-terminal sequence, but the tandem-repeat type galectins (galectin-4, -6, -8, -9, and -12) are composed of two non-identical CRDs joined by a short linker peptide sequence. The single chimera-type galectin (galectin-3) has one CRD and an extended N-terminal tail containing several repeats of proline-tyrosine-glycine-rich motif. The structures of several galectin CRDs have been reported, and all exhibit a β-sandwich fold containing two antiparallel β-sheets (3–6). However, their quaternary structures differ. Galectin-1 and -2 form non-covalently associated homodimers through extended β-sheet interactions (7). The association state of galectin-3 is regulated by its N-terminal domain, and it can exist in monomeric or oligomeric forms (8). Finally, because the tandem-repeat type galectins possess two different CRDs, they may adopt more complex assembly states.

Galectins are found in both the cytoplasm and extracellular regions where they regulate inflammation, cell adhesion, cell proliferation, and cell death (9). Galectins lack a traditional signal sequence, and several are secreted by an unorthodox mechanism to exert their extracellular functions (10). There is a variety of potential glycoconjugate targets for galectins in mammalian cells, but the molecular mechanisms of carbohydrate recognition remain unclear.

Galectin-9, a tandem-repeat type galectin, is a 40-kDa protein consisting of 353 amino acids. The sequence identity between the N- and C-terminal CRDs is 35%. The C-terminal CRD of galectin-9 has a high affinity for small β-galactosides, but the overall binding affinity for complex glycoconjugates varies substantially. The structure of the galectin-9 NCRD reveals the basic mechanism of carbohydrate recognition.

---

*This work was supported by the Protein 3000 project from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‘advertisement’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2D6K, 2D6L, 2D6M, 2D6N, 2D6P, and 2D6Q) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2.

To whom correspondence should be addressed. Tel.: 81-29-879-6177; Fax: 81-29-879-6179; E-mail: ryuichi.kato@kek.jp.

The abbreviations used are: CRD, carbohydrate recognition domain; NCRD, N-terminal CRD; CCRD, C-terminal CRD; LacNAc, N-acetyllactosamine; T-antigen, Thomsen-Friedenreich antigen; LN2, N-acetyllactosamine dimer; Tris, tris(hydroxymethyl)aminomethane; CHES, N-tris(hydroxymethyl)methylamine; Galβ1–3Galβ1–4GlcNAcβ1–3Galβ1–4GlcNAcβ1–3Galβ1–4Glc; MD1a, NeuAcα2–3Galβ1–4GlcNAcβ1–3Galβ1–4GlcNAcβ1–3Galβ1–4Glc; GD1a, NeuAcα2–3Galβ1–4GlcNAcβ1–3Galβ1–4GlcNAcβ1–3Galβ1–4Glc; GD1b, Galβ1–3Galβ1–4GlcNAcβ1–4NeuAcα2–6NeuAcα2–3Galβ1–4Glc; GaNAcβ1–3Galβ1–4Glc.
Crystal Structure of Mouse Galectin-9 N-terminal CRD

CRD (CCRD) is highly homologous to rat galectin-5 CRD with an amino acid sequence identity of 70%, but the N-terminal CRD (NCRD) is only moderately homologous with the known galectins. Among these, the galectin-9 NCRD shows the highest sequence identity (40%) with the galectin-3 CRD.

Galectin-9 was first cloned from tumor cells from Hodgkin disease, a condition characterized by blood and tissue eosinophilia (11). Moreover, the recombinant galectin-9 causes thymocyte apoptosis in mouse cells, suggesting a possible role for galectin-9 in negative selection during T-cell development (12, 13). Interestingly, galectin-9 was shown to be related to a novel eosinophil chemoattractant produced by T lymphocytes, previously designated “ecalectin” (14). Mutation studies showed that both the NCRD and CCRD of galectin-9 were required for the eosinophil chemoattraction activity (15). Additionally, galectin-9 interacts with Tim-3, which is specifically expressed on the surface of T helper type 1 (T\(_h\),1) cell, through recognition of Tim-3 carbohydrates, and the Tim3-galectin9 pathway induces cell death in T\(_h\),1 cells. This suggests that galectin-9 plays a role in down-regulating the effector T\(_h\),1 responses (16).

Galectin-9 interacts with carbohydrate(s) covalently attached to the surface of Tim-3, but the molecular and structural basis for this recognition is unknown. In vitro analyses showed that galectin-9 has a high affinity for a variety of oligosaccharides containing β-galactosides (17), and the NCRD and CCRD of galectin-9 have different oligosaccharide-binding affinities. The biological activities of galectin-9 may be related to the ligand binding specificity of each CRD and the multivalent binding conferred by two CRDs. To date, the structures of many CRDs from fungi to human have been solved, but there is no structural information about the structure of tandem-repeat type galectin CRDs. Such information should greatly clarify the mechanism of carbohydrate recognition by the CRDs and the multivalent properties that lead to multiple functions for a single protein.

Compared with the galectin-9 CCRD, the NCRD shows striking affinities for complex glycoconjugates such as Fossmann pentasaccharide and polymerized N-acetyllactosamine (17, 32). The specific interactions of the galectin-9 NCRD with the carbohydrates is thought to be the clue for understanding the physiological mechanism of galectin-9. We report here the crystal structures of the mouse galectin-9 N-terminal CRD in the absence and the presence of carbohydrate ligands. These structures show both the basic mechanism of carbohydrate binding and suggest a potential mechanism for the specificity of carbohydrate recognition and binding. Additionally, the galectin-9 NCRD dimerizes in both crystal and solution. From these observations, we discuss the relationship between the structure and function of galectin-9.

EXPERIMENTAL PROCEDURES

Materials—Glutathione-Sepharose 4B, benzamidine-Sepharose, and Superdex75 columns for protein purification were purchased from Amersham Biosciences. Carbohydrate chemicals of lactose (Galβ1–4Glc) and Thomsen-Friedenreich antigen (T-antigen, Galβ1–3GalNAc) were from Sigma-Aldrich and Merck Ltd., respectively. N-Acetyllactosamine (LacNAc, Galβ1–4GlcNAc) was synthesized by a previously described method (18). N-Acetyllactosamine dimer (LN2) was also synthesized by the procedure described below. All crystallization reagents were purchased from Hampton Research (Aliso Viejo, CA) and deCODE genetics. Other chemicals were obtained from Wako Pure Chemical Industries Ltd. (Japan) and Sigma-Aldrich.

Synthesis of LN2 Carbohydrate—UDP-GlcNAc and UDP-Gal were kind gifts from Yamasa Corp. (Japan). Fetal bovine serum was purchased from Dainippon Pharma Co., Ltd. (Japan). LacNAc (460 mg) UDP-GlcNAc-2Na (264 mg) and UDP-Gal-2Na (244 mg) were dissolved in 20 ml of 150 mM sodium cacodylate buffer (pH 6.8) containing MnCl2 (64 mg), ATP-2Na (20 mg), and 0.02% NaN3 (w/v) followed by addition of 3.2 g of crude enzyme preparation obtained by 80% saturated ammonium sulfate precipitation from fetal bovine serum. The mixture was incubated for 8 days at 310 K, and the reaction was terminated by boiling for 5 min. The resulting precipitate was removed by centrifugation (10,000 × g, 15 min), and the supernatant was loaded onto a charcoal-Celite column (φ2.5 × 25 cm) equilibrated with H_2O at a flow rate of 2.5 ml/min. The column was washed with 150 ml of H_2O and eluted with a linear 0–50% ethanol gradient (total 2,000 ml). Absorbance was monitored at 210 nm, and peak fractions were collected, concentrated, and applied onto a Toyopearl HW-40S column (φ2.5 × 65 cm) equilibrated with H_2O at a flow rate of 0.5 ml/min. The column was eluted with H_2O, and the fraction containing the product was concentrated and lyophilized to give LN2 carbohydrate (40.1 mg).

Protein Purification and Crystallization—The N-terminal CRD of mouse galectin (15) was expressed as a glutathione S-transferase fusion protein in Escherichia coli strain BL21(DE3) cells using plasmid pGEX4T-1 (Amersham Biosciences). The cells were disrupted by sonication at 277 K. The supernatant was applied to a glutathione S-transferase affinity column of glutathione-Sepharose 4B and washed with 50 mM Tris-HCl buffer (pH 8.0) containing 500 mM NaCl and 1 mM dithiothreitol. The fusion protein bound to the resin was eluted with 10 mM glutathione-containing buffer, and glutathione S-transferase was removed from the fusion protein by cleaving with human α-thrombin (Amersham Biosciences) at 10 units/ml for 12 h at 293 K. The cleaved proteins were collected for further purification by benzamidine-Sepharose and Superdex 75. The purified protein was a single band on SDS-PAGE stained with Coomassie Brilliant Blue.

Crystals were grown using the hanging drop vapor diffusion method from drops containing equal volumes of protein (6 mg/ml) in 10 mM Tris (pH 8.0), 100 mM NaCl, 1 mM dithiothreitol, and precipitant composed of 0.1 M CHES (pH 9.5), 15% ethanol at 289 K. The crystals (apo form I) attained dimensions of 0.1 × 0.1 × 0.02 mm\(^3\) within 1–2 weeks and diffracted to 2.5-Å resolution. They grew in space group \(P_4_1_2_1_2\) with unit cell dimensions of \(a = b = 58.1\), and \(c = 221.7\) Å. There are two molecules in the asymmetric unit. Under almost identical conditions (0.1 M Tris-HCl (pH 7.5), 15% ethanol), another crystal (apo form 2) was obtained in another space group \(P_2_1_2_1_2\) with unit cell dimensions of \(a = 56.4, b = 58.6, c = 48.4\) Å, and only one molecule in the asymmetric unit. The complex crystals with carbohydrate were prepared by two methods. Lactose
crystal complexes were obtained by co-crystallization in 0.1 M sodium citrate (pH 5.0), 5% polyethylene glycol 6000, 10 mM lactose, and T-antigen complex crystals were obtained by soaking the apo form1 crystal in 0.1 M sodium citrate (pH 5.0), 5% polyethylene glycol 6000, 10 mM galactose in 10 mM of each solution. In the galactose complex, the electron density of sugar moiety was poor, but the addition of longer carbohydrates greatly improved the quality of the models were shown to be reasonable using PROCHECK program (22). Final statistics of the crystallographic refinement are summarized in Table 1. The figures were drawn with the programs MOLSCRIPT (23), Raster3D (24), GRASP (25), and PyMOL. Coordinates for the galectin-9 N-terminal CRD are being deposited in the Protein Data Bank of the Research Collaboratory for Structural Bioinformatics. The Protein Data Bank accession numbers for the apo form1, apo form2, lactose complex, LacNAC complex, T-antigen complex and LN2 complex structures are 2D6K, 2D6L, 2D6M, 2D6N, 2D6P, and 2D6O, respectively.

Surface Plasmon Resonance Measurement—SPR binding assay was performed at 298 K using BIACORE 2000 (Biacore). HBS buffer (10 mM HEPES (pH 7.2), 150 mM NaCl, 0.005% (v/v) Surfactant P20) was used as a running buffer at a flow rate of 5 μl/min. The galectin-9 NCRD was directly immobilized on CM4 sensor chips (Biacore) by amine coupling. The binding response was measured at 1–50 μM concentrations of galectin-9 NCRD as analyte. The net response was calculated by subtracting the background response from the binding response. Steady-state responses (R<sub>eq</sub>) were determined from the net response of sensorgrams using BIAevaluation 3.2 program (Biacore). The R<sub>eq</sub> values were plotted against galectin-9 NCRD concentrations and were fitted to a simple 1:1 steady-state binding model using the BIAevaluation 3.2 software, R<sub>eq</sub> = C·R<sub>max</sub>/(K<sub>d</sub> + C), where C is the analyte galectin-9 NCRD concentration, R<sub>max</sub> is the maximum binding response, and K<sub>d</sub> is the equilibrium dissociation constant. RESULTS

Overall Structure—We determined the x-ray crystal structures of mouse galectin-9 NCRD in the absence of ligand in two crystal forms and in the presence of four different carbohydrate complexes were solved by molecular replacement using the apo form1 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form2 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form1 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form2 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form1 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form2 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form1 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form2 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form1 structure as a search model. All carbohydrate complexes were solved by molecular replacement using the apo form2 structure as a search model.
drates (lactose, LacNAc, T-antigen, and LN2). The crystal structure of galectin-9 NCRD apo form1 was determined by the molecular replacement method using the galectin-3 structure (PDB code: 1A3K) as a search model. The carbohydrate binding site is shown by a dotted box. The dimeric structure of the galectin-9 N-terminal CRD is shown. Two monomers in an asymmetric unit in the apo form1 crystal are shown in red (chain-A) and green (chain-B), respectively. The amino acid residues involved in the dimer formation are shown in ball-and-stick model. The carbon, oxygen, nitrogen, and sulfur atoms are shown in white, red, blue, and yellow spheres, respectively. Hydrogen bonds are depicted by red dotted lines. D, electrostatic potential maps of the dimer surfaces of the galectin-9 N-terminal CRD (upper) and galectin-1 CRD (lower) (PDB code: 1GZW). Positive (blue) and negative (red) potentials are mapped on the van der Waals surfaces in the range $-10 k_B T$ to $10 k_B T$, where $k_B$ is Boltzmann’s constant and $T$ is the absolute temperature. The orientation of the galectin-9 N-terminal CRD dimer is same as Fig. 1B.

Dimer Formation—In the asymmetric unit of apo form1, two molecules, referred as chains A and B for clarity, are related by a 2-fold non-crystallographic axis perpendicular to the
The contact area of the dimer is 615 Å², and this suggests that the CCRD does not prevent dimer formation. In our structures, this contact is conserved in all six crystal structures. This positional flexibility may affect the hydrogen network. These waters, which are held in the carbohydrate binding cleft, help to stabilize the spatial arrangement of the amino acid side chains involved in carbohydrate recognition in the absence of the ligand. Because the water molecules in the ligand binding cleft are reported to mimic the carbohydrate binding mode in galectin-1, chick galectin CG16 and fungal galectin CG1L2, galectins may generally use such a water stabilization mechanism.

β-sheets. They form a continuous 12-stranded antiparallel β-sheet through interactions between the β-strands of chain-A S6 and chain-B S6 (Fig. 1B). On the dimer interface, the main-chain oxygen and nitrogen atoms of Arg-86 form hydrogen bonds with the corresponding main-chain atoms of Arg-86 of the other monomer (Fig. 1C). Previously, a modified galectin-9 construct lacking the linker region was generated (29), and substitution of Arg-86 with alanine increased the solubility of this protein (data not shown). Thus, a local conformational change caused by this mutation may affect the intermolecular interaction of two galectin-9 monomers, suggesting that dimer formation occurs at the S6 strands in solution. The main-chain oxygen atoms of Glu-84 of both molecules also form hydrogen bonds with the main-chain nitrogen atoms of Met-88 of the other molecules. The N and C termini of each monomer are positioned at the opposite side of the dimer interface (Fig. 1A). This suggests that the CCRD does not prevent dimer formation of the NCRD. The contact area of the dimer is 615 Å², and this value is slightly smaller than that of galectin-1 (670 Å², PDB code: 1GZW). Although galectin-1 and -2 form homodimers, the interfaces are formed by extended β-sheet interactions across the two monomers at both sides (F1 and S1) (3, 4). The architecture is quite different from that of the galectin-9 NCRD. In our structures, this contact is conserved in all six crystals, which contain different space groups with only slight differences in the monomer orientations.

Apo Form Structure—The apo and lactose binding structures are very similar to each other with an r.m.s.d. of Ca atoms of 0.4 Å. The carbohydrate binding site of apo form1 is occupied by four well ordered water molecules, which correspond to O4 (Wat-1), cyclic O5 (Wat-2), and O6 (Wat-3) of the β-galactoside and O6 of the reducing sugar (Wat-4) in the complexes with lactose and other carbohydrates (Fig. 2A). In the other apo form structure (apo form2), the positions of the water molecules are slightly shifted, and the hydrogen bond network is also slightly altered (Fig. 2B). The position of wat4 is common in both structures, but wat1 and wat2 move by 1.0 Å without losing the hydrogen bond network. Wat-3 moves by 0.4 Å and makes a new hydrogen bond with Asn-74. Because Wat-2 does not directly interact with the protein in both structures, this positional flexibility may affect the hydrogen network.

**Carbohydrate Recognition Mechanism**—Galectin-9 recognizes many β-galactoside-containing carbohydrates (17, 33, 34). To elucidate the recognition mechanism at an atomic level, we determined four complex structures with different saccharides. The positions of the β-galactoside moiety at the non-reducing end are virtually the same in all the carbohydrate complexes examined (Fig. 3A). The β-galactoside moiety is most deeply buried in the binding site formed by β strands S4—S6. O4 of the galactose plays a central role in forming hydrogen bonds, accepting protons from two highly conserved residues His-60 and Asn-62. O6 of the galactose also interacts with Asn-74 via another hydrogen bond. Two residues with planar side chains, His-60 and Trp-81, provide contacts that help align the carbohydrate. Trp-81 participates in a stacking interaction with the galactose ring similar to that seen in a number of other galactose and lactose binding lectins (35).

In the lactose complex, O6 of the glucose moiety is recognized by Arg-64, Glu-84, and Arg-86 through hydrogen bonds (Fig. 3B). Arg-64 also makes a hydrogen bond with O4 in the galactose moiety. Replacement of Arg-64 with Ala in mouse galectin-9 impairs its capacity to bind to Tim-3 (16). Moreover, ecallentin, which was previously isolated as an eosiinophil chemoattractant from a human T-cell-derived expression library, is a variant of human galectin-9 with amino acid sequence identity of 66%, and substitution of Arg-65 by Asp, which corresponds to Arg-64 in mouse galectin-9, disrupts its ability to bind lactose (15). The importance of Arg-64 for galectin-9 activity shown by these experiments is explained well by our crystal structures.

Next, we determined the complex structures of galectin-9 NCRD with LacNAC and T-antigen. In the LacNAC complex, the N-acetylglucosamine (GlcNAc) moiety is exposed to solvent and is recognized by Arg-64 and Glu-84, which make hydrogen bonds with O3 of the GlcNAc (Fig. 3C, white). The methyl group makes a van der Waals contact with the guanosine head group of Arg-86. In contrast, T-antigen adopts a...
Crystal Structure of Mouse Galectin-9 N-terminal CRD

Complex Structure with Tetrasaccharide—Galectin-9 preferentially binds to glycans carrying repeating units of LacNAc (17, 33). We obtained crystals of the galectin-9 NCRD in complex with LacNAc dimer (LN2) by the soaking method. The electron density of the four carbohydrate residues was clearly identified in the initial map (Fig. 4A). Interestingly, the coordination of His-60, Asn-62, Arg-64, Asn-74, and Trp-81, which interact with the terminal galactose at the non-reducing end of the molecule, is very similar to that of the LacNAc monomer complex. Conversely, the LacNAc moiety at the reducing end protrudes and is recognized by the neighboring molecules of galectin-9 NCRD. The Ca atom of Gly-68 in the neighboring molecule makes van der Waals contact with C1 in GlcNAc at the reducing end of LN2. This additional interaction of the forth sugar residue with the neighboring galectin-9 NCRD molecule likely increases the affinity to glycans with repeating LacNAc units.

Protein-Protein Interaction—We showed that the galectin-9 NCRD forms a dimer in all the six crystal structures obtained (two apo-forms and four carbohydrate complexes). To determine whether the NCRD-NCRD interaction occurs in solution, we performed an SPR analysis using BIACORE. The resonance unit of galectin-9 NCRD immobilized on the sensor chip increased with increasing concentrations of analyte galectin-9 NCRD. The plot of analyte concentration versus steady-state resonance unit ($R_{eq}$) is well fitted with a 1:1 steady-state binding model with a dissociation constant ($K_D$) of 20 μM (Fig. 5). Thus, the galectin-9 NCRD exhibits homophilic interactions in solution, and suggests that the interacting molecules may form a dimer in solution as in the obtained crystals.

The carbohydrate binding site of galectin-9 NCRD is located close to the dimer interface. This striking feature of galectin-9 NCRD is in stark contrast to that seen for galectin-1 and -2. The dimer surfaces of human galectin-1 and toad ovary galectin have a long negatively charged cleft in the cavity containing the carbohydrate binding pocket (30, 36). In contrast, the galectin-9 NCRD has a large positively charged patch on the dimer surface (Fig. 1D). The differences in the electrostatic potential of their surfaces may reflect differences in their physiological targets.

FIGURE 3. Structures of the galectin-9 N-terminal CRD-ligand complexes at the carbohydrate binding site. A, superposition of four carbohydrate complex structures. The main chains of the N-terminal CRDs and the carbohydrates are depicted by rod model. The amino acid residues that interact with β-galactoside at the non-reducing end are shown by ball-and-stick model. The complexes with lactose, T-antigen, LacNAc, and LN2 are colored by orange, cyan, green, and yellow, respectively. B, structure of the carbohydrate binding site of the lactose complex. The lactose molecule and the amino acid residues that interact with the lactose are shown in ball-and-stick model. The atoms of carbon, oxygen, and nitrogen are shown in white, red, and blue spheres, respectively. Hydrogen bonds are depicted by dotted lines. C, superposition of LacNAc complex (protein: cyan; ligand: white) and T-antigen complex (protein: green; ligand: yellow). The amino acid residues involved in GlcNAc and GalNAc recognition are shown in ball-and-stick models. Hydrogen bonds in LacNAc and T-antigen complexes are shown in red and blue dotted lines, respectively.

different conformation compared with LacNAc at the reducing end. The O4 in N-acetylglactosamine (GlcNAc) makes hydrogen bonds to Arg-64 and Glu-84 (Fig. 3C, yellow). The methyl group of GalNAc is situated away from Arg-86 and located near Arg-43. As the distance between O7 of GalNAc and NH1 of Arg-43 is too long to interact directly with each other, a water-mediated interaction may exist between them. We could not unambiguously assign water molecules around the carbohydrate binding site because of the relatively low resolution of the structure. Comparing the LacNAc and T-antigen complex structures, the position of O3 of GlcNAc is almost identical to that of O4 of GalNAc both of which are recognized by Arg-64 and Glu-84, but the orientation of the acetyl group differs (Fig. 3C). In the case of LacNAc complex, the GlcNAc moiety points to Arg-86, and this may explain why the galectin-9 NCRD can bind both types of sugar chains, β1–3 and β1–4.
DISCUSSION

Galectins are important regulators of a variety of cellular functions, and there are several galectin sub-families. This is the first crystal structural analysis of a tandem-repeat type galectin. Although the exact structure of target carbohydrates recognized by mouse galectin-9 in vivo is unclear, two exhaustive in vitro ligand analyses for the human galectin-9 NCRD were reported (17, 33). The overall sequence identity between the mouse and human galectin-9 NCRDs is 70% (Fig. 6). But the amino acid residues at around the carbohydrate binding site coincide well with each other. Thus, it is likely that the physiological targets and roles of both galectin-9 are essentially the same.

Our structural study revealed that the mouse galectin-9 NCRD is a typical β-sandwich fold as previously reported for other galectin structures, however, the dimer architecture of the crystals differs substantially from the prototype galectins, such as galectin-1 and CG-16 (30, 31). Galectin-1 forms a head-to-head dimer in which the N-terminal S1 strand is used for the dimer formation, whereas the galectin-9 NCRD forms a tail-to-tail dimer in which the S6 strands located at the opposite side of the N-terminal of the molecule interact with other for dimerization. Additionally, there is a significant local conformational difference between the galectin-1 and galectin-9 NCRD structures. The N-terminal region specific for the galectin-9 NCRD shields the F1 strand, which is responsible for dimer formation in galectin-1 (supplemental Fig. S1A). Within the S1 strand, Ser-7, which forms a hydrogen bond at the dimer interface of galectin-1, is replaced by Pro-20, and the proline ring sticks out to F1 strand (supplemental Fig. S1B). This change prevents the galectin-9 NCRD from forming head-to-head dimers. In contrast, the S6 strand in galectin-1 is kinked at Glu-74, and as a result, the position of Ala-75, which corresponds to Met-88 in galectin-9, moves and prevents tail-to-tail dimerization in galectin-1 (supplemental Fig. S1C).

The overall structure of the galectin-9 NCRD is also quite similar to those of the galectin-3 and galectin-7 CRDs, which exist as monomers in solution, except for the N-terminal tail and S6 strand. The positions of oxygen atoms in the carbonyl backbone of the galectin-3 and -7 S6 strands differ slightly from that of the galectin-9 NCRD (supplemental Fig. S2), and this explains the different dimer arrangements of these proteins. The amino acid sequences of the S6 strands of the galectin-9 NCRD and CCRD are not identical. This suggests that the galectin-9 CCRD may not form tail-to-tail dimers like the NCRD. Interestingly, galectin-5, which is a prototype galectin with 78% sequence identity to the galectin-9 CCRD, weakly agglutinates rat erythrocytes, suggesting oligomerization (37). The galectin-9 CCRD may form oligomers, but the protein-protein interaction mode is likely different from that of the NCRD.

Carbohydrate recognition is the first and most critical step in galectin function. The amino acid residues involved in β-galactoside recognition are well conserved among proto-, chimera-, and tandem-repeat type
galectins, whereas the residues involved in disaccharide recognition differ. Arg-86 of the galectin-9 NCRD recognizes the glucose moiety in lactose and the GlcNAc moiety in LacNAc via hydrogen bonds. The galectin-8 NCRD has weaker affinities for lactose and LacNAc than the galectin-9 NCRD as assessed by SPR (38). Arg-86 in the galectin-9 NCRD is replaced by isoleucine in galectin-8, and this is presumably responsible for the weaker interactions. In contrast, the affinity for T-antigen may be similar among the tandem-repeat type galectins, because Arg-43, which interacts with the carbohydrate of the reducing end of the galectin-9 NCRD, is conserved in all tandem-repeat-type galectins. The 9-amino acid deletion from the C terminus of the human galectin-9 NCRD (Val-140 to Gln-148) does not allow lactose binding (29). This region (F1 strand) is a part of the $\beta$-sandwich arrangement of the galectin-9 NCRD and causes its lack of carbohydrate binding activity.

The human galectin-9 NCRD has significant affinity for glycolipid-type glycans, such as ganglioside GA1, GM1, GD1a, GD1b, and Gb4 (17). Because the Galβ1–3GalNAc structure is commonly found in these gangliosides, our crystal structure of the galectin-9 NCRD with the T-antigen complex explains the mechanism of this interaction with high affinity. Conversely, the affinity of the human galectin-9 NCRD for GM3 is dramatically weaker than that for GM1. We built a model structure of the mouse galectin-9 NCRD-GM1 complex (data not shown) from the structures of the T-antigen complex and cholera toxin B-pentamer-GM1 complex (PDB code: 1CT1 (39)). This model was very similar to the galectin-1-GM1 complex obtained by NMR analysis (40), where the lipid portion of GM1 was observed at the opposite side of the protein molecule. This orientation would support the interaction of galectins with glycolipids on the cell surface.

The human galectin-9 NCRD also exhibits high affinity for two other types of glycolipids, Forssman pentasaccharide (GalNAc1–3GalNAcβ1–3Galβ1–4Glc) and A-hexasaccharide (GalNAc1–3[Fuc1–2]Galβ1–3GlcNAcβ1–3Galβ1–4Glc) (17). Although the binding sites for these oligosaccharides remain unclear, these glycans have a common structure, which is the same as lactose, at the reducing end of the carbohydrate chains. If the lactose moieties of these carbohydrates bind to the galectin-9 NCRD analogous to lactose, the remaining carbohydrate chain would interact with the extended cleft formed by the S1–S3 strands. Because the amino acid sequences of the S1–S3 strands are not conserved among the tandem-repeat type galectins, ligand specificity may be determined by this region. Moreover, the positively charged surface of the galectin-9 NCRD may facilitate recognition of glycans attached to negatively charged extracellular surfaces under physiological conditions.

In many cases, galectins act as mediators of cell adhesion by binding to glycoconjugates at the cell surface or the extracellular...
lar matrix. Tandem-repeat type galectins may act as a bridge between specific carbohydrates, because they have two CRDs within one molecule. In the case of mouse galectin-9, mutation of either Arg-64 in the NCRD or Arg-238 in the CCRD, which corresponds to Arg-64 of the NCRD, decreased binding to Tim-3, whereas the double mutant completely abolished binding (16). Thus, both CRDs appear required for the strong interaction with Tim-3. Additionally, in the case of human galectin-9, both the individual N- and C-terminal CRDs exhibit eosinophil chemoattractant activity, however, this activity was substantially lower than full-length wild-type galectin-9. Likewise, each human galectin-9 CRD exhibits hemagglutination activity, but this activity is also lower than that of the wild-type, full-length protein (15). Conversely, recombinant chimeric proteins consist of two NCRDs or two CCRDs joined by a linker, have virtually the same eosinophil chemoattractant activity as wild-type galectin-9 (33). These results suggest that two CRDs connected by a linker are required for galectin-9 activity, but the precise domain identity and carbohydrate binding specificity are flexible.

The ability of galectins to cross-link ligand molecules is essential for cell adhesion, signal transduction through receptor clustering, and formation of multivalent galectin-glycoprotein networks on the cell surface (41–45). Galectins regulate the degree of protein cross-linking thereby affecting a variety of physiological activities. Galectin-14, which was cloned from ovine eosinophil-rich leukocytes, is a prototype galectin with a physiological activity. Galectin-1, and many plant lectins, form homogeneous protein networks on the cell surface (41–45). Galectins regulate protein interactions, and our current study cannot assess the contribution of these domains to galectin-9 dimerization. Galectin-9 controls various biological events, such as cell aggregation, chemoattraction of eosinophils, and induction of apoptosis of thymocytes, immunocytic T cells, and melanoma cells (13–16, 26). The multiple functions of galectin-9 may arise from the strict carbohydrate recognition and cross-linking between galectin-9 and target molecules. A recent report revealed that the carbohydrate on Tim-3, a T_H1-specific cell surface molecule, contains a region targeted toward mouse galectin-9, and the interaction of these molecules regulates T_H1 immunity (16). Structural studies on complexes between galectin-9 and its newly identified ligands will provide further insights into the molecular and cellular functions.

Acknowledgments—We thank the staff of beamlines BL-5A, BL-6A, and AR-NW12A at Photon Factory, Tsukuba, Japan and BL41XU at SPring-8, Harima, Japan for providing the data collection facilities and their support.

REFERENCES

Crystal Structure of Mouse Galectin-9 N-terminal CRD

Crystal Structure of the Galectin-9 N-terminal Carbohydrate Recognition Domain from *Mus musculus* Reveals the Basic Mechanism of Carbohydrate Recognition

Masamichi Nagae, Nozomu Nishi, Takeomi Murata, Taichi Usui, Takanori Nakamura, Soichi Wakatsuki and Ryuichi Kato

doi: 10.1074/jbc.M606648200 originally published online September 21, 2006

Access the most updated version of this article at doi: 10.1074/jbc.M606648200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2006/09/25/M606648200.DC1

This article cites 48 references, 13 of which can be accessed free at
http://www.jbc.org/content/281/47/35884.full.html#ref-list-1