Structural Basis for Carbohydrate Recognition and Anti-inflammatory Modulation by Gastrointestinal Nematode Parasite *T. leonina* Galectin

Eun Young Hwang, Mi Suk Jeong, Sang Kyun Park, Sung Chul Ha, Hak Sun Yu, and Se Bok Jang

From the Department of Molecular Biology, College of Natural Sciences, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeongsu, Busan 46241, the Pohang Accelerator Laboratory, Pohang University of Science and Technology, 80, Jigok-ro 127 beon-gil, Nam-gu, Pohang-si, Gyeongsangbuk-do 37673, and the Department of Parasitology, School of Medicine, Pusan National University, 49, Busandaehak-ro, Mulgeum-eup, Yangsan-si, Gyeongsangnam-do 50612, Korea

Edited by Luke O'Neill

**Toxascaris leonina** galectin (Tl-gal) is a galectin-9 homologue protein isolated from an adult worm of the canine gastrointestinal nematode parasite, and Tl-gal-vaccinated challenge can inhibit inflammation in inflammatory bowel disease-induced mice. We determined the first X-ray structures of full-length Tl-gal from the concave surfaces of both carbohydrate recognition domains (CRDs) in Tl-gal. All binding sites were found in the Hemagglutination activities of Tl-gal were inhibited by interactions with carbohydrates and mutations. Hemagglutination activities of Tl-gal were inhibited by interactions with carbohydrates and mutations. We found that the mutation of Tl-gal (E80A/E215A) at the carbohydrate binding region induced protein aggregation and was caused in many diseases. The short linker region between the N-terminal and C-terminal CRDs of Tl-gal was very stable against proteolysis and maintained its biological activity. This structural information is expected to elucidate the carbohydrate recognition mechanism of Tl-gal and improve our understanding of anti-inflammatory mediators and modulators of immune response.

Parasite-derived proteins can inhibit ovalbumin-specific Th2 responses, and immunization with parasite protein has been shown to exert greater protection than allergen transfection. The complex interplay of parasites between infections and autoimmunity has recently been the focus of many studies. The relationship between intestinal helminthic parasites has been of particular interest because it has been used to cure autoimmune diseases, especially Crohn’s disease (2). *Toxascaris leonina* is a gastrointestinal nematode parasite found in adult canines, and *T. leonina* galectin (Tl-gal) is a galectin-9 homologue gene isolated from the *T. leonina* parasite. Tl-gal can inhibit inflammatory reactions through inhibition of Th1 and Th2 cytokine production by increasing production of the cytokines TGF-β and IL-10 (3). Tl-gal-vaccinated challenge can inhibit inflammation in inflammatory bowel disease-induced mice, and the colons of the Tl-gal treated mice were longer than those of inflammation-induced dextran sulfate sodium-treated mice.

Tl-gal shows 15–35% sequence identity with other tandem repeat type galectins (galectin-4, -6, -8, -9, and -12) of humans, mice, and rats and exhibits the highest similarity with human galectin-9, which was confirmed as a tumor antigen of unknown function in Hodgkin’s disease (4). The protein was shown to be identical to the potent eosinophil chemottractant, ecalcin, cloned from a human T-cell line (5). Galectin-9 is an element of the galectin family of carbohydrate-binding proteins, and mouse galectin-9 induces thymocyte apoptosis in a lactose-inhibitable manner (6). The ability of galectin to bind with the carbohydrate also modulates the secretion of galectins because they are devoid of classical secretion signals (7). Extracellularly, the ability to interact with carbohydrates facilitates cell-cell and cell-matrix interactions. This ability also regulates receptor lattice formations on the cell membrane, which in turn regulate the signaling pathway (8). Recent studies have shown that the C-terminal CRD of galectin-9 is a major factor influencing receptor recognition and death pathway signaling in T-cells. In addition, the N-terminal CRD of galectin-9 affects the activation of dendritic cells (9, 10). Tandem repeat type galectin-9 is composed of two CRDs and linker peptides and can also be in multimeric form (11–13). The protein is very susceptible to proteolysis. There are three isoforms of galectin-9, Gal-9(L), Gal-9(M), and Gal-9(S), which vary according to the length of the linker domain between the...
NCRD and CCRD. The differences in eosinophil chemoattractant activity between these three isoforms are dependent on the carbohydrate binding activity of their CRDs; however, these isoforms show no differences in cancer cell adhesion or in the expression levels of their T-cells (14–16).

The molecular basis of the specific carbohydrate recognition and mechanism of immune evasion of full-length galectin have yet to be elucidated. Although TI-gal shows a lower carbohydrate binding ability than other galectins (i.e. rat galectin), the carbohydrate binding ability of TI-gal has been confirmed (3). The particular preference for carbohydrate ligands will help to clarify the molecular mechanism by which TI-gal displays its many biological functions. In this study, we determined the first X-ray crystal structures of full-length TI-gal having a short linker domain with carbohydrate ligands (lactose, N-acetyllactosamine (LacNAc), lacto-N-tetraose (LNT), sialyllactose (SiaLac), and glucose) on both CRDs of TI-gal. The anti-inflammatory mediator activity of TI-gal was inhibited by interaction with carbohydrates and mutation. In addition, the important key binding residues of TI-gal responsible for recognizing carbohydrate ligands based on three-dimensional structures were mutated, and the mutation induced protein aggregation. Studies of TI-galectin alone and in complexes with carbohydrate ligands will provide important clues to their structure-function relationships. The information presented herein improves our understanding of anti-inflammatory mediators and modulators of immune response.

Results

Overall Structures of the TI-gal and Carbohydrate Complexes—
To measure the immune regulatory function of TI-gal in inflammatory response, recombinant TI-gal using Escherichia coli was constructed. The physiological function of TI-gal probably depends on its carbohydrate recognition ability. Numerous carbohydrates have a more polar and a less polar face. The more polar faces hydrogen-bond with the protein, whereas the less polar faces form hydrophobic interactions with nonpolar residues. These interactions produce high-affinity binding and high specificity of galectins for their carbohydrates, which represents a kind of information transfer that is clearly central in many processes within and between cells. We investigated the interaction and structural basis between full-length TI-gal and five carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose).

The galectin-9 chemoattractant activity is dependent on its carbohydrate binding activity and requires both NCRD and CCRD (14–16). TI-gal is a galectin-9 homologue protein that has carbohydrate binding activity (3). In this study, we successfully determined the structures of high resolution full-length TI-gal and its complexes with carbohydrates at high resolutions of 1.7–2.1 Å (Table 1). We also obtained the TI-gal mutant crystals (W77F/W212F and R61A/R196A) at the carbohydrate binding sites; however, they did not provide diffraction data. Mutant protein (E80A/E215A) of TI-gal at the other carbohydrate binding sites was aggregated during purification. Protein aggregation is a biological situation in which misfolded proteins aggregate intra- or extracellularly (17, 18). Exposed hydrophobic partitions of the unfolded protein can interact with the exposed hydrophobic patches of other unfolded proteins, resulting in protein aggregation. The protein aggregates are frequently poisonous and have been implicated in many diseases, including ALS, Parkinson’s, Alzheimer’s, and prion disease (19, 20).

Full-length TI-gal, composed of the N-terminal CRD (residues 1–142) and the C-terminal CRD (residues 150–278), was held together via a short linker (residues 143–149) (Fig. 1A). In previous studies, we determined the first full-length TI-galectin crystal structure at 2.0 Å in space group P2₁2₁2₁ (21). TI-galectin is composed of two antiparallel β-sheets that form β-sandwich arrangements and jelly-roll folds in each CRD. In this study, the structures of wild-type TI-gal at 1.8 Å in P2₁, and carbohydrate complexes revealed that a β-strand was changed to a long loop between β5’ and β6’, whereas a short α-helix loop was formed between β4’ and β5’ of the CCRD relative to the previous TI-gal structure. As a result, the NCRD of TI-gal had 11 β-strands and a small α-helix between β9 and β10, whereas the CCRD had 10 β-strands and no α-helix (Figs. 1B and 2A). All carbohydrates (lactose, LacNAc, LNT, and SiaLac) were bound to concave surfaces surrounded by β4, β5, and β6 of the TI-gal NCRD, and those were also bound to the surfaces formed by β4’, β5’, and the loop between β5’ and β6’ of the TI-gal CCRD (Figs. 1B and 2). The electron density of carbohydrates was of excellent quality (supplemental Figs. 3 and 4). The crystallographic statistics of full-length TI-gal with carbohydrates are shown in Table 1. With the exception of the glucose complex in P1, there were two molecules of TI-gal with three carbohydrates per asymmetric unit in the P2₁ space group. The complex of TI-gal and glucose contained four molecules of TI-gal and four glucose in the CCRD. Interestingly, the glucose did not bind to NCRD in TI-gal.

Binding Affinity between TI-gal and Carbohydrates—Because tryptophan participates in carbohydrate binding, the change in its essential fluorescence can be observed in galectin-4 ligand binding assays (22). In the present study, TI-gal has two tryptophan residues (Trp-77 and Trp-212) involved in carbohydrate binding of the NCRD and CCRD. The maximum fluorescence emission spectra of full-length TI-gal CRDs were observed at 348 nm using an excitation wavelength of 280 nm. After binding of carbohydrates, the fluorescence maxima shifted to a shorter 343-nm wavelength, and the emission intensities changed (Figs. 1C and 3A). The binding between TI-gal and LNT showed a larger fluorescence maxima shift and higher fluorescence intensity than wild-type TI-gal. The binding between TI-gal and carbohydrates could be caused by substantial conformational changes in the complex compared with wild-type TI-gal protein. The less strict hydrophobic circumstances produced conformational changes related to TI-gal and carbohydrates and changes in fluorescence intensity. The mutations (W77F/W212F and R61A/R161A) of TI-gal induced no shifts in the fluorescence maxima. Interestingly, the fluorescence emission intensity of TI-gal mutants (W77F/W212F) showed a greater decrease than that of the wild type. This occurred because of the mutation of tryptophan to phenylalanine as well as the presence of an additional fluorescence shoulder at 303 nm that may have been a weak emission of phenylalanine (Fig. 3B).
Crystal Structures of Tl-gal and Carbohydrates

| Table 1: Crystallographic statistics |
|-------------------------------------|
| Values in parentheses are for the highest resolution shell. |

| Data collection statistics | P2₁ | P2₁ | P2₁ | P2₁ | P2₁ | P1 |
|----------------------------|-----|-----|-----|-----|-----|----|
| Space group                |     |     |     |     |     |    |
| Unit cell dimensions       |     |     |     |     |     |    |
| a, b, c (Å)                | 60.4, 84.0, 78.4 | 60.6, 84.4, 78.4 | 60.8, 84.9, 78.4 | 60.6, 84.3, 78.6 | 60.6, 84.0, 78.8 | 74.6, 78.6, 83.0 |
| α, β, γ (degrees)          | 90, 109.7, 90 | 90, 109.2, 90 | 93.6, 109.5, 90 | 90, 109.2, 90 | 90, 109.7, 90 | 90, 110.7, 106 |
| Resolution (Å)             | 30.0 (1.86-1.80) | 30.1 (1.76-1.70) | 30.2 (2.07-2.00) | 30.0 (2.07-2.00) | 30.2 (2.8-2.10) | 30.2 (2.07-2.00) |
| Completeness (%)           | 97.8 (96.5) | 98.7 (99.9) | 94.0 (98.1) | 99.5 (100.0) | 99.7 (100.0) | 96.4 (96.2) |
| Observed reflections       | 655,614 | 725,964 | 426,572 | 2,427,634 | 863,235 | 580,092 |
| Unique reflections         | 67,664 | 81,697 | 50,859 | 50,425 | 43,403 | 114,308 |
| Rmerge (%)                 | 7.9 (38.1) | 7.3 (33.0) | 5.3 (14.6) | 5.6 (23.3) | 7.9 (41.0) | 5.5 (22.4) |
| Redundancy                 | 3.8 | 3.7 | 3.9 | 7.2 | 3.7 | 1.9 |

| Refinement statistics      |     |     |     |     |     |    |
|----------------------------|-----|-----|-----|-----|-----|----|
| Resolution range (Å)       | 30.1 | 30.1 | 30.2 | 30.2 | 30.2 | 30.2 |
| Rmerge/Rfree (%)           | 22.3/25.8 | 23.7/26.8 | 23.4/27.2 | 22.6/25.7 | 22.3/27.1 | 23.8/27.6 |
| Proteins/Water             | 556/519 | 556/487 | 556/441 | 556/412 | 556/323 | 1112/697 |
| Carbohydrate (mol)         | 3 | 3 | 3 | 3 | 3 | 4 |
| RMSDs                      |     |     |     |     |     |    |
| Bond lengths (Å)           | 0.005 | 0.008 | 0.009 | 0.016 | 0.014 | 0.008 |
| Angles (degrees)           | 1.396 | 1.702 | 1.673 | 4.479 | 1.763 | 1.626 |
| Average B-factor (Å²)      | 50.3 | 64.5 | 59.0 | 55.8 | 51.3 | 51.9 |
| Ramachandran plot (%)      |     |     |     |     |     |    |
| Most favored region        | 80.3 | 84.5 | 83.7 | 83.9 | 83.5 | 82.8 |
| Additional allowed region  | 11.0 | 10.0 | 11.4 | 11.0 | 12.1 | 12.5 |
| Generously allowed region  | 2.2 | 2.9 | 2.3 | 2.0 | 2.1 | 2.7 |

| Values in parentheses are for the highest resolution shell. |

Isothermal titration calorimetry (ITC) analysis was performed to measure carbohydrate binding ability between full-length Tl-gal and carbohydrates (supplemental Fig. 7 and supplemental Table 2). Tl-gal bound to all carbohydrates and showed higher binding affinity with higher molecular weight LNT than other carbohydrates. The oligosaccharide epitope of the ganglio-series glycosphingolipids was embedded within higher molecular weight glycans (23). We found that LNT physically bound to Tl-gal with an apparent KD of 24 μM and that small molecular weight lactose showed the lowest binding affinity in this study. Glucose was bound relatively well to Tl-gal despite its small size. Tl-gal mutants (W77F/W212F and R61A/R196E) showed low binding affinity or no affinity compared with the binding of the wild-type Tl-gal-carbohydrate complexes based on the KD values (data not shown).

Carbohydrate Recognition Mechanism of Tl-gal—The direct binding of Tl-gal and carbohydrates is shown in Fig. 4 and supplemental Table 3. Carbohydrates interact with some amino acids of the binding sites discovered in the HXXKR and WGXER motifs on the CRDs. The carbohydrates interacted with Tl-gal on concave surfaces with charged or polar residues in the CRDs. The relative charge distribution of the concave surface is positive and basic. Tl-gal and lactose complex interacted through hydrogen bonds between lactose and polar amino acids His57, Asn70, and Trp77 in the NCRD and His192, Asn194, Asn205, and Trp121 in the CCRD. Moreover, oxygens of lactose interacted with negatively charged carboxylate groups of Glu80 and Glu215, which formed interactions with nitrogens of charged residues Arg61 and Arg82 in the NCRD and Arg196 in the CCRD. Positively charged Arg61 and Arg196 of Tl-gal have the ability to form multiple bonds for binding with negatively charged oxygens of lactose, which were stabilized strongly through binding with lactose. The interactions of Arg or Glu residues with carbohydrates are strongly electrostatic.

In the complex of Tl-gal and LacNAc, similar to the interaction between Tl-gal and lactose, nitrogen or oxygen of the charged residues Arg61/Arg196 and Glu80/Glu215 in Tl-gal formed strong multiple bonds with oxygen of LacNAc. In addition, Arg82 interacted weakly with LacNAc. Polar amino acids His57, Asn70, His192, Asn205, and Trp212 also interacted with the oxygen of LacNAc; however, Trp77 and Asn194 did not.

Tl-gal interacted most strongly with high molecular weight LNT among these carbohydrates. As in the lactose complex, polar residues containing hydrogen bonds in the Tl-gal and LNT complex were His57, Asn70, His192, Asn194, and Asn205. Trp77 and Trp212 did not participate in the interaction between Tl-gal and LNT. The charged residues of Arg61, Glu80, Arg196, and Glu215 play important roles through the strong binding with carbohydrates. In addition, Glu199 and Arg82 did participate in the interaction with LNT. The Tl-gal and SiaLac complex showed similar binding with lactose complex. Polar Ser29 residue participated in binding to SiaLac; however, the residues of Asn205 did not. Interestingly, in the case of Tl-gal and glucose complex, four glucoses did not bind to NCRD, but they did bind to CCRD through interactions with the His192, Asn194, Arg196, Asn205, Trp212, and Glu215 residues of Tl-gal.

As a result, Tl-gal can recognize five carbohydrates through interaction with the residues on the β4/β5/β6 and the loop region of β5-β6 in the NCRD and β4'/β5' and the loop region of β5'-β6' in the CCRD.

Conformational Changes in Tl-gal and Tl-gal-Carbohydrates and Structural Differences between NCRD and CCRD in Tl-gal—To investigate the conformational differences between wild-type Tl-gal and Tl-gal-carbohydrate complexes (lactose, LacNAc, LNT, SiaLac, and glucose), their structures were superimposed (Fig. 5, A–C). Their average root mean square deviations (RMSDs) were found to be 0.346, 0.259, 0.387, 0.175, and 0.280 Å, respectively. Superposition of wild-type Tl-gal and...
five Tl-gal-carbohydrate complexes showed differences in the loop regions of the CRDs (supplemental Fig. 2 and Table 4). The RMSDs of the wild-type Tl-gal and Tl-gal-carbohydrate complexes are 0.117, 0.146, 0.144, 0.161, and 0.204 Å in NCRD and 0.071, 0.101, 0.122, 0.136, and 0.169 Å in CCRD, respectively. The RMSDs of NCRDs between Tl-gal and Tl-gal with carbohydrates showed larger differences than those of the CCRDs. From the superpositions of these structures, we found that residues with RMSDs exceeding 0.8–3.5 Å were mainly the Asn^{44}–Gly^{11} and Glu^{218}–Lys^{220} residues. Residues having large RMSDs were primarily located in the loop regions between β3-β4 sheets on the NCRD and between β5′-β6′ sheets on the CCRD. In other words, large conformational changes caused by binding to carbohydrates were shown in the loop regions around the carbohydrate recognition regions of the CRDs. The binding sites of carbohydrates were almost congruous with each other in the five carbohydrate-binding complexes.

FIGURE 1. Structures of full-length Tl-gal and carbohydrates. A, schematic diagram showing the domains of Tl-gal (residues 1–278). B, sequence alignments of the NCRD and CCRD of the Tl-gal and the secondary structures are shown. Secondary elements are shown by an arrow (β-sheets) and coil (α-helix). The linker loops are shown as black lines. The residues conserved across two domains are indicated in red. Every 10 residues are indicated by a point. Similarity residues are indicated by blue boxes. The amino acids involved in the interactions with carbohydrates are indicated by stars. C, the chemical structures of lactose, LacNAc, LNT, SiaLac, and glucose are shown.
Crystal Structures of Tl-gal and Carbohydrates
The four carbohydrates (lactose, LacNAc, LNT, and SiaLac) were bound to both CRDs of Tl-gal. However, sugars were only bound to CCRD of T-gal, despite their small sizes. Three carbohydrate (lactose, LacNAc, and LNT) complexes showed similar binding residues and conformational changes as wild-type Tl-gal. Tl-gal complex with SiaLac showed little conformational change, except for the Ser46 residue (3.5 Å RMSD) (supplemental Fig. 6). In the complex of T-gal and glucose, the residues (Leu15–Pro18) at the long loop between the β1 and β2 sheets on the NCRD showed some RMSDs from wild-type Tl-gal, whereas other residues showed deviation patterns similar to those of other carbohydrates complexes. Because of this stability of CCRD, the glucose might be strongly bound to the CCRD in Tl-gal.

To understand the structural differences between NCRD and CCRD and confirm the conserved binding motif in the CRDs of Tl-gal, amino acid sequence alignment of the CRDs of Tl-gal was performed. This alignment revealed that NCRD and CCRD shared 37.1% homology. The amino acid sequences of Tl-gal with secondary structures are shown in Fig. 1B. The HXXR and WGXEER motifs were conserved in the CRDs involved with carbohydrate binding sites (Fig. 1B).

**Linker Domain and Molecular Packing of Tl-gal—**Tandem repeat type galectins (galectin-4, -6, -8, -9, and -12) have two CRDs in the N-terminal and C-terminal regions linked by about 30 peptides. This linker region is flexible and unstable, so conformational changes are often observed. We recently reported a novel crystal structure of full-length Tl-gal containing both CRDs (21). In this study, we determined the X-ray structure of wild-type full-length Tl-gal in a high resolution ligand-free form and its complexes with five different carbohydrates. Interestingly, the linker domain of Tl-gal was shorter than that of other tandem repeat type galectins. The conserved amino acids in the linker domain of galectins were shown, and surprisingly, there were no conformational changes in the linker domains between wild-type Tl-gal and carbohydrate complexes (supplemental Fig. 5). We found that the linker domain of Tl-gal is very stable. We expressed and purified NCRD or CCRD from Tl-gal, however, we could not successfully make crystals from these domains. This property is novel and indicates that the full-length structure of Tl-gal is more stable either domain of Tl-gal or other galectins alone.

Molecular packing involves reconstructing neighboring asymmetric units from the crystallographic experiment that produced the given structure. Molecular interactions of Tl-gal and carbohydrate complexes are shown in asymmetric units (Fig. 2H). In these structures, the Tl-gal complexes were tightly packed with symmetrically related Tl-gal molecules. The structure had different molecular packing compared with that of known tandem repeat type galectin-9.

**Effect on Mutagenesis of Tl-gal—**In the complexes of Tl-gal and carbohydrates, the arginine, glutamic acid, and tryptophan on CRDs were structurally important residues for binding of these complexes. To investigate the function of Arg, Glu, and Trp on the conserved and binding motifs of CRDs, mutagenesis (R61A/R196A, E80A/E215A, and W77F/W212F) was performed and confirmed by SDS-PAGE (Fig. 2I). The mutation of W77F/W212F induced exposure to the hydrophobic amino acid Phe. The crystal of the W77F/W212F and R61A/R196A mutation could not collect diffraction data and might have unstable geometry of the three-dimensional structure because of the mutation. The mutation of E80A/E215A induced the exposure of the charged amino acid Glu to the hydrophobic amino acid of Ala and caused protein aggregations. These mutations in Tl-gal might have broken electrostatic or polar interactions and induced a decrease in the secondary structure. They are also likely to interfere with proper folding of the mutant proteins and induce large conformational changes and alterations in the function of Tl-gal (supplemental Fig. 2).

**Protease Resistance of Tl-gal—**To determine the protease susceptibility of Tl-gal, trypsin treatment was applied for 12 h at 37 °C. Almost no wild-type full-length Tl-gal was degraded during trypsin treatment, presumably because of there being a short linker region, as shown in Fig. 3C. The short linker region of wild-type Tl-gal showed stability against protease degradation. However, two mutant forms, W77F/W212F and R61A/R196A, were almost entirely degraded by trypsin after 15 min at 37 °C (Fig. 3, D and E). The mutants were degraded more rapidly than the wild type and were unstable for trypsin. These results showed that the linker and carbohydrate binding regions of Tl-gal play important roles in the structural stability of Tl-gal protein.

**Hemagglutination Activity of Tl-gal Was Inhibited by Carbohydrate—**Galectins have hemagglutination activity caused by their carbohydrate binding activity. To determine whether five carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose) could affect Tl-gal-carbohydrate binding activity, the hemagglutination activity on human erythrocytes was examined. As shown in Fig. 6, the amounts of Tl-gal for the hemagglutination assay were 1–0.03 μg/well. At high concentrations of Tl-gal, hemagglutination activity increased; however, the activity was inhibited by the addition of carbohydrates. Tl-gal mutants (W77F/W212F and R61A/R196A) showed no activities on a hemagglutination assay and were not changed by the addition of carbohydrates.

**Elevation of IL-10 and TGF-β Production in Splenocytes by Tl-gal Was Inhibited after Adding Carbohydrates—**Expression of IL-10 and TGF-β, which was related to regulatory T-cells, was elevated by Tl-gal treatment in vivo and in vitro (3). To identify changes in the immunological function of
Crystal Structures of Tl-gal and Carbohydrates

A

Fluorescence intensity

Wavelength (nm)

B

Fluorescence intensity

Wavelength (nm)

C

Fluorescence intensity

Wavelength (nm)

D

Fluorescence intensity

Wavelength (nm)

E

full-length wild-type Tl-gal

M 0 min 5 min 15 min 30 min 1 hr 3 hrs 6 hrs 12 hrs

F

full-length double-mutant (W77F/W212F) Tl-gal

M 0 min 5 min 15 min 30 min 1 hr 3 hrs 6 hrs 12 hrs

G

full-length double-mutant (R61A/R196A) Tl-gal

M 0 min 5 min 15 min 30 min 1 hr 3 hrs 6 hrs 12 hrs
Tl-gal with or without added carbohydrates, IL-10 and TGF-β production of splenocytes was evaluated after the addition of carbohydrates. In splenocytes, IL-10 and TGF-β concentrations increased significantly in response to Tl-gal treatments but were inhibited by the addition of all carbohydrates. In addition, mutant Tl-gal treatments were inhibited with or without carbohydrates (Fig. 6).

Discussion

All galectins possess a conserved carbohydrate recognition domain, which enables interaction with other proteins or carbohydrates. Tl-gal is a galectin-9 homologue gene isolated from the canine gastrointestinal nematode parasite, T. leonina (3). We investigated the interaction and structural basis between full-length Tl-gal and five carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose). Full-length Tl-gal, composed of the NCRD (residues 1–142) and CCRD (residues 150–278), was held together via a short linker (143–149). The NCRD of Tl-gal had 11 β-strands and a small α-helix between β9 and β10, whereas the CCRD had 10 β-strands and no α-helix (Figs. 1B and 2A). The chemical structures and electron density maps of the carbohydrates are shown in Fig. 1C and supplemental Fig. 3. Based on alignment of the NCRD and CCRD of Tl-gal, conserved motifs (HXXXX and WGXEE) were confirmed in the CRDs (Fig. 1B). This alignment showed that the NCRD and CCRD were 37.1% homologous. In this study, the carbohydrates (lactose, LacNAc, LNT, and SiaLac) interacted with the conserved motifs of both CRDs and glucoses strongly bound only to the motifs on the CCRD of the Tl-gal. Tl-gal showed greater binding affinity with higher molecular weight LNT than other carbohydrates.

The overall structures and detailed binding sites of the Tl-gal and carbohydrate complexes were also revealed (Figs. 2 and 4). All carbohydrates (lactose, LacNAc, LNT, and SiaLac) were bound to concave surfaces surrounded by β4, β5, and β6 of the Tl-gal NCBD, but these were also bound to the surfaces formed by β4’, β5’, and the loop between β5’ and β6’ of the Tl-gal CCRD (Figs. 1B and 2). The binding sites formed concave surfaces with positive charges due to the presence of arginine on the CRDs in Tl-gal. Large conformational changes caused by binding to carbohydrates were observed in the loop regions on the β3-β4 and β5’-β6’ sheets neighboring the carbohydrate recognition regions of the CRDs. The binding sites of carbohydrates were almost congruous with each other in the five carbohydrate-binding complexes. The structural basis of wild-type full-length Tl-gal with carbohydrates suggests that Tl-gal functions as a novel anti-inflammation agent for patients with inflammatory bowel disease or autoimmune diseases, most notably Crohn’s disease.

The diffraction data of the mutants of W77F/W212F and R61A/R196A were almost congruous with each other in the five carbohydrate-binding complexes. The binding sites for the Tl-gal complexes were located in conserved motifs on the CRDs. Furthermore, the binding affinity of Tl-gal with carbohydrates was also measured using ITC analysis. In recent studies, galectin-9 homologue protein, may also play a critical role in human disease. Moreover, Tl-gal induces increase in the levels of TGF-β and IL-10 and can ameliorate intestinal inflammation (3). Five carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose) can affect Tl-gal-carbohydrate binding to carbohydrates were observed in the loop regions on the β3-β4 and β5’-β6’ sheets neighboring the carbohydrate recognition regions of the CRDs. The binding sites of carbohydrates were almost congruous with each other in the five carbohydrate-binding complexes. The structural basis of wild-type full-length Tl-gal with carbohydrates suggests that Tl-gal functions as a novel anti-inflammation agent for patients with inflammatory bowel disease or autoimmune diseases, most notably Crohn’s disease.

The diffraction data of the mutants of W77F/W212F and R61A/R196A at the carbohydrate binding sites were not successfully collected, and the mutant protein (E80A/E215A) of Tl-gal at the other carbohydrate binding sites was aggregated during purification by the misfolded protein. In this study, charged Arg⁶¹/Arg²⁹⁶ and Glu⁸⁰/Glu²¹⁵ were found to be crucial residues for protein folding and structure. In addition, polar His, Asn, and Trp were also important for the interaction with carbohydrates through hydrogen bonding. The binding between Tl-gal and LNT showed the largest fluorescence maximum shift and highest fluorescence intensity relative to wild-type Tl-gal. Tl-gal bound to all carbohydrates and showed higher binding affinity with higher molecular weight LNT than other carbohydrates. Tl-gal mutants (W77F/W212F and R61A/R196A) showed low binding affinity or no affinity compared with the binding of wild-type Tl-gal-carbohydrate complexes.

Alignment revealed that Tl-gal possessed a shorter linker domain (composed of seven amino acids) than other galectins (supplemental Fig. 5). In a recent study, the structure of the full-length mutant galectin-8 was determined, and the linker domain of mutant galectin-8 was replaced by only two amino acids, signifying that the length of the linker domain in galectins had an effect on the formation of crystals (24). Conversely, the short linker domain of Tl-gal was influenced by the structural stability. In this study, crystals of NCRD and CCRD of Tl-gal were not grown. Superposition of the linker domains in wild-type Tl-gal and the Tl-gal complexes with carbohydrates revealed that they were almost identical. The full-length Tl-gal also showed structural stability when analyzed using UV-visible spectra of the protein (supplemental Fig. 8). Based on these results, Tl-gal is suggested to be a novel anti-inflammation agent with improved immune efficacy, especially in an acidic intestinal environment.

In this study, structures of the full-length Tl-gal complexes with carbohydrates were determined for the first time, and the binding modes of the complexes were revealed. The binding sites for the Tl-gal complexes were located in conserved motifs on the CRDs. Furthermore, the binding affinity of Tl-gal with carbohydrates was also measured using ITC analysis. In recent studies, galectin-9 was found to play an important role in human disease (25); therefore, Tl-gal, a galectin-9 homologue protein, may also play a critical role in human disease. Moreover, Tl-gal induces increase in the levels of TGF-β and IL-10 and can ameliorate intestinal inflammation (3). Five carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose) can affect Tl-gal-carbohydrate binding.
Crystal Structures of Tl-gal and Carbohydrates

A

NCRD

Lactose

B

NCRD

Lactose

C

NCRD

LacNAc

D

CCRD

LacNAc

E

NCRD

LNT

F

CCRD

LNT

G

NCRD

Sialac

H

CCRD

Sialac

I

CCRD

Glucose
binding activity, and hemagglutination activity of TI-gal on human erythrocytes was inhibited by interaction with carbohydrates. Consequently, the results of this study will help to reveal the binding mechanism between TI-gal and carbohydrates and binding partners, which may contribute to the design of therapeutic strategies for human disease.

FIGURE 4. Interactions between TI-gal and carbohydrates. The surface and atomic representations of interactions between TI-gal and carbohydrates are shown. TI-gal NCRD-carbohydrates (lactose, LacNAc, LNT, and SiaLac) (A, C, E, and G). TI-gal CCRD-carbohydrates (lactose, LacNAc, LNT, SiaLac, and glucose) (B, D, F, H, and I). Relative distribution of the surface charge is shown with the acidic region in red, the basic region in blue, and the neutral region in white. Hydrogen bonds in TI-gal and carbohydrate complexes are shown as black dotted lines.

FIGURE 5. Superpositions of the TI-gal and TI-gal carbohydrate complexes. A, three complexes with lactose, LacNAc, and LNT are colored cyan, magenta, and yellow, respectively. The wild-type TI-gal is colored green. B, NCRD of three complexes (lactose, LacNAc, and LNT). C, CCRD of three complexes. D, NCRD after superposition of NCRD (red) and CCRD (blue) of TI-gal with lactose. E, CCRD after superposition of NCRD (red) and CCRD (blue) of TI-gal with lactose. The side chains of several binding amino acids of TI-gal are shown in green, along with lactose. F and G, ribbon diagram and C trace of superposition between wild-type TI-gal NCRD and CCRD.
Experimental Procedures

Expression, Purification, and Mutagenesis of Tl-gal—Full-length Tl-gal (amino acids 1–278) incorporated into pET-28a expression vector was transformed into E. coli BL21 (DE3). Recombinant protein expression was induced by isopropyl β-D-1-thiogalactopyranoside for 16 h at 25 °C (21). The cells were then harvested and resuspended in lysis buffer A (50 mM Tris-HCl (pH 8.0) and 200 mM NaCl). The supernatant was then sonicated and purified using a nickel-nitrilotriacetic acid column and gel filtration chromatography in conjunction with an FPLC system (GE Healthcare), after which the proteins were confirmed by SDS-PAGE. For site-directed mutagenesis of Tl-gal, double-stranded oligonucleotides were used (supplemental Table 1). PCR was then performed in a mixture containing template, each primer, dNTPs, and Pfu DNA polymerase. Next, the amplified DNA was digested with DpnI enzyme and trans-
formed into cells. Finally, the mutants of TI-gal were purified using the same steps as above.

**Crystallization of Wild-type TI-gal and TI-gal-Carbohydrate Complexes**—Protein crystallization was performed at 293 K by the hanging drop vapor diffusion method. The crystal of wild-type TI-gal was obtained by incorporation of equal volumes of the protein and reservoir solution. The complexes of TI-gal with carbohydrates were obtained by co-crystallization with equal volumes of the protein, reservoir solution, and 20 mM carbohydrates. The carbohydrates used in this study were lactose (Sigma-Aldrich), LacNAc (Sigma-Aldrich), LNT (Santa Cruz Biotechnology, Inc.), SiaLac (Santa Cruz Biotechnology), and glucose (Sigma-Aldrich). The crystal of the lactose complex became larger in a reservoir composed of polyethylene glycol 4000, 0.1 M NaHEPES, pH 7.5, and 0.1 M ammonium sulfate, whereas for the others, 10% polyethylene glycol 8,000, 100 mM sodium/potassium phosphate (pH 6.2), and 200 mM NaCl were used (supplemental Fig. 1). After about 5 days, wild-type rectangular crystals appeared and grew to maximum size within 2 weeks. The crystals of TI-gal with carbohydrates appeared after 10 days and grew within 3 weeks.

**Data Collection, Structure Determination, and Refinement**—Data for the wild-type TI-gal and TI-gal complexes with carbohydrates were collected at the Pohang Light Source, Republic of Korea. Briefly, crystals were soaked in a cryoprotectant solution containing precipitant and 20% glycerol and then mounted in a loop under a liquid nitrogen stream at 100 K. Diffraction data were subsequently collected and processed using HKL-2000 (26). The structures of TI-gal and carbohydrate complexes were then determined by the molecular replacement method using the AMoRe and EPMR programs (27, 28). In addition, the program O was employed to improve the model by iterative model building (29). Final refinements to R and Rfree of 22.3 and 25.8% for wild-type TI-gal and 23.7 and 26.8% for lactose complex became larger in a reservoir composed of polyethylene glycol 4000, 0.1 M NaHEPES, pH 7.5, and 0.1 M ammonium sulfate, whereas for the others, 10% polyethylene glycol 8,000, 100 mM sodium/potassium phosphate (pH 6.2), and 200 mM NaCl were used (supplemental Fig. 1). After about 5 days, wild-type rectangular crystals appeared and grew to maximum size within 2 weeks. The crystals of TI-gal with carbohydrates appeared after 10 days and grew within 3 weeks.

**Isothermal Titration Calorimetry**—For ITC analysis, 0.2 mM protein was loaded onto an ITC200 microcalorimeter (GE Healthcare) cell at 25 °C. The titration syringe was filled with 4 mM carbohydrate ligand solution solubilized in buffer A. The titration of TI-gal with carbohydrates was conducted using the same steps as above.

**Trypsin Treatment**—The wild-type full-length TI-gal and mutants of TI-gal were incubated at 37 °C for 5 min in digestion buffer containing 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 1 mM CaCl2. After trypsin treatment (0.36 μg), 36 μg of each protein was incubated at 37 °C for 12 h. The proteins were confirmed by SDS-PAGE.

**Fluorescence Spectroscopy Analysis**—Fluorescence spectroscopy analysis was performed using a SCINCO FluoroMate FS-2 with path length cuvettes of 1 cm and excitation and emission slits of 20 nm. The emission spectra were obtained at 305–465 nm with an excitation wavelength of 295 nm and a protein concentration of 5 μM at 25 °C.

**Hemagglutination Assay**—To determine the activity of TI-gal with or without carbohydrates, we conducted a hemagglutination assay as described previously (31). Briefly, the assay was performed in 96-well microtiter plates with serial 2-fold dilutions (1–0.03 μg/100 μl) of the TI-gal in PBS in the absence or presence of carbohydrates. Samples (75 μl) were mixed with a 2% (v/v) suspension of ABO blood type B erythrocytes (25 μl) and incubated at room temperature for 1 h. Agglutination activity was determined by the sedimentation state of the erythrocytes. Agglutinated erythrocytes formed a “mat” on the bottom of the well.

**Cytokine Analysis**—Splenocytes were isolated from the spleens of 6-week-old C56/BL6 mice. After sacrifice, the detached spleens were disrupted in 3 ml of RPMI 1640 medium containing 10% fetal bovine serum. Cells were obtained by centrifugation at 1,200 rpm for 3 min, after which 3 ml of ammonium-chloride-potassium hypotonic red blood cell lysis solution (Sigma-Aldrich) was added to the cells for 2 min. Next, 3 ml of RPMI 1640 medium was added, and the cells were centrifuged at 1,200 rpm for 3 min. Following centrifugation, 5 × 106 cells in 1 ml of RPMI 1640 medium were treated with 1 μg of TI-gal with or without carbohydrates for 72 h at 37 °C. The supernatants were then harvested, and IL-10 and TGF-β levels were determined via ELISA in accordance with the manufacturer’s instructions. The absorbance of the final reactant was determined at 450 nm using an ELISA plate reader.

**Author Contributions**—E. Y. H. performed the biological experiments, made the crystals, and conducted the data collection. S. K. P. and H. S. Y. conducted hemagglutination and cytokine analysis, and S. C. H. helped with data collection. S. B. J. and M. S. J. designed the research, solved the crystal structures, and analyzed the data. S. B. J., M. S. J., E. Y. H., and H. S. Y. wrote the manuscript.

**Acknowledgments**—We thank Dr. Sung Chul Ha at the Pohang Light Source (PLS-7A), Republic of Korea, for assistance with the X-ray diffraction experiments. This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology, Grants 2015R1D1A1A01059594 (to S. B. J.) and 2016R1D1A1B02011142 (to J. M. S.).

**References**

1. Lee, K. H., Park, H. K., Jeong, H. J., Park, S. K., Lee, S. J., Choi, S. H., Cho, M. K., Ock, M. S., Hong, Y. C., and Yu, H. S. (2008) Immunization of proteins from *Toxascaris leonina* adult inhibits allergic specific Th2 response. *Vet. Parasitol.* 156, 216–225
2. Reddy, A., and Fried, B. (2007) The use of *Trichurus suis* and other helminth therapies to treat Crohn’s disease. *Parasitol. Res.* 100, 921–927
3. Kim, J. Y., Cho, M. K., Choi, S. H., Lee, K. H., Ahn, S. C., Kim, D. H., and Yu, H. S. (2010) Inhibition of dextran sulfate sodium-induced intestinal inflammation via enhanced IL-10 and TGF-β production by galecin-9 homologues isolated from intestinal parasites. *Mol. Biochem. Parasitol.* 174, 53–61

**Crystal Structures of TI-gal and Carbohydrates**
Crystal Structures of Tl-gal and Carbohydrates

4. Türeci, O., Schmitt, H., Fadle, N., Pfreundschuh, M., and Sahin, U. (1997) Molecular definition of a novel human galectin which is immunogenic in patients with Hodgkin’s disease. J. Biol. Chem. 272, 6416–6422

5. Matsumoto, R., Matsumoto, H., Seki, M., Hata, M., Asano, Y., Kanegasaki, S., Stevens, R. L., and Hirashima, M. (1998) Human ecallactin, a variant of human galectin-9, is a novel eosinophil chemoattractant produced by T lymphocytes. J. Biol. Chem. 273, 16976–16984

6. Wada, J., Ota, K., Kumar, A., Walner, E. L., and Kanwar, Y. S. (1997) Developmental regulation, expression, and apoptotic potential of galectin-9, a β-galactoside binding lectin. J. Clin. Invest. 99, 2452–2461

7. Hughes, R. C. (1999) Secretion of the galectin family of mammalian carbohydrate-binding proteins. Biochim. Biophys. Acta 1473, 172–185

8. Yang, R. Y., Rabinovich, G. A., and Liu, F. T. (2008) Galectins: structure, function and therapeutic potential. Expert Rev. Mol. Med. 10, e17

9. Bi, S., Earl, L. A., Jacobs, L., and Baum, L. G. (2008) Structural features of human galectin-9, a novel eosinophil chemoattractant produced by T cells. J. Biol. Chem. 283, 12248–12258

10. Li, Y., Feng, J., Geng, S., Geng, S., Wei, H., Chen, G., Li, X., Wang, L., Wang, R., Peng, H., Han, G., Shen, B., and Li, Y. (2011) The N- and C-terminal carbohydrate recognition domains of galectin-9 contribute differently to its multiple functions in innate immunity and adaptive immunity. Mol. Immunol. 48, 670–677

11. Heusschen, R., Griffioen, A. W., and Thijssen, V. L. (2013) Galectin-9 in tumor biology: a jack of multiple trades. Biochim. Biophys. Acta 1836, 177–185

12. Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W. E., Yagi, F., and Kasai, K. (2002) Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. Biochim. Biophys. Acta 1572, 232–254

13. Thijssen, V. L., and Griffioen, A. W. (2014) Galectin-1 and -9 in angiogenesis: a sweet couple. Glycobiology 24, 915–920

14. Chabot, S., Kashio, Y., Seki, M., Shirato, Y., Nakamura, T., Kishi, N., Nakamura, T., Matsumoto, R., and Hirashima, M. (2002) Regulation of galectin-9 expression and release in Jurkat T cell line cells. Glycobiology 12, 111–118

15. Matsushita, N., Nishi, N., Seki, M., Matsumoto, R., Kuwabara, J., Liu, F. T., Hata, Y., Nakamura, T., and Hirashima, M. (2000) Requirement of divalent galactoside-binding activity of ecallactin/galectin-9 for eosinophil chemoatraction. J. Biol. Chem. 275, 8355–8360

16. Sato, M., Nishi, N., Shoji, H., Seki, M., Hashidate, T., Hirabayashi, J., Kasai Ki, K., Hata, Y., Suzuki, S., Hirashima, M., and Nakamura, T. (2002) Functional analysis of the carbohydrate recognition domains and a linker peptide of galectin-9 as to eosinophil chemoattractant activity. Glycobiology 12, 191–197

17. Aguzzi, A., and O’Connor, T. (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. Nat. Rev. Drug Discov. 9, 237–248

18. Stefani, M., and Dobson, C. M. (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J. Mol. Med. 81, 678–699

19. De Felice, F. G., Vieira, M. N., Meirelles, M. N., Morozova-Roche, L. A., Dobson, C. M., and Ferreira, S. T. (2004) Formation of amyloid aggregates from human lysozyme and its disease-associated variants using hydrostatic pressure. FASEB J. 18, 1099–1101

20. Tanzi, R. E., and Bertram, L. (2005) Twenty years of the Alzheimer’s disease amyloid hypothesis: a genetic perspective. Cell 120, 545–555

21. Jeong, M. S., Hwang, H. G., Yu, H. S., and Jang, S. B. (2013) Structure of full-length Toxascaris leonina galectin with two carbohydrate-recognition domains. Acta Crystallogr. D Biol. Crystallogr. 69, 168–175

22. Iglesias, M. M., Rabinovich, G. A., Ivanovic, V., Sotomayor, C., and Wollenstein-Todel, C. (1998) Galectin-1 from ovine placenta: amino-acid sequence, physicochemical properties and implications in T-cell death. Eur. J. Biochem. 252, 400–407

23. Collins, P. M., Bum-Erdene, K., Yu, X., and Blanchard, H. (2014) Galectin-3 interactions with glycosphingolipids. J. Mol. Biol. 426, 1439–1451

24. Yoshida, H., Teraoka, M., Nishi, N., Nakakita, S., Nakamura, T., Hirashima, M., and Kamitori, S. (2010) X-ray structures of human galectin-9 C-terminal domain in complexes with a biantennary oligosaccharide and sialyllactose. J. Biol. Chem. 285, 36969–36976

25. Wiersma, V. R., de Bruyn, M., Helfrich, W., and Bremer, E. (2013) Therapeutic potential of galectin-9 in human disease. Med. Res. Rev. 33, E102–E126

26. Otwinowski, Z., and Mino, r. W. (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326

27. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search. Acta Crystallogr. D Biol Crystallogr. 55, 484–491

28. Navaza, J. (1994) AmoRe: an automated package for molecular replacement. Acta Crystallogr. A 50, 157–163

29. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47, 110–119

30. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol Crystallogr. 54, 905–921

31. Wang, J., Yan, R., Xu, L., and Li, X. (2007) The second glutamic acid in the C-terminal CRD affects the carbohydrate-binding properties of recombinant galectins of Haemonchus contortus. Vet. Parasitol. 148, 247–255