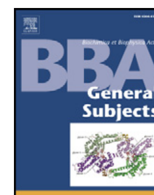
Biochimica et Biophysica Acta 1780 (2008) 1131–1142

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen



Caenorhabditis elegans galectins LEC-1–LEC-11: Structural features and sugar-binding properties

Yoko Nemoto-Sasaki [a,*], Ko Hayama [b], Hiroyuki Ohya [a], Yoichiro Arata [a,1], Mika Kato Kaneko [c], Naruya Saitou [d], Jun Hirabayashi [b], Ken-ichi Kasai [a]

[a] *Department of Biological Chemistry, Teikyo University School of Pharmaceutical Sciences, 1091-1 Suarashi, Sagamiko, Sagamihara, Kanagawa 229-0195, Japan*
[b] *Lectin Application and Analysis Team, Research Center for Medical Glycoscience (RCMG), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8568, Japan*
[c] *Glycogene Function Team, Research Center for Medical Glycoscience (RCMG), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8568, Japan*
[d] *Division of Population Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*

ARTICLE INFO

Article history:
Received 5 March 2008
Received in revised form 28 June 2008
Accepted 7 July 2008
Available online 14 July 2008

Keywords:
Caenorhabditis elegans
Galectin
Lectin
Frontal affinity chromatography
Carbohydrate-binding property

ABSTRACT

Galectins form a large family of β-galactoside-binding proteins in metazoa and fungi. This report presents a comparative study of the functions of potential galectin genes found in the genome database of *Caenorhabditis elegans*. We isolated full-length cDNAs of eight potential galectin genes (*lec-2–5* and *8–11*) from a λZAP cDNA library. Among them, *lec-2–5* were found to encode 31–35-kDa polypeptides containing two carbohydrate-recognition domains similar to the previously characterized *lec-1*, whereas *lec-8–11* were found to encode 16–27-kDa polypeptides containing a single carbohydrate-recognition domain and a C-terminal tail of unknown function. Recombinant proteins corresponding to *lec-1–4*, *-6*, and *8–10* were expressed in *Escherichia coli*, and their sugar-binding properties were assessed. Analysis using affinity adsorbents with various β-galactosides, i.e., *N*-acetyllactosamine (Galβ1-4GlcNAc), lacto-*N*-neotetraose (Galβ1-4GlcNAcβ1-3Galβ1-4Glc), and asialofetuin, demonstrated that LEC-1–4, -6, and -10 have a significant affinity for β-galactosides, while the others have a relatively lower affinity. These results indicate that the integrity of key amino acid residues responsible for recognition of lactose (Galβ1-4Glc) or *N*-acetyllactosamine in vertebrate galectins is also required in *C. elegans* galectins. However, analysis of their fine oligosaccharide-binding properties by frontal affinity chromatography suggests their divergence towards more specialized functions.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Galectins are a group of lectins that show β-galactoside-binding activity and have at least one evolutionarily conserved carbohydrate-recognition domain (CRD) [1]. They were first found in the electric organ of the electric eel in 1975 [2] and then identified in a wide range of species, including invertebrates and fungi [3–6]. The first invertebrate galectin was identified in the primitive model animal *Caenorhabditis elegans*, a nematode, as a 32-kDa protein (N32, LEC-1) consisting of two CRDs. The CRDs showed 25–35% amino acid-sequence similarities to vertebrate galectin CRDs. This was also the first study on a "tandem-repeat type" galectin [7]. Later, another nematode galectin (N16, LEC-6) consisting of two identical 16-kDa subunits was isolated from the worm, classified as a "proto-type," and characterized [8].

Thus far, 14 members of the galectin family have been identified in mammals (e.g., human, mouse, rat, bovine, etc.) and named galectins-1–10 and 12–15 in the order of identification. The name galectin-11 is not used at present [9]. Galectins are categorized into three structural types on the basis of the global architecture (not sequence similarity), i.e., "proto" (galectins-1, -2, -5, -7, -10, -13, -14, and -15), "chimera" (galectin-3), and "tandem-repeat" types (galectins-4, -6, -8, -9, and -12) [9,10]. Galectins that have been characterized thus far have the following properties in common: (1) they bind β-galactosides, e.g., *N*-acetyllactosamine (LacNAc; Galβ1-4GlcNAc); (2) they are soluble proteins with no apparent membrane-anchoring region; (3) they do not require metal ions for sugar binding; and (4) they have characteristics distinctive of typical cytoplasmic proteins, i.e., N-terminal acetylation, absence of a signal sequence for secretion, no glycosylation, preservation of free cysteine residue(s) rather than formation of disulfide bond(s) [4].

Abbreviations: CRD, carbohydrate-recognition domain; FAC, frontal affinity chromatography; Lac, lactose; LacNAc, *N*-acetyllactosamine; LN2, 3, and 5, *N*-acetyllactosamine oligomers $(Gal\beta1\text{-}4GlcNAc)_2$, $(Gal\beta1\text{-}4GlcNAc)_3$, and $(Gal\beta1\text{-}4GlcNAc)_5$, respectively; LNT, lacto-*N*-tetraose; LNnT, lacto-*N*-neotetraose; LNFP-I, II, and III, lacto-*N*-fucopentaose I, II, and III, respectively; LNDFH, lacto-*N*-difucohexaose; NA2, 3, and 4, biantennary, triantennary, and tetraantennary asialo type *N*-glycans, respectively; ORF, open reading frame; PA, pyridylaminated

* Corresponding author. Tel.: +81 42 685 3741; fax: +81 42 685 3742.
E-mail address: ynsasaki@pharm.teikyo-u.ac.jp (Y. Nemoto-Sasaki).
[1] Present address: Laboratory of Molecular Immunology, Faculty of Pharmaceutical Sciences, Josai University, Sakado 350-0295, Japan.

0304-4165/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagen.2008.07.003

With regard to structure and function, extensive studies on vertebrate galectins have established the presence of eight conserved amino acids (His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73; the numbers correspond to those of human galectin-1) that play a crucial role in the binding of sugars [11–18]. Therefore, it is of evolutionary interest to know whether this holds true in the case of invertebrate galectins. It is also important from a comparative viewpoint to know the divergence of the galectin proteins in invertebrates.

Based on the above concept, we initiated "comparative galectinomics" focusing on the nematode *C. elegans* (some data has been presented previously, [19,20]). The *C. elegans* genome [21] contains a significant number of potential galectin genes, as previously suggested [22,23]. Thus far, however, it has not been investigated from a comparative standpoint to determine whether all of the "potential" galectin genes are actually expressed *in vivo* and function as β-galactoside-binding lectins. In this study, we targeted 11 potential *C. elegans* galectin genes (designated *lec-1* to *lec-11*), including those previously characterized (*lec-1* and *lec-6*), for further functional analysis. The analysis included the isolation of full-length cDNA clones to compare their amino acid sequences and production of recombinant proteins to assess their sugar-binding properties. We isolated full-length cDNA clones, except for *lec-7*. Although we failed to produce a recombinant protein of LEC-5, we successfully produced recombinant proteins of LEC-2–4 and LEC-8–11. By using affinity chromatography, we showed that LEC-2–4 and -10 are able to bind β-galactosides. Thus, it became evident that most potential *C. elegans* galectin genes (*lec-1–4*, *-6*, and *-10*) are functional. Moreover, LEC-8, -9, and -11 showed extremely weak affinities for the β-galactosides examined thus far, though they might have affinities for other carbohydrates besides those used in this study. The results clearly indicated that β-galactoside-binding properties have been strongly conserved across species.

2. Materials and methods

2.1. Database search for the C. elegans galectin-like genes

The TBLASTN search was performed for potential galectin genes in the *C. elegans* genome database with amino acid sequences of LEC-1 (GenBank accession number: M94671) and LEC-6 (D63575) as queries. Among the genes showing significant scores (E values<0.1), those encoding polypeptides containing at least four of the eight critical amino acids (His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73; residue numbers correspond to those of human galectin-1) were considered for further analysis.

2.2. Reverse transcriptase-polymerase chain reaction

Total RNA was prepared from 0.2 g (wet weight) of wild-type Bristol N2 strain (mixed stages) by using the SV Total RNA isolation system (Promega), and reverse transcription (RT) was performed with a 1st-strand™ cDNA synthesis kit (Clontech). Appropriate sets of internal primers were designed to amplify the DNA fragments of approximately 0.2 kbp (for details, see Supplementary Table 1). Polymerase chain reaction (PCR) was performed by using LA-*Taq* DNA polymerase (Takara Bio), and the cycling conditions were as follows: 92°C, 30 s; 62°C, 30 s; and 72°C, 1 min. Thereafter, a final extension was conducted at 72°C for 10 min.

2.3. Screening of cDNAs of lec-2–5 and 7–11

The amplified 0.2-kbp cDNA fragments were used as probes to screen a previously described λZAP cDNA library [24] after labeling with [α-^{32}P] dCTP using the Megaprime™ DNA labeling system (Amersham). Plaque hybridization was performed either independently (*lec-2*, *-10*, and *-11*) or in groups (*lec-3–5* and *7–9*) by mixing relevant probes. In vivo excision for conversion to the plasmid pBluescript was performed as previously described [25]. Colony PCR was performed to screen positive clones. These positive clones were subjected to sequence analysis by a conventional dideoxy method as described previously [8]. To determine the trans-spliced leader sequence in the 5′-untranslated region, nested PCR was performed as described previously [26].

2.4. Prediction of hydrophobic regions and signal peptides

For prediction of hydrophobic regions and signal peptides, the SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/), the SignalP ver. 3.0 (http://www.cbs.dtu-dk/services/SignalP/), and the PSORT II programs (http://psort.ims.u-tokyo.ac.jp/) were used.

2.5. Phylogenetic analysis of LEC-1–11 and DC2.3a

The Clustal W program was used for multiple alignments [27] of the LEC-1–11 and DC2.3a amino acid sequences, and the neighbor-joining method [28] was used to construct a phylogenetic tree as described previously [29]. To validate the confidence levels in the constructed tree, bootstrap analysis was performed. By bootstrap analysis, a total of 1000 trees were generated from the initial data set, and the percentage of trees containing a particular clade was measured.

2.6. Production of recombinant proteins LEC-2–4 and 8–11

Full-length coding regions of the *lec-2*, *-3b*, *-4*, *-8*, *-9*, *-10*, and *-11* genes were amplified by a standard PCR procedure. A plasmid containing a full-length cDNA of each galectin-like gene was used as a template; the primers used are described in Supplementary Table 2. In the case of *lec-8*, a truncated mutant LEC-8ΔC, which lacks a C-terminal tail (residues 142–180), was also constructed by using lec-8F and lec-8ΔCR primers (Supplementary Table 2). Expression of recombinant proteins was performed as described previously [30,31]. Briefly, amplified PCR fragments were ligated into a pET21a or pET21d vector (Novagen) that was digested with relevant restriction enzymes (i.e., NdeI/XhoI or NdeI/BamHI). The resultant expression constructs were used to transform *Escherichia coli* BL21 (DE3) cells. Production of recombinant proteins was induced by the addition of isopropyl-β-D-thiogalactoside.

2.7. Preparative affinity chromatography on asialofetuin-agarose

Preparative asialofetuin-agarose chromatography was performed as described previously [30,31]. Briefly, after induction of recombinant proteins (37°C for 4 h or 25°C for 16 h), *E. coli* cells were harvested by centrifugation and disrupted by sonication (30 min, on ice) with five volumes of MEPBS (4 mM β-mercaptoethanol, 2 mM EDTA, 20 mM sodium phosphate, pH 7.2, 150 mM NaCl). After centrifugation (15,000 rpm, 4°C, 25 min), the derived supernatant solution was applied to an asialofetuin-agarose column (10 ml bed volume, 9 mg fetuin immobilized per ml of agarose gel). After extensive washing of the column, the adsorbed protein was eluted with MEPBS containing 20 mM lactose. The protein concentration was determined and each fraction was subjected to analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using a 14% gel. In the case of experiments for LEC-8–11, Tris-buffered saline (TBS; 50 mM Tris–HCl, pH 7.5, 150 mM NaCl) was used instead of MEPBS. For recombinant proteins, which have a relatively weak affinity for asialofetuin, i.e., LEC-4, -8, -9, and -11, retarded fractions containing these proteins were pooled, concentrated by ultrafiltration using the Millipore Centricon Plus 20 (PL-10, 10,000 Da), and subjected to rechromatography on the same column. In the cases of LEC-8, -10, and -11, the use of metal-chelate

affinity chromatography proved to be useful (Hayama, K.; unpublished data).

2.8. Frontal affinity chromatography

The principle and basic procedures of frontal affinity chromatography (FAC) have been described previously [31–36];

$$K_d = B_t/V_f - V_0 - [A]_0 \quad (1)$$

where K_d is the dissociation constant (M) between the analyte (A in this study; either recombinant galectins or pyridylaminated-oligosaccharides (PA-oligosaccharides)) and immobilized ligand (B in this study; either glycans or recombinant galectins). B_t is the total amount of effectively immobilized ligand (mol), whereas $[A]_0$ is the initial concentration of the analyte A (M). V_f is the elution volume of A (ml), which can be determined as described previously [34] by monitoring the fluorescence signal based on either tryptophan (excitation and emission wavelengths, 285 and 350 nm, respectively) if A is a recombinant galectin or PA-oligosaccharides (excitation and emission wavelengths, 310 nm and 380 nm, respectively) if A is a PA-oligosaccharide. Under the assumption that $[A]_0$ is small relative to K_d, Eq. (1) is simplified into Eq. (2):

$$K_d = B_t/(V_f - V_0) \quad (2)$$

For analytical affinity chromatography, three types of column (4×10 mm; column volume, 0.126 ml) packed with either asialofetuin-agarose (2 mg/ml gel), Synsorb 34 (*N*-acetyllactosamine, LacNAc; Galβ1-4GlcNAc), or Synsorb 101 (lacto-*N*-neotetraose, L*N*nT; Galβ1-4GlcNAcβ1-3Galβ1-4Glc) were purchased from Wako Pure Chemicals, Tokyo, Japan. Binding of the purified recombinant proteins LEC-1, -2, -3b, -4, -6, -8ΔC, -9, and -10 to these columns was analyzed by FAC. FAC was carried out as previously described [33,34]. Briefly, 2 ml of purified recombinant galectin solution was dissolved in EDTA-PBS at a concentration of 1 mg/ml and applied to the columns. The elution of the protein was monitored as described above. The flow rate and the column temperature were kept at 0.25 ml/min and 20°C, respectively.

To analyze the fine oligosaccharide-binding properties, purified recombinant proteins of LEC-2, -3b, -4, -8ΔC, -9, and -10 were immobilized on HiTrap *N*-hydroxysuccinimide-activated (NHS-activated) Sepharose cartridge columns (1 ml; GE Healthcare Biosciences) according to the manufacturer's instructions at appropriate concentrations (1–7 mg/ml gel), as previously described [31]. PA-oligosaccharides were obtained as described previously [35]. The association constants between the lectin-immobilized agarose beads and PA-oligosaccharides were determined by the above FAC equation, as previously described [31].

3. Results

3.1. Novel galectin-like genes in C. elegans

The presence of many galectin-like genes other than *lec-1* (originally named N32) [7] and *lec-6* (similarly named N16) [8] in *C. elegans* was first implicated by Cooper, D.N.W. (University of California, San Francisco; personal communication) [22,23]. Our TBLASTN analysis using the *lec-1* sequence as the query resulted in the identification of a number of genes with significant *E* values (<0.1): W09H1.6 (=*lec-1*), F52H3.7, ZK892.1, C44F1.3, ZK1248.16, C53D6.7, DC2.3, R07B1.2, R07B1.10, C16H3.2, W01A11.4, T02G6.7, F41D3.6, F38A5.3, F46A8.4, F49F1.11, F49F1.10, F49F1.9, and F46A8.3, where W09H1.6 encodes N32. When the *lec-6* sequence was used as the query, the identified genes largely overlapped: i.e., Y55B1AR.1 (=*lec-6*), ZK892.1, F52H3.7, W01A11.4, W09H1.6, R07B1.10, C16H3.2, C44F1.3, ZK1248.16, DC2.3, F41D3.6, F46A8.3, F49F1.11, F46A8.4, F49F1.10, T02G6.7, R07B1.2, F49F1.9, and F38A5.3, where Y55B1AR.1 encodes N16. Since the overall sequence similarity between *lec-1* and *lec-6* is relatively low (~25% identity), the observed overlap is meaningful. These galectin-like genes are classified under the protein family (Pfam, http://pfam.sanger.ac.uk/) with the identifier PF00337.

Among the extracted candidate genes, we carefully examined the preservation of key amino acids, i.e., His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73 (numbers correspond to those of human galectin-1), that likely form the sugar-binding site of the *C. elegans* galectins. On the basis of this assumption, we chose the following nine genes for further analysis: F52H3.7, ZK892.1, C44F1.3, ZK1248.16, R07B1.2, R07B1.10, C16H3.2, W01A11.4, and F38A5.3. According to GeneFinder prediction, F52H3.7, ZK892.1, C44F1.3, and ZK1248.16 were strongly suggested to encode tandem-repeat-type galectins like LEC-1, and were thus designated as *lec-2–5*, respectively. On the other hand, R07B1.2, R07B1.10, C16H3.2, W01A11.4, and F38A5.3 were predicted to encode a single galectin CRD and were therefore designated as *lec-7–11* (Supplementary Table 3). We excluded DC2.3 as a candidate from the cDNA cloning study because of errors in the predicted sequence. However, an updated prediction revealed that this gene met the criteria, and the critical amino acids were found to be conserved. Therefore, we used this sequence data for comparison with the cloned cDNAs.

3.2. mRNA expression and cDNA cloning of lec-2–5 and 7–11

mRNA expression of the candidate galectin genes (*lec-2–5* and *7–11*) was examined by RT-PCR. Transcripts of *lec-2*, *-4*, *-5*, *-8*, *-9*, *-10*, and *-11* were detectable after 30 cycles. However, transcripts of *lec-3* and *lec-7* were not detectable even after 35 cycles (Fig. 1).

After confirming the expression of the candidate genes, we attempted to isolate cDNA clones by screening a λ phage library previously used [7,8]. The 0.2-kbp PCR fragments obtained were radiolabeled and used as probes. Full-length cDNAs were successfully obtained for *lec-2*, *-4*, *-5*, *-8*, *-9*, *-10*, and *-11* using an initial phage suspension containing 6×10^5 phages. However, no positive clone was detected despite repeated trials for *lec-3* and *lec-7*. In case of *lec-3*, however, the entire coding region predicted for this gene could be amplified by direct RT-PCR using lec-3F and lec-3R primers (Supplementary Table 2; data not shown). Further, no amplification product was obtained for *lec-7* even after 35 cycles of amplification.

Many mRNAs of *C. elegans* are known to attach to the trans-spliced leader sequence at the 5′-end [37–39]. In our previous study, the *lec-1* transcript has been shown to possess the 22-nucleotide consensus trans-spliced leader sequence "SL1" at positions −52 to −31 (where "A" of the initiation codon ATG indicates position 1) [26]. Nucleotide sequencing of *lec-2*, *-4*, and *-5* cDNAs showed that all of them

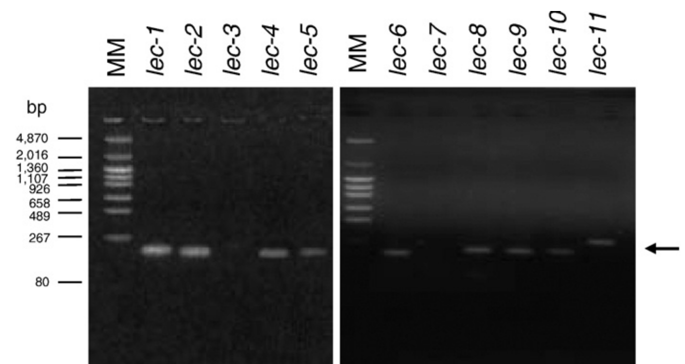


Fig. 1. RT-PCR analysis of *lec-1–11*. To confirm the mRNA expression of the *C. elegans* potential galectin genes *lec-1–11*, forward and reverse primers were designed for each gene, as described in Materials and methods. Approximately 0.2 kbp products were analyzed after 35 PCR cycles. The arrow indicates the position of the 0.2-kbp amplified products.

contained an SL1 sequence in their 5′-non-coding regions (Supplementary Fig. 1). RT-PCR with two nested reverse (antisense) primers specific for either *lec-3*, *-6*, *-8*, *-9*, *-10*, or *-11* and either SL1 or SL2 primers revealed that *lec-3*, *-6*, *-9*, *-10*, and *-11* cDNAs contain the SL1 sequence (Supplementary Fig. 1). On the other hand, we failed to detect the trans-spliced leader sequence of *lec-8*. Since the obtained *lec-8* cDNA clone had only a 17-bp 5′-non-coding region (Supplementary Fig. 1), it is possible that *lec-8* is also subjected to trans-splicing.

On the basis of the established cDNA sequences for *lec-1–6* and *8–11*, genome organizations of these genes have been provided in Supplementary Fig. 2. Exon/intron boundaries were in accordance with the consensus "GT/AG" rule (data not shown).

Except for *lec-3* (described below), all the cloned galectin-like gene transcripts were unique, and thus, each encoded only one distinct amino acid sequence. Nucleotide sequences of *lec-2–5* and *8–11* cDNAs, including the SL1 sequence, as well as the deduced amino acid sequences were registered in GenBank DNA databases. Accession numbers are shown in Supplementary Table 3.

3.3. Splicing variants of lec-3

As described above, *lec-3* cDNA was directly amplified by RT-PCR by using forward (lec-3F) and reverse primers (lec-3R) designed on the basis of the GeneFinder prediction. However, 5′- rapid amplification of cDNA ends (RACE) analysis revealed that the *lec-3* gene has an additional splicing variant, designated *lec-3b* (the former transcript is named *lec-3a* hereafter). Except for the difference in the 5′-region, both transcripts had the same sequence (Supplementary Fig. 1). These splicing variants encoded polypeptides that differed only in their N-terminal amino acid sequences, i.e., MAEPKSF in LEC-3a and MPSVWHCAKYI in LEC-3b (Supplementary Fig. 1).

3.4. Deduced amino acid sequences of LEC-1–11 and DC2.3a

Amino acid sequences deduced from cloned cDNA structures are shown in Fig. 2. The cDNAs of *lec-2*, *-3a*, *-3b*, *-4*, and *-5* encode tandem-repeat-type polypeptides, consisting of 278, 297, 301, 283, and 314 amino acids, respectively, similar to *lec-1*. The cDNA of *lec-9* encodes a mono-CRD type polypeptide, similar to that encoded by *lec-6*, with 140 amino acids. On the other hand, the cDNAs of *lec-8*, *-10*, and *-11* encode 180, 192, and 232 amino acid proteins, respectively. They have an N-terminal CRD and an unknown functional region at the C-terminal, and thus, they are considered to be novel chimera-type galectin-like proteins. Although LEC-9 lacks a C-terminal tail, we consider LEC-9 to be a member of the chimera-type galectin-like proteins based on the amino acid similarity (See Discussion). The

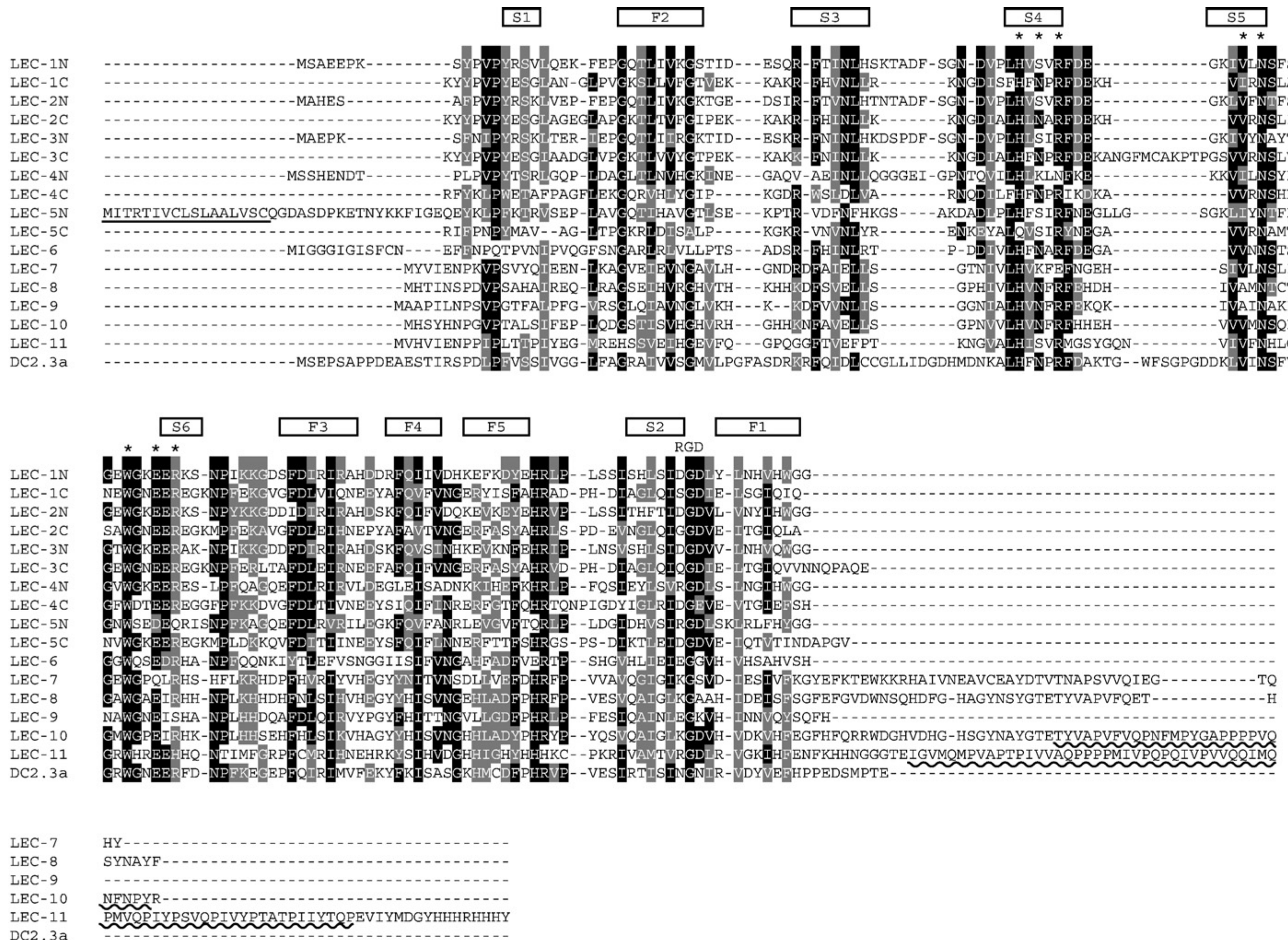


Fig. 2. Amino acid-sequence alignments of LEC-1–11 and DC2.3a. Amino acid sequences deduced from the cDNA structures of *lec-1–11* were aligned to produce maximal identity by the Clustal W program. Tandem-repeat domains of LEC-1–5 are separately aligned. Sequences of LEC-7 and DC2.3a reported by the *C. elegans* sequencing consortium have also been included for comparison. Highly conserved and similar amino acids are shown as white characters on a black and shaded background, respectively, using BOXSHADE program. The eight highly conserved residues (H, N, R, V, N, W, E, and R), which have been shown to play an important role in the binding of sugars, are denoted by asterisks. The position of a consensus tripeptide for adhesion in LEC-4N, LEC-5N, and LEC-11 is denoted as "RGD." The potential signal peptide of LEC-5 is underlined. The positions of the β-strands forming two β-sheets (S1–S6 and F1–F5) of the N-terminal CRD of LEC-1 based on the X-ray crystal structure [44] are shown at the top of the sequences. Hydrophobic segments observed in C-terminal tail regions of LEC-10 and -11 are underlined with wavy lines.

deduced amino acid sequences of LEC-7 and DC2.3a registered by the *C. elegans* sequencing consortium are also included for comparison in Fig. 2.

3.5. Predicted signal peptide of LEC-5

Galectins identified to date do not have any signal peptides, except for a few exceptions such as LEC-5 [23]. Moreover, Kaji et al. [40] and Fan et al. [41] recently identified LEC-5 as an *N*-linked glycoprotein. Therefore, we searched for signal peptides and transmembrane regions within the deduced amino acid sequences of the cloned galectin-like genes using the programs SOSUI, PSORT, and SignalP. The SOSUI program predicted that LEC-1–11 are soluble proteins. The PSORT program predicted that LEC-5 had a signal peptide, and the predicted cleavage site was between Cys17 and Gln18 at the N-terminal. No other genes examined were predicted to have signal peptides. The same results were obtained by SignalP-NN analysis. The SignalP-HMM program also predicted that LEC-5 had a signal peptide, but the predicted cleavage site was between Gly19 and Asp20.

3.6. Phylogenetic relationship of LEC-1–11 and DC2.3a

The deduced amino acid sequences of LEC-1–11 and DC2.3a were aligned by Clustal W and a phylogenetic tree was constructed (Fig. 3). The N-terminal and C-terminal CRDs of LEC-1–5 were separately analyzed. The polypeptides were classified into three groups: (i) tandem-repeat-type galectin-like proteins LEC-1, -2, -3, -4, and -5; (ii) LEC-6, -7, -8, -9, -10, and DC2.3a; and (iii) LEC-11. We constructed the phylogenetic tree using the *lec-11* gene as an outgroup. In the subgroup (i), the N-terminal and C-terminal CRDs were separately clustered. This result indicates that these CRDs were generated tandemly in the ancestral gene of this subgroup. After divergence of the *lec-1*, *-2*, *-3*, *-4*, and *-5* genes, *lec-1*, *-2*, and *-3* showed relatively slower evolutionary rates than those of the *lec-4* and *-5* genes. The topologies of the N-terminal and C-terminal CRDs of LEC-1, -2, and -3 are not the same in the tree. However, the bootstrap value at the node connecting the N-terminal CRD of LEC-1, -2, and -3 is very low (35%)

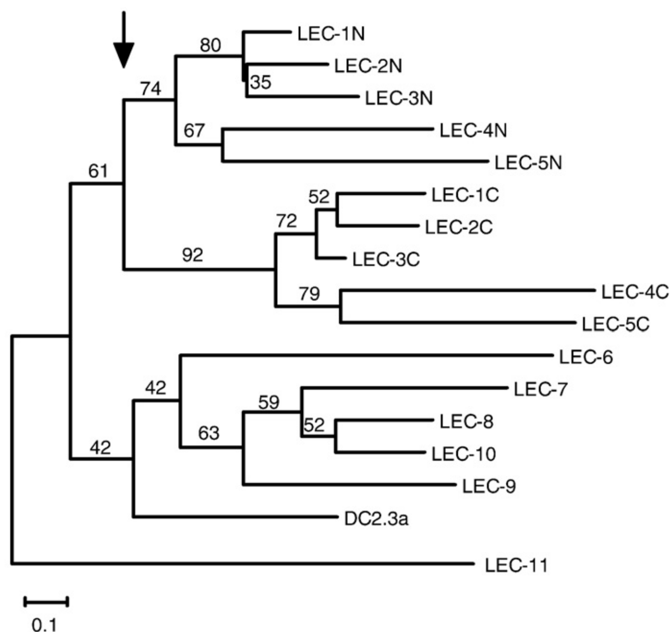


Fig. 3. A phylogenetic tree of the *C. elegans* galectin family. The phylogenetic tree was constructed as described under Materials and methods. For the analysis of tandem-repeat-type galectin-like proteins LEC-1–5, independent analysis was performed for N-terminal CRDs and C-terminal CRDs. The numbers on the individual branches of the tree indicate bootstrap values. The scale bar corresponds to 0.1 substitutions per site.

Table 1
Conservation of critical amino acids and binding to asialofetuin

		Critical amino acids								Number of substitution	Relative affinity for asialofetuin[a]
		H.	N.	R...	V.	N...	W..	E.	R		
LEC-1	N-CRD	H.	S.	R...	V.	N...	W..	E.	R	1	100
	C-CRD	H.	N.	R...	I.	N...	W..	E.	R	1	
LEC-2	N-CRD	H.	S.	R...	V.	N...	W..	E.	R	1	112
	C-CRD	H.	N.	R...	V.	N...	W..	E.	R	0	
LEC-3	N-CRD	H.	S.	R...	V.	N...	W..	E.	R	1	122
	C-CRD	H.	N.	R...	V.	N...	W..	E.	R	0	
LEC-4	N-CRD	H.	K.	N...	I.	N...	W..	E.	R	3	67
	C-CRD	H.	N.	R...	V.	N...	W..	E.	R	0	
LEC-5	N-CRD	H.	S.	R...	I.	N...	W..	D.	Q	4	NT[c]
	C-CRD	Q.	S.	R...	V.	N...	W..	E.	R	2	
LEC-6		H.	N.	R...	V.	N...	W..	E.	R	0	198
LEC-7[b]		H.	K.	E...	V.	N...	W..	Q.	R	3	NT[c]
LEC-8		H.	N.	R...	A.	N...	W..	E.	R	1	21
LEC-9		H.	N.	R...	A.	N...	W..	E.	S	2	9
LEC-10		H.	N.	R...	V.	N...	W..	E.	R	0	137
LEC-11		H.	S.	R...	V.	N...	W..	E.	H	2	NT[c]
DC2.3a[b]		H.	N.	R...	V.	N...	W..	E.	R	0	NT[c]

[a] Data from the results of Supplementary Fig. 4A are shown.
[b] LEC-7 and DC2.3a are included for comparison. See Supplementary Table 3.
[c] NT, not tested.

and that at the C-terminal CRD is also relatively low (52%). These three genes appeared to have originated at the same time.

3.7. Production and purification of recombinant galectin-like proteins

To investigate the sugar-binding properties of the *C. elegans* galectin-like proteins, recombinant proteins were expressed in *E. coli*, and their soluble extracts were applied to a column of asialofetuin agarose, as in the previous cases of LEC-1 [7] and LEC-6 [8]. We successfully produced recombinant proteins of LEC-2, -3b, -4, -8, -9, -10, and -11, but failed to produce a recombinant LEC-5 protein. Aliquots of flow-through, retarded, and adsorbed fractions were subjected to SDS-PAGE to verify the binding ability of individual galectin-like proteins to asialofetuin. LEC-2, -3b, and -10 were found to be strongly adsorbed on the column and were eluted specifically with lactose (Supplementary Fig. 3). A considerable portion of proteins leaked from the column before lactose elution in the cases of LEC-4, -8, and -11 (Supplementary Fig. 3). LEC-9 was shown to have the lowest affinity for asialofetuin because it was scarcely adsorbed on the column. Nevertheless, its elution was significantly retarded when compared with the bulk of the *E. coli* proteins.

In this study, a deletion mutant LEC-8ΔC lacking the C-terminal tail region was also used because its intact form (intact LEC-8) was very sensitive to proteolysis (data not shown). However, the results for intact LEC-8 and LEC-8ΔC were almost the same (data of LEC-8 not shown). The above galectin-like proteins (LEC-1, -2, -3b, -4, -6, -8ΔC, -9, -10, and -11) were purified to an apparent homogeneity by single or repeated steps of asialofetuin-agarose chromatography.

3.8. Analytical affinity chromatography

In order to examine the sugar-binding ability of LEC-1–11 quantitatively, FAC [32] was performed with a recently reinforced system using a miniature column packed with asialofetuin-agarose, and the elution fronts were analyzed as described previously [33,34]. The obtained results (Supplementary Fig. 4A) were consistent with the behavior of these proteins in preparative experiments using asialofetuin-agarose (Supplementary Fig. 3), and were found to be associated with the integrity of the critical amino acids demonstrated in the case of vertebrate galectins (His44, Asn46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73; numbers correspond to those of human galectin-1), as shown in Table 1. LEC-6 and -10, in which these amino acids are

completely conserved, showed the highest affinity for asialofetuin. LEC-9, which has two substitutions, showed significant but poor affinity. LEC-4 has three substitutions in the N-terminal CRD, whereas its C-terminal CRD is perfectly conserved. It showed intermediate affinity. Therefore, galectins with the eight critical amino acids were found to retain their sugar-binding affinity toward asialofetuin.

However, since fetuin carries distinct sugar chains, i.e., three *N*-linked [42] and three *O*-linked oligosaccharides [43], it is not clear whether the above discussion is applicable to only complex-type *N*-glycans or can be extended to more extensive β-galactosides. To clarify this point, we also analyzed the binding affinity of individual galectins for immobilized *N*-acetyllactosamine (LacNAc, Galβ1-4GlcNAc) and lacto-*N*-neotetraose (L*N*nT, Galβ1-4GlcNAcβ1-3Galβ1-4Glc). The results were largely consistent with each other and also with that of the asialofetuin-agarose experiment (Supplementary Fig. 4). However, there were some differences: (1) LEC-2 showed a relatively lower affinity for both LacNAc and L*N*nT relative to LEC-1. (2) LEC-6 showed a higher affinity for LacNAc than LEC-1, while the former had a relatively poor affinity toward L*N*nT. These differences may reflect specialized features of individual galectins, as described before [35].

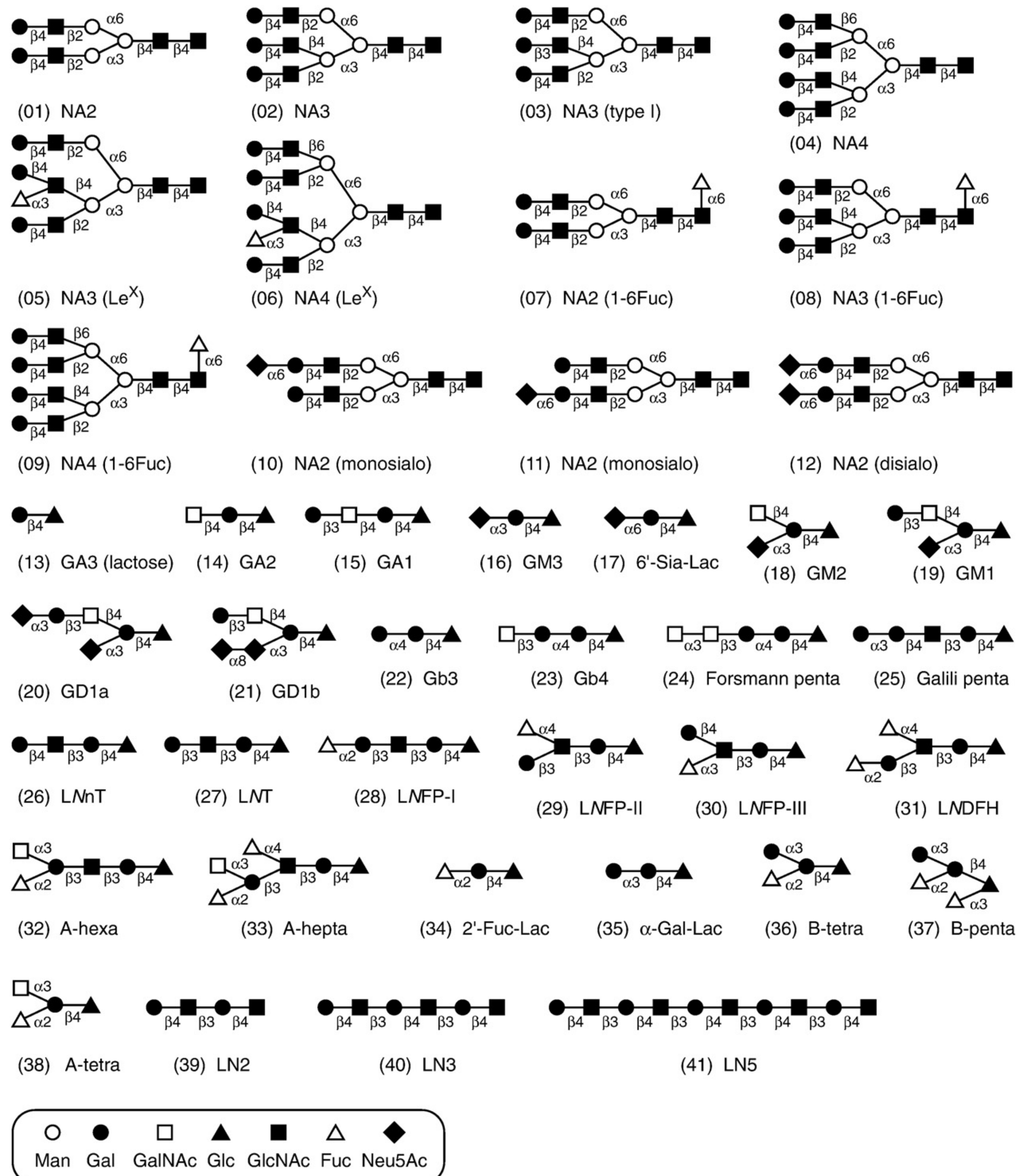


Fig. 4. Schematic representation of PA-oligosaccharides used in FAC analysis. Note that the reducing terminals are pyridylaminated for FAC analysis. Symbols used to represent monosaccharides are shown in the inset.

3.9. Fine oligosaccharide-binding properties of LEC-1, -2, -3b, -4, -6, and -10

We have reported the detailed oligosaccharide-binding properties of extensive galectins, including LEC-1 [31], LEC-6, and galectins from other species [35] by FAC using a panel of 41 PA-oligosaccharides (for structures, see Fig. 4). In this study, we extended the FAC analysis to other members of the *C. elegans* galectin family, i.e., LEC-2, -3b, -4, -8, -9, and -10, which were obtained in sufficient amounts as soluble forms. Among the tested proteins, LEC-2, -3b, -4, and -10 showed significant affinity for some of the PA-oligosaccharides used in this study. For convenience, the sugar-binding affinities of these *C. elegans* galectins, including the previously characterized LEC-1 and LEC-6, were compared (Table 2, Fig. 5).

3.9.1. Conserved features of all LECs

All LECs and CRDs shown in Table 2 and Fig. 5 recognized basic oligosaccharides such as lacto-*N*-neotetraose (L*N*nT, 26), lacto-*N*-tetraose (L*N*T, 27), and a complex-type *N*-glycan with biantennary Galβ1-4GlcNAc (NA2, 01).

To examine the effect of the number of branching, affinities for tri- (NA3, 02) and tetraantennary glycans (NA4, 04) were compared with those for a biantennary (NA2, 01) glycan. Relative affinities with respect to L*N*nT (26) are shown in Table 3. All LECs and CRDs showed higher affinities for the tri- (02) and tetraantennary glycans (04) than for the biantennary (01) glycan. LEC-3b and -6 showed much higher affinities for the tetraantennary (04) glycan than for the triantennary glycan (02) (Table 3).

Next, the effect of the number of repeats of *N*-acetyllactosamine was analyzed. Relative affinities of $(LacNAc)_2$ (39), $(LacNAc)_3$ (40), and $(LacNAc)_5$ (41) to L*N*nT (26) are shown in Table 3. All the LECs and CRDs examined, except for LEC-10, exhibited enhanced affinities as the number of repeats increased. Affinities of LEC-10 for these oligolactosamines were low, and no significant difference was observed.

Substitutions at the 6-OH groups of nonreducing terminal galactoses with sialic acids (01 vs. 10–12) eliminated their binding affinities (Table 4). On the other hand, α1–2 fucosylation of nonreducing terminal galactose of L*N*T increased (LEC-1N, -3b, and -10) or did not change (LEC-1, -1C, -2, -4, and -6) the affinity (27 vs. 28) (Table 4).

Substitutions at 3-OH groups of the GlcNAc residues of Galβ1-4GlcNAc (type 2 chain) by fucose in complex-type *N*-glycans reduced their binding affinities (02 vs. 05, 04 vs. 06) (Table 2, Fig. 5). The decrease in affinity caused by α1–3 fucosylation was also seen at the GlcNAc residue of the type 2 chain in lacto-*N*-neotetraose (L*N*nT) (26 vs. 30) (Table 2, Fig. 5). In the case of α1–4 fucosylation at the GlcNAc residue of Galβ1-3GlcNAc (type 1 chain) in lacto-*N*-tetraose (L*N*T), the affinity decreased (LEC-1, -1N, -1C, -4, -6, and -10) or remained

Table 2

Dissociation constants (K_d's) determined by frontal affinity chromatography for PA-oligosaccharides and galectins[a]

PA-sugar no.	Trivial name	LEC-1[b]			LEC-2	LEC-3	LEC-4	LEC-6[b]	LEC-10
		Whole	LEC-1N	LEC-1C					
1	NA2	220	680	150	61	300	74	93	130
2	NA3	100	370	84	26	110	26	40	70
3	NA3 (type 1)	80	260	52	20	100	22	36	81
4	NA4	110	380	77	23	41	27	27	63
5	NA3 (Le^X)	230	920	210	70	100	80	100	140
6	NA4 (Le^X)	230	780	150	58	240	55	54	110
7	NA2 (1-6Fuc)	230	580	200	51	300	71	82	130
8	NA3 (1-6Fuc)	120	370	82	25	140	30	46	78
9	NA4 (1-6Fuc)	110	370	78	21	100	25	38	60
10	NA2 (monosialo)	960	–	950	110	1000	280	590	470
11	NA2 (monosialo)	–	–	1300	–	1400	350	710	550
12	NA2 (disialo)	–	–	–	–	–	4100	–	–
13	GA3 (lactose)	–	–	–	–	–	600	2000	5500
14	GA2	–	–	–	–	–	890	–	–
15	GA1	680	–	670	740	1900	56	580	–
16	GM3	–	–	–	–	–	1800	3400	–
17	6′-Sia-Lac	–	–	–	NT[c]	–	–	–	–
18	GM2	–	–	–	–	–	1300	–	–
19	GM1	880	–	1100	–	3800	54	630	2700
20	GD1a	–	–	–	–	–	540	–	–
21	GD1b	–	–	–	–	–	86	1600	–
22	Gb3	–	–	–	–	3800	–	–	–
23	Gb4	530	2900	500	–	2100	130	1300	–
24	Forsmann penta	–	–	–	300	700	44	3400	470
25	Galili penta	54	72	100	17	70	170	160	130
26	L*N*nT	210	1000	200	140	630	25	320	290
27	L*N*T	91	440	77	35	250	83	190	320
28	LNFP-I	60	130	65	36	90	65	170	120
29	LNFP-II	300	–	240	38	230	570	–	–
30	LNFP-III	780	–	890	910	–	630	–	–
31	LNDFH	240	–	190	12	570	780	–	2700
32	A-hexa	58	48	810	24	37	72	1600	75
33	A-hepta	250	1800	170	12	650	830	–	2700
34	2′-Fuc-Lac	–	–	–	–	3800	500	3800	1800
35	α-Gal-Lac	290	420	97	680	440	92	850	4300
36	B-tetra	190	350	270	340	77	33	1500	860
37	B-penta	–	–	–	–	–	600	–	–
38	A-tetra	–	–	–	–	330	360	–	240
39	LN2	120	940	97	NT[c]	430	210	370	1100
40	LN3	68	590	51	19	230	110	230	1200
41	LN5	31	350	18	10	170	40	140	1400

[a] K_d's are given in μM.

[b] Data from our previous report are included for comparison (Ref. [31,35]).

[c] NT, not tested.

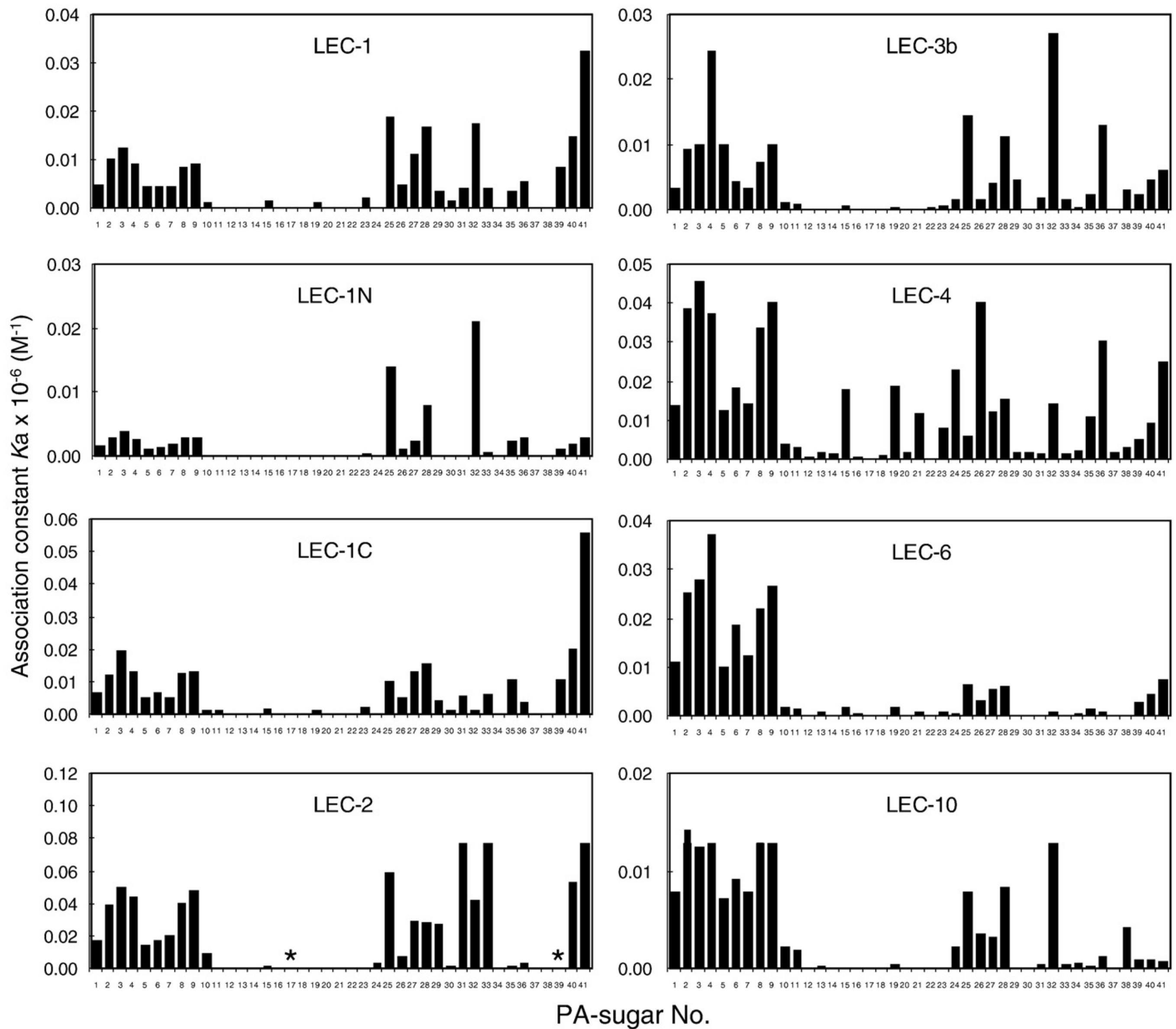


Fig. 5. Comparison of fine oligosaccharide-binding properties of the *C. elegans* galectins, LEC-1, -1N, -1C, -2, -3b, -4, -6, and -10. Purified recombinant galectins were immobilized on NHS-activated Sepharose 4FF, and FAC analysis was performed with 41 PA-oligosaccharides (Fig. 4), as described in Materials and methods. Association constants were calculated from the results shown in Table 2. Note that the ordinate scales are variable, reflecting the overall binding strengths of galectins. *Not tested.

unchanged (LEC-2 and -3b) (27 vs. 29) (Table 4). Fucosylation of the reducing terminal GlcNAc of complex-type *N*-glycans had no effect (01 vs. 07, 02 vs. 08) (Table 2, Fig. 5).

3.9.2. Specialized features of individual LECs

(1) LEC-2

LEC-2 exhibited some distinct features as compared to the other LECs and CRDs. As shown in Table 4, the substitution at the 4-OH group of GlcNAc in the type 1 chain by fucose (L*N*DFH, Le^b (31)/L*N*FP-I (28) or A-heptasaccharide (33)/A-hexasaccharide (32)) increased the affinity of LEC-2. The substitution at the nonreducing terminal Gal of lacto-*N*-fucopentaose II by fucose (L*N*DFH, Le^b (31)/L*N*FP-II, Le^a (29)) resulted in 3.2-fold increase in the affinity of LEC-2 (Table 2, Fig. 5).

(2) LEC-3b

LEC-3b showed striking enhancement of affinity for NA4 (04) as compared to that for NA3 (02) (Table 3). Relatively strong affinities for A-hexasaccharide (32) and B-tetrasaccharide (36) were observed as compared to those of the other LECs and CRDs (Table 2, Fig. 5).

(3) LEC-4

LEC-4 exhibited a strong affinity for the oligosaccharides GA1 (15), GM1 (19), and GD1b (21) containing Galβ1-3GalNAc at the nonreducing terminal, while the other LECs and CRDs used in this study showed a weak or no affinity (Table 2, Fig. 5). LEC-4 also strongly recognized the Forssman pentasaccharide (24), which was weakly recognized or not recognized by other LECs or CRDs. In contrast, LEC-4 showed a relatively weak affinity for the Galili pentasaccharide (25) among the *C. elegans* galectins (Table 2, Fig. 5). The addition of galactose to the 3-OH group of the terminal galactose of lacto-*N*-neotetraose (L*N*nT) (26) decreased the affinity of LEC-4, while it increased the affinities of the other LECs (Galili (25)/L*N*nT (26)) (Table 4). Moreover, only LEC-4 exhibited a distinct preference for *N*-acetyllactosamine (type 2 chain, 26) over lacto-*N*-biose (type 1 chain, 27) (Table 5). Although LEC-10 showed similar affinity for the type 1 and type 2 chains, other LECs and CRDs preferred the type 1

Table 3
Effects of branching and repeating of lactosamine units on galectin affinity[a]

Number of lactosamine unit	Ratio	LEC-1			LEC-2	LEC-3	LEC-4	LEC-6	LEC-10
		Whole	LEC-1N	LEC-1C					
Branch									
2	NA2 (01)/LNnT (26)	0.95	1.5	1.3	2.3	2.1	0.34	3.4	2.2
3	NA3 (02)/LNnT (26)	2.1	2.7	2.4	5.4	5.7	0.96	8.0	4.1
4	NA4 (04)/LNnT (26)	1.9	2.6	2.6	6.1	15	0.93	12	4.6
Repeat									
2	LN2 (39)/LNnT (26)	1.8	1.1	2.1	NT[b]	1.5	0.12	0.86	0.26
3	LN3 (40)/LNnT (26)	3.1	1.7	3.9	7.4	2.7	0.23	1.4	0.24
5	LN5 (41)/LNnT (26)	6.8	2.9	11	14	3.7	0.63	2.3	0.21

[a] Relative affinities with respect to LNnT (26) were calculated based on K_d's presented in Table 2.
[b] NT, not tested.

chain to the type 2 chain (Table 5). Taken together, LEC-4 exhibited the most distinctive carbohydrate-binding property among the *C. elegans* galectins analyzed in the present study.

(4) LEC-6

The main feature of LEC-6 was the relative preference for complex-type *N*-glycans (Table 2, Fig. 5). Its low affinity for A-hexasaccharide (32) was another notable feature. Substitutions at the 3-OH group of the terminal galactose of lacto-*N*-fucopentaose I (LNFP-I, 28) by *N*-acetylgalactosamine decreased the affinities of LEC-1C and LEC-6, while they increased or did not change the affinities of the other LECs (A-hexasaccharide (32)/LNFP-I (28)) (Table 4).

(5) LEC-10

Although the preference of A-hexasaccharide (32) for LNFP-I (28) was remarkable (Table 4), the overall LEC-10 affinity for oligosaccharides was very similar to that of LEC-6 (Table 2, Fig. 5). LEC-10 was unique in that it preferred A-tetrasaccharide (38) to B-tetrasaccharide (36) (Table 5).

4. Discussion

4.1. Novel galectin-like genes in C. elegans

According to a genome database search by Cooper, D.N.W. [23], a number of potential galectin genes have been nominated in some model organisms, including *C. elegans*. However, it is unknown whether such candidate genes are actually expressed *in vivo* and whether they have substantial sugar-binding activity as that of the galectins, i.e., β-galactoside-binding proteins.

We isolated the full-length cDNAs of potential galectin genes (*lec-2–5* and *8–11*) in the present study. We also added the sequence data of *lec-7* and DC2.3a registered by the *C. elegans* sequencing consortium for comparison with the cloned galectin-like genes (Table 1, Figs. 2, 3, Supplementary Table 3, Fig. 2).

There was no discrepancy with regard to the nucleotide sequences of the galectin-like cDNAs obtained in this study and those published in the genome database. Four galectin-like genes, *lec-1*, *-2*, *-3*, and *-5* were mapped to chromosome II, and *lec-4* and *-6* were mapped to chromosome III. *lec-8–9*, *lec-10*, and *lec-11* were mapped to chromosome X, V, and IV, respectively (Supplementary Table 3). Apparently, *lec-7* and *-8* are gene duplication products because both mapped to the X chromosome within 2 kbp, and they were identified in the same cosmid with identifiers R07B1.2 (*lec-7*) and R07B1.10 (*lec-8*), respectively. On the WormBase website (http://www.wormbase.org/), a *C. elegans* database, there exists 1–5 gene models of transcripts for each galectin-like gene, and some of these models have been confirmed by the *C. elegans* sequencing consortium. cDNAs isolated by this study (except for *lec-3b* as described below) correspond to some of the gene models shown in WormBase (Supplementary Table 3). There are five splicing variants of *lec-3* registered by the *C. elegans* sequencing consortium in WormBase (ZK892.1a-e; GenBank cDNA accession no. NM_063758, NM_001027360-3). *lec-3a* corresponds to ZK892.1a (NM_063758) (Supplementary Table 3), whereas *lec-3b* has no corresponding cDNA in WormBase.

4.2. Predicted features from deduced amino acid sequences

From the deduced amino acid sequences, we can describe some features of the individual galectin-like proteins. From a structural viewpoint, the C-terminal CRD of LEC-3 and DC2.3a are assumed to have an extra loop consisting of 12 and 10 amino acids, respectively, between two β-strands, S4 and S5 (Fig. 2) on the basis of X-ray crystallography of LEC-1 [44]. This loop region probably affects the carbohydrate-binding property because this region is adjacent to the carbohydrate-binding site. Insertion of such an extra-loop region has also been found in a tandem-repeat-type galectin (Hco-GAL-4) of the parasitic nematode *Haemonchus contortus* [45].

A notable feature shared by LEC-4, -5, and -11 is the presence of an “RGD” sequence at equivalent positions (Fig. 2). Interestingly, ovine galectin-15/OVGAL11 also has an RGD motif at an equivalent position

Table 4
Effects of substitutions in lactosamine units on galectin affinity[a]

Substitution	Ratio	LEC-1			LEC-2	LEC-3	LEC-4	LEC-6	LEC-10
		Whole	LEC-1N	LEC-1C					
Siaα2-6Gal	NA2 (monosialo) (10)/NA2 (1)	0.23	0	0.16	0.55	0.30	0.26	0.16	0.28
	NA2 (monosialo) (11)/NA2 (1)	0	0	0.12	0	0.21	0.21	0.13	0.24
	NA2 (disialo) (12)/NA2 (1)	0	0	0	0	0	0.02	0	0
Fucα1-2Gal	LNFP-I (28)/LNT (27)	1.5	3.4	1.2	0.97	2.8	1.3	1.1	2.7
GalNAcα1-3Gal	A-hexa (32)/LNFP-I (28)	1.0	2.7	0.08	1.5	2.4	0.90	0.11	1.6
Fucα1-4GlcNAc	LNFP-II (29)/LNT (27)	0.30	0	0.32	0.92	1.1	0.15	0	0
	LNDFH (31)/LNFP-I (28)	0.25	0	0.34	3.0	0.16	0.08	0	0.04
	A-hepta (33)/A-hexa (32)	0.23	0.03	4.8	2.0	0.06	0.09	0	0.03
Galβ1-3Gal	Galili (25)/LNnT (26)	3.9	14	2.0	8.2	9.0	0.15	2.0	2.2

[a] Relative affinities for relevant pairs were calculated based on K_d's presented in Table 2.

Table 5
Binding preference of galectins[a]

Ratio	LEC-1			LEC-2	LEC-3	LEC-4	LEC-6	LEC-10
	Whole	LEC-1N	LEC-1C					
*L*NT (27)/*L*NnT (26) (type 1/type 2)	2.3	2.3	2.6	4.0	2.5	0.30	1.7	0.91
A-tetra (38)/B-tetra (36)	0	0	0	0	0.23	0.09	0	3.6

[a] Relative affinities were calculated based on K_d's presented in Table 2.

[10]. This position apparently corresponds to the boundary between S2 and F1 β-strands of N-terminal CRDs based on the X-ray crystal structure [44] (Fig. 2). Since the region is expected to be exposed on the surface of the protein, it is possible that the RGD motif actually plays a role in cell/substratum adhesion like the other RGD-containing proteins such as fibronectin, vitronectin, laminin, and von Willebrand factor [46]. In that case, their potential receptors, i.e., integrins, may function together. Actually, ovine galectin-15/OVGAL11 is reported to mediate cell adhesion in an RGD sequence-dependent manner [47].

PSORT and SignalP analysis revealed that only LEC-5 had a signal peptide among the *C. elegans* galectin-like proteins examined. Therefore, LEC-5 can be regarded as an unusual galectin along with the sponge galectin GCLT1, which also has the putative signal peptide [48]. Recently, Kaji et al. [40] and Fan et al. [41] identified LEC-5 as an *N*-linked glycoprotein. The glycosylation site identified was Asn110 [41]. These results suggest that LEC-5 is a unique galectin family member because most of the galectins isolated to date exhibit no glycosylation. Although localization of the LEC-5 protein in *C. elegans* cells has not been elucidated, LEC-5 is probably secreted extracellularly through a classical endoplasmic reticulum (ER)/Golgi pathway that is different from those used by the most other galectins. Trp111 in the N-terminal CRD of LEC-5 is one of the indispensable amino acids for carbohydrate binding among galectins from various species. Therefore, the glycosylation at Asn110 may inhibit carbohydrate binding of the N-terminal CRD of LEC-5 because of steric hindrance. In the present study, we failed to produce a recombinant LEC-5 protein and analyze its galactose-binding ability. At present, we cannot estimate the proportion of endogenous *N*-glycosylated LEC-5. However, it seems important to assess whether the endogenous *N*-glycosylated LEC-5 protein has a carbohydrate-binding ability or some other biological function independent of carbohydrate binding.

A notable feature in LEC-7, -8, -10, and -11 is the presence of an additional C-terminal region of varied length, i.e., 40, 41, 53, and 90 amino acids, respectively. The similarity in this region between LEC-8 and LEC-10 is apparently high (48%), whereas that between the others are much lower (5–22%).

As another distinct feature is that significantly high contents of histidine were observed in LEC-8 (14%), LEC-10 (14%), and LEC-11 (11%). In addition, LEC-11 showed notable distinctions in its tail region. The tail region of LEC-11 is considerably hydrophobic and it lacks the charged residues spanning from Ile153 to Pro216 (Fig. 2). Nevertheless, the region does not seem to form the usual membrane-integrated region since it has 17 proline residues that are α-helix breakers. Analysis with the SOSUI program also confirmed that the predicted LEC-11 is a soluble protein. The hydrophobic tail is followed by a distinct sequence "His–His–His–Arg–His–His–His" (residues 225–231).

Although the size of LEC-9 is very similar to that of LEC-6, similarity between LEC-6 and LEC-9 is considerably low (28% identity). On the other hand, LEC-9 shows much higher similarities to LEC-7, -8, and -10 (36, 52, and 46% identities, respectively) despite the lack of a C-terminal tail. At present, it is not known whether LEC-9 forms a noncovalent dimer like LEC-6. Therefore, we consider LEC-9 as a member of chimera-type galectin-like proteins such as LEC-7, -8, -10, and -11.

4.3. Oligosaccharide-binding property

The present study demonstrated that most of the *C. elegans* galectin-like genes investigated in this work (*lec-1*, *-2*, *-3*, *-4*, *-6*, and *-10*) proved to be functional in terms of their β-galactoside-binding ability. However, it is notable that some expressed proteins have only a poor sugar-binding activity (e.g., LEC-8, -9, and -11). The former type of galectin-like proteins have a perfectly or almost perfectly conserved set of eight critical amino acid residues that are indispensable for the sugar-binding function of vertebrate galectins (i.e., His44, Ans46, Arg48, Val59, Asn61, Trp68, Glu71, and Arg73; numbers correspond to those of human galectin-1). Therefore, the presence of these conserved essential amino acids could be a key in identifying functional galectins when screening other databases. DC2.3a, which was not cloned in the present investigation, seems to be another galectin candidate that has a β-galactoside-binding ability because all the critical amino acids are conserved in this protein. It was also shown that individual members of the *C. elegans* galectin family are divergent in terms of both molecular structure and sugar-binding properties.

FAC in the present study indicated that the *C. elegans* galectins share basic carbohydrate-binding properties with mammalian galectins [35], which are summarized as follows. (1) All LECs and CRDs in this study recognize basic oligosaccharides such as lacto-*N*-neotetraose, lacto-*N*-tetraose, and a complex-type *N*-glycan with biantennary Galβ1-4GlcNAc to some extent (K_d values; 25–1000 μM). (2) All LECs and CRDs show higher affinities for tri- and tetraantennary glycans than for biantennary glycan. Similarly, affinity is enhanced as the number of repeats of *N*-acetyllactosamine increases. (3) Substitutions at 6-OH groups of nonreducing terminal galactoses and at 3-OH groups of the GlcNAc residues of Galβ1-4GlcNAc (type 2 chain) results in a decrease in the affinities of all LECs and CRDs. In contrast, substitutions at 4-OH groups of the GlcNAc residue of Galβ1-3GlcNAc (type 1 chain) results in decreased affinities in most cases, while there are some exceptions like LEC-1C, LEC-2, and LEC-3.

LEC-2 shows relatively high affinities for lacto-*N*-difucohexaose (*L*NDFH, Le^b) and A-heptasaccharide. Not only *C. elegans* galectins but also those from other species showed no or relatively weak affinity for these saccharides in our previous study [35]. Here, LEC-2 is considered a novel-type galectin. Thus far, the carbohydrate-binding properties of invertebrate galectins have not been extensively investigated in detail. The result suggests the possible binding of invertebrate galectins to glycans that cannot be recognized by mammalian galectins.

Among the *C. elegans* galectins examined in the present study, LEC-4 exhibits many distinct features. For instance, LEC-4 has a relatively high affinity for variable glycolipid-type oligosaccharides. Moreover, only LEC-4 prefers the type-2 chain compared to the type-1 chain among the *C. elegans* galectins. However, these features have also been observed in some galectins from other species [35].

In the above FAC study, LEC-8 and -9 were not observed to have any significant sugar-binding activity. However, since almost all the PA-oligosaccharides used are of mammalian origin, it is probable that LEC-8 and -9 have a remarkable affinity for some endogenous glycans existing in *C. elegans*, which have structures different from those in mammals. Alternatively, chimera-type galectin-like proteins of *C. elegans* may have diverse functions that are unknown because they have a C-terminal tail whose function is unknown. It was reported that the expression of *lec-8* mRNA was elevated after a *Serratia marcescens* infection [49]. This result implies that LEC-8 may play a role in defense against bacterial infections.

4.4. Expression of galectin-like genes in C. elegans

With regard to the expression of the *C. elegans* galectin-like genes *in vivo*, we detected mRNAs of *lec-1*, *-2*, *-4*, *-5*, *-6*, *-8*, *-9*, *-10*, and *-11* by RT-PCR. According to the Nematode Expression Pattern DataBase (NEXTDB) (http://nematode.lab.nig.ac.jp/), mRNA expressions of *lec-1–6* and *8–11* are spatiotemporally diverse. The 32-kDa galectin protein was shown to be abundantly localized in the cuticle and in the terminal valve of the pharynx of adult worms [50]. However, thus far, the analysis of the spatiotemporal localization of each galectin protein has not been performed from a comparative viewpoint.

4.5. Concluding remarks

We previously identified several glycoprotein ligands for LEC-6 from an extract of *C. elegans* [51,52]. In the present study, other *C. elegans* galectins were found to possess the conserved fundamental properties, i.e., ability to recognize β-galactosides such as *N*-acetyllactosamine. It is intriguing that a relatively simple multicellular organism like *C. elegans* has numerous galectins. In the present study, we revealed that they have diverse carbohydrate-binding properties. These results suggest that individual galectins have distinct roles *in vivo*. To further confirm the concept, we wish to examine protein localizations and glycoconjugate ligands for each *C. elegans* galectin *in vivo* as a future study.

Acknowledgements

The authors thank Drs. Tadasu Urashima (Obihiro University of Agriculture and Veterinary Medicine) and Kei-ichi Yoshida (Seikagaku Kogyo, Co. Ltd.) for the gift of oligosaccharides used in the FAC analysis. The authors also thank Dr. Tomoharu Takeuchi (Teikyo University School of Pharmaceutical Sciences) for helpful discussions. This work was supported by a grant (No. 10178102; to J.H.) from the program Grants-in-Aid for Scientific Research on Priority Areas of the Ministry of Education, Science, Sports, and Culture of Japan; Kato Memorial Bioscience Foundation; and Mizutani Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbagen.2008.07.003.

References

[1] S.H. Barondes, V. Castronovo, D.N. Cooper, R.D. Cummings, K. Drickamer, T. Feizi, M.A. Gitt, J. Hirabayashi, C. Hughes, K. Kasai, Galectins: a family of animal β-galactoside-binding lectins, Cell 76 (1994) 597–598.

[2] V.I. Teichberg, I. Silman, D.D. Beitsch, G. Resheff, A β-D-galactoside binding protein from electric organ tissue of *Electrophorus electricus*, Proc. Natl. Acad. Sci. U. S. A. 72 (1975) 1383–1387.

[3] S.H. Barondes, D.N. Cooper, M.A. Gitt, H. Leffler, Galectins. Structure and function of a large family of animal lectins, J. Biol. Chem. 269 (1994) 20807–20810.

[4] K. Kasai, J. Hirabayashi, Galectins: a family of animal lectins that decipher glycocodes, J. Biochem. (Tokyo) 119 (1996) 1–8.

[5] H. Leffler, Introduction to galectins, Trends Glycosci Glycotechnol 9 (1997) 9–19.

[6] D.N. Cooper, R.P. Boulianne, S. Charlton, E.M. Farrell, A. Sucher, B.C. Lu, Fungal galectins, sequence and specificity of two isolectins from *Coprinus cinereus*, J. Biol. Chem. 272 (1997) 1514–1521.

[7] J. Hirabayashi, M. Satoh, K. Kasai, Evidence that *Caenorhabditis elegans* 32-kDa β-galactoside-binding protein is homologous to vertebrate ß-galactoside-binding lectins. cDNA cloning and deduced amino acid sequence, J. Biol. Chem. 267 (1992) 15485–15490.

[8] J. Hirabayashi, T. Ubukata, K. Kasai, Purification and molecular characterization of a novel 16-kDa galectin from the nematode *Caenorhabditis elegans*, J. Biol. Chem. 271 (1996) 2497–2505.

[9] H. Leffler, S. Carlsson, M. Hedlund, Y. Qian, F. Poirier, Introduction to galectins, Glycoconj. J. 19 (2004) 433–440.

[10] C.A. Gray, D.L. Adelson, F.W. Bazer, R.C. Burghardt, E.N. Meeusen, T.E. Spencer, Discovery and characterization of an epithelial-specific galectin in the endometrium that forms crystals in the trophectoderm, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7982–7987.

[11] W.M. Abbott, T. Feizi, Soluble 14-kDa β-galactoside-specific bovine lectin. Evidence from mutagenesis and proteolysis that almost the complete polypeptide chain is necessary for integrity of the carbohydrate recognition domain, J. Biol. Chem. 266 (1991) 5552–5557.

[12] J. Hirabayashi, K. Kasai, Effect of amino acid substitution by sited-directed mutagenesis on the carbohydrate recognition and stability of human 14-kDa ß-galactoside-binding lectin, J. Biol. Chem. 266 (1991) 23648–23653.

[13] J. Hirabayashi, K. Kasai, Further evidence by site-directed mutagenesis that conserved hydrophilic residues form a carbohydrate-binding site of human galectin-1, Glycoconj. J. 11 (1994) 437–442.

[14] Y.D. Lobsanov, M.A. Gitt, H. Leffler, S.H. Barondes, J.M. Rini, 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. 268 (1993) 27034–27038.

[15] D.I. Liao, G. Kapadia, H. Ahmed, G.R. Vasta, O. Herzberg, Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 1428–1432.

[16] Y. Bourne, B. Bolgiano, D.I. Liao, G. Strecker, P. Cantau, O. Herzberg, T. Feizi, C. Cambillau, Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides, Nat. Struct. Biol. 1 (1994) 863–870.

[17] D.D. Leonidas, E.H. Vatzaki, H. Vorum, J.E. Celis, P. Madsen, K.R. Acharya, Structural basis for the recognition of carbohydrates by human galectin-7, Biochemistry 37 (1998) 13930–13940.

[18] J. Seetharaman, A. Kanigsberg, R. Slaaby, H. Leffler, S.H. Barondes, J.M. Rini, X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution, J. Biol. Chem. 273 (1998) 13047–13052.

[19] J. Hirabayashi, Y. Arata, K.I. Kasai, Galectins from the nematode *Caenorhabditis elegans* and the genome project, Trends Glycosci. Glycotechnol. 9 (1997) 113–122.

[20] J. Hirabayashi, Y. Arata, K. Hayama, K.I. Kasai, Galectins from the nematode *Caenorhabditis elegans* and the glycome project, Trends Glycosci. Glycotechnol. 13 (2001) 533–549.

[21] *C. elegans* Sequencing Consotium, Genome sequence of the nematode *C. elegans*: a platform for investigating biology, Science 282 (1998) 2012–2018.

[22] D.N. Cooper, S.H. Barondes, God must love galectins; he made so many of them, Glycobiology 9 (1999) 979–984.

[23] D.N. Cooper, Galectinomics: finding themes in complexity, Biochim. Biophys. Acta 1572 (2002) 209–231.

[24] R.J. Barstead, R.H. Waterston, The basal component of the nematode dense-body is vinculin, J. Biol. Chem. 264 (1989) 10177–10185.

[25] J.M. Short, J.M. Fernandez, J.A. Sorge, W.D. Huse, λZAP: a bacteriophage λ expression vector with *in vivo* excision properties, Nucleic Acids Res. 16 (1988) 7583–7600.

[26] Y. Arata, J. Hirabayashi, K. Kasai, Structure of the 32-kDa galectin gene of the nematode *Caenorhabditis elegans*, J. Biol. Chem. 272 (1997) 26669–26677.

[27] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.

[28] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[29] M. Kaneko, T. Kudo, H. Iwasaki, Y. Ikehara, S. Nishihara, S. Nakagawa, K. Sasaki, T. Shiina, H. Inoko, N. Saitou, H. Narimatsu, α1,3-fucosyltransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX, FEBS Lett. 452 (1999) 237–242.

[30] Y. Arata, J. Hirabayashi, K. Kasai, The two lectin domains of the tandem-repeat 32-kDa galectin of the nematode *Caenorhabditis elegans* have different binding properties. Studies with recombinant protein, J. Biochem. (Tokyo) 121 (1997) 1002–1009.

[31] Y. Arata, J. Hirabayashi, K. Kasai, Sugar binding properties of the two lectin domains of the tandem repeat-type galectin LEC-1 (N32) of *Caenorhabditis elegans*. Detailed analysis by an improved frontal affinity chromatography method, J. Biol. Chem. 276 (2001) 3068–3077.

[32] K. Kasai, S. Ishii, Quantitative analysis of affinity chromatography of trypsin. A new technique for investigation of protein–ligand interaction, J. Biochem. (Tokyo) 77 (1975) 261–264.

[33] J. Hirabayashi, Y. Arata, K. Kasai, Reinforcement of frontal affinity chromatography for effective analysis of lectin–oligosaccharide interactions, J. Chromatogr. A 890 (2000) 261–271.

[34] Y. Arata, J. Hirabayashi, K.I. Kasai, Application of reinforced frontal affinity chromatography and advanced processing procedure to the study of the binding property of a *Caenorhabditis elegans* galectin, J. Chromatogr. A 905 (2001) 337–343.

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

[36] J. Hirabayashi, Y. Arata, K. Kasai, Frontal affinity chromatography as a tool for elucidation of sugar recognition properties of lectins, Methods Enzymol. 362 (2003) 353–368.

[37] M. Krause, D. Hirsh, A trans-spliced leader sequence on actin mRNA in *C. elegans*, Cell 49 (1987) 753–761.

[38] J. Spieth, G. Brooke, S. Kuersten, K. Lea, T. Blumenthal, Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions, Cell 73 (1993) 521–532.

[39] M. Krause, Transcription and translation, in: H.F. Epstein, D.C. Shakes (Eds.), Methods in Cell Biology, vol. 48, *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, Academic Press, Inc., San Diego, 1995, pp. 483–512.
[40] H. Kaji, H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kasai, N. Takahashi, T. Isobe, Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify *N*-linked glycoproteins, Nat. Biotechnol. 21 (2003) 667–672.
[41] X. Fan, Y.M. She, R.D. Bagshaw, J.W. Callahan, H. Schachter, D.J. Mahuran, Identification of the hydrophobic glycoproteins of *Caenorhabditis elegans*, Glycobiology 15 (2005) 952–964.
[42] M.G. Yet, C.C. Chin, F. Wold, The covalent structure of individual *N*-linked glycopeptides from ovomucoid and asialofetuin, J. Biol. Chem. 263 (1988) 111–117.
[43] A.S. Edge, R.G. Spiro, Presence of an *O*-glycosidically linked hexasaccharide in fetuin, J. Biol. Chem. 262 (1987) 16135–16141.
[44] T. Nonaka, K. Sekihashi, S. Nishizaki, Y. Arata, J. Hirabayashi, K.I. Kasai, Y. Mitsui, Crystal structure of the tandem-repeat galectin of the nematode *Caenorhabditis elegans*, Glycoconj. J. 16 (1999) S119.
[45] C.J. Greenhalgh, A. Loukas, D. Donald, S. Nikolaou, S.E. Newton, A family of galectins from *Haemonchus contortus*, Mol. Biochem. Parasitol. 107 (2000) 117–121.
[46] E. Ruoslahti, RGD and other recognition sequences for integrins, Annu. Rev. Cell Dev. Biol. 12 (1996) 697–715.
[47] J.L. Farmer, R.C. Burghardt, F.D. Jousan, P.J. Hansen, F.W. Bazer, T.E. Spencer, Galectin 15 (LGALS15) functions in trophectoderm migration and attachment, FASEB J. 22 (2008) 548–560.
[48] C. Wagner-Hülsmann, N. Bachinski, B. Diehl-Seifert, B. Blumbach, R. Steffen, Z. Pancer, W.E. Müller, A galectin links the aggregation factor to cells in the sponge (*Geodia cydonium*) system, Glycobiology 6 (1996) 785–793.
[49] G.V. Mallo, C.L. Kurz, C. Couillault, N. Pujol, S. Granjeaud, Y. Kohara, J.J. Ewbank, Inducible antibacterial defense system in *C. elegans*, Curr. Biol. 12 (2002) 1209–1214.
[50] Y. Arata, Y. Akimoto, J. Hirabayashi, K. Kasai, H. Hirano, An immunohistochemical study of the 32-kDa galectin (ß-galactoside-binding lectin) in the nematode *Caenorhabditis elegans*, Histochem. J. 28 (1996) 201–207.
[51] J. Hirabayashi, K. Hayama, H. Kaji, T. Isobe, K. Kasai, Affinity capturing and gene assignment of soluble glycoproteins produced by the nematode *Caenorhabditis elegans*, J. Biochem. (Tokyo) 132 (2002) 103–114.
[52] J. Hirabayashi, K. Kasai, Separation technologies for glycomics, J. Chromatogr. B. Biomed. Appl. 771 (2002) 67–87.