Special Issue

Journal Home page : www.jeb.co.in ★ E-mail : editor@jeb.co.in

Journal of Environmental Biology

JEB™ p-ISSN: 0254-8704
e-ISSN: 2394-0379
CODEN: JEBIDP

DOI : http://doi.org/10.22438/jeb/40/5(SI)/SI-17

In silico characterization of single and tandem-repeat galectin from terrestrial slug, *Incilaria fruhstorferi*

Paper received: 30.10.2018 *Revised received*: 06.03.2019 *Accepted*: 16.03.2019

Authors Info

B.B. Patnaik[2], H.J. Hwang[1], S. Baliarsingh[2], J.M. Chung[1], M.K. Sang[1], H.R. Min[1], J.E. Park[1], H.C. Cho[1], S.W. Kang[3], S.Y. Park[4], H.S. Park[5], J.S. Lee[6] and Y.S. Lee[1]*

[1]Department of Life Sciences and Biotechnology, College of Natural Sciences, Soonchunhyang University, 31538, Korea

[2]School of Biotech Sciences, Trident Academy of Creative Technology, Bhubaneswar-751 024, India

[3]Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Jeongeup, 56212, Korea

[4]Animal & Plant Research Department, Nakdonggang National Institute of Biological Resources, Sangju, Gyeonbuk, 37242, Korea

[5]Research Institute, GnC Bio Company Ltd., Daejeon, 34069, Korea

[6]Institute for basic sciences, College of Natural Sciences, Soonchunhyang University, Asan, Chungnam, 31538, Korea

***Corresponding Author Email :** yslee@sch.ac.kr

Edited by

Professor Hyun Woung Shin

Reviewed by

Dr. Yeon Soo Han

Dr. Yong Hun Jo

Abstract

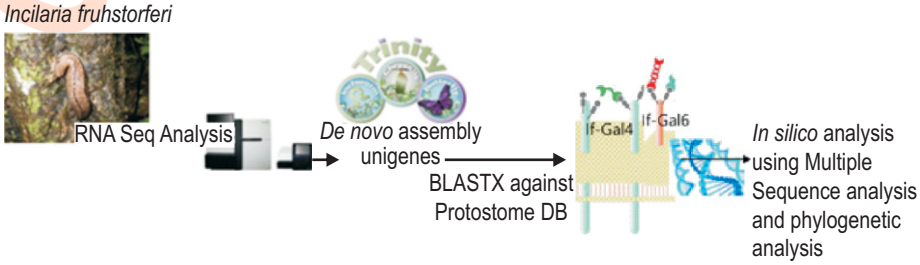
Aim: Galectins are β-galactoside binding proteins currently represented by 15 members (Galectin-1 through Galectin-15). The members are differentiated on the basis of the number of carbohydrate recognition domains (CRDs), the linker region connecting dual CRDs, and the conserved amino acid (aa) residues within CRDs. In this study, galactosyl-binding lectin family proteins were screened and characterized from transcriptome of terrestrial slug *Incilaria fruhstorferi* using bioinformatics analysis.

Methodology: *I. fruhstorferi* unigene sequences showing identity to galectins in the orthologous species were used for prediction of ORFs using FGENESH software. Subsequently, the ORFs were validated for conserved domain analysis using BLASTp and SMART programs. The multiple sequence alignments and percent identity/distance was elucidated using Clustal X2 and represented in GeneDoc sequence visualization program. The phylogenetic tree was inferred using MEGA7.

Results: A single-repeat and a tandem-repeat galectin (If_Gal6 and If_Gal4, respectively) were retrieved from the transcriptome profile of terrestrial slug, *I. fruhstorferi*. The predicted If_Gal4 and If_Gal6 comprised of 693 nucleotides translated to 230 aa residues and 435 nucleotides translated to 144 aa residues, respectively. If_Gal4 possessed two CRDs, each showing the highly conserved aa responsible for galactosyl-binding activity. Surprisingly, If_Gal6 is a single-CRD protein in contrast to the common tandem-repeat galectins illustrated for Gal6 in most other invertebrate and vertebrate species. At the level of aa sequence, If_Gal4 showed maximum of 54% identity and minimum distance of 0.459 with the sea snail, *Haliotis discus* Galectin-4 (Hd_Gal4). Cluster analysis showed higher similarities of If_Gal4 with selected molluscan Galectin-4 aa sequences. If_Gal6 showed 41% identity with the air-breathing freshwater snail, *Biomphalaria glabrata* Galectin-6 (Bg_Gal6) supposed to exist as dual-CRD molluscan galectin.

Interpretation: Discovery and structural characterization of a single-CRD galectin from *I. fruhstorferi* is second in molluscs after the sequence was reported from the Pacific oyster, *Crassostrea gigas*.

Key words: Carbohydrate-recognition domains, Galectin, *Incilaria fruhstorferi,* Phytogenetic analysis



How to cite : Patnaik, B.B., H.J. Hwang, S. Baliarsingh, J.M. Chung, M.K. Sang, H.R. Min, J.E. Park, H.C. Cho, S.W. Kang, S.Y. Park, H.S. Park, J.S. Lee, Y.S. Lee: *In silico* characterization of single and tandem-repeat galectin from terrestrial slug, *Incilaria fruhstorferi*. *J. Environ. Biol.*, **40**, 940-947 (2019).
DOI : http://doi.org/10.22438/jeb/40/5(SI)/SI-17

Introduction

Galectins are lectins with carbohydrate-recognition domains (CRDs) binding specifically to β-galactoside residues. To date, this large family of lectins are represented by 15 members (Galectin-1 through Galectin-15) classified based on the number and length of CRDs, a linker peptide connecting the CRDs, and highly conserved residues participating in the galactosyl binding activity (Yoshino *et al*., 2008; Vasta *et al*., 2015). Galectin-1, -2, -5, -7, -10, -11, -13, -14, and -15 are prototype galectins bearing single CRD, galectins-4, -6, -8, -9, and -12 are tandem-repeat galectins with 2 distinct CRDs separated by a linker region and galectin-3 a chimeric lectin with a single CRD at the C-terminus and a large N-terminal protein binding domain (Nio-Kobayashi, 2017; Diaz-Alvarez and Ortega, 2017). Despite sequence conservation at the level of CRDs, galectins have been implicated in variety of functions including anti-apoptosis activity, cell proliferation, development and innate immunity (Honjo *et al.,* 2001; Yang and Liu, 2003; Diaz-Alvarez and Ortega, 2017). Hence, understanding galectin repertoire in non-model species can be useful in extracting fitness information, especially under immunological perspective and making informed choices for conserving the species in the wild habitats.

Multiple members of galectin superfamily regulating biological, cellular and immunological processes are known in metazoan invertebrates. Prototype galectin-1 and -2 isolated from marine sponge, *Cinachyrella* sp. that mediates ionotropic glutamate receptors are suggestive of neurophysiological function (Ueda *et al*., 2013). Galectins have also been purified from the nematode species such as *Caenorhabditis elegans* (Hirabayashi *et al.,* 1996), *Haemonchus contortus* (Greenhalgh *et al.,*2000), and *Toxoascaris leonina* (Hwang *et al.,*2016) among others. In *Drosophila melanogaster,* six candidate galectin genes have been reported (Pace *et al.,* 2002). Generally, insect galectins are associated with embryonic development and immunity but a few of them are the mediators of parasite infection in insect hosts (Kamhawi *et al.,* 2004; Huang *et al.,* 2007). Extensive galectin repertoire has also been discovered from crustacean species such as Chinese mitten crab, *Eriocheir sinensis* (Wang *et al.,* 2016), shrimp, *Litopenaeus vannamei* (Hou *et al.,* 2015; Tian *et al.,* 2018) and Kuruma shrimp, *Marsupenaeus japonicus* (Shi *et al.,* 2014). Further, in the phylum Mollusca previous studies have revealed the domain organization, expression analysis and structure-function relationships of galectins (Vasta *et al.,* 2015; Wang *et al.,* 2018). Although the tandem-repeat galectins are highly populated in molluscs, single CRD and chimeric lectins have also been known (Zhang *et al.,* 2011). A galectin with unique four-CRD organization and broader binding specificity has been known from the hemocyte of eastern oyster, *Crassostrea virginica* (Feng *et al.,* 2015).

Despite compelling evidences related to the structural characterization of galectin-like proteins in molluscan species, there are still many more that needs to be explored especially in relation to the adaptation potential of non-model species. *Incilaria fruhstorferi* (*syn.Meghimatium fruhstorferi* or *Philomycus fruhstorferi* Collinge, 1901) is an air-breathing slug grouped under family Philomycidae. The genomic information of the species is limited at National Center for Biotechnology Information (NCBI), thus reducing the scope for understanding the genic regulation of phenotypic traits. We have provided the first set of genomic resources by *de novo* transcriptome characterization of *I. fruhstorferi* focusing on growth, immunity, and reproduction related genes (Patnaik *et al.,* 2019). The immunity related genes of *I. fruhstorferi* were classified to categories such as the pathogen-recognition receptors (PRRs), Toll signaling pathway, endogenous ligands, immune effectors, antimicrobial peptides, cytokines receptors and others. Towards exploring the gene structure dynamics of lectin families and their related function as PRRs in innate immunity, we screened unigene sequences homologous to galectin-4 and galectin-6 in *I. fruhstorferi*. Earlier, transcriptome characterization of Spanish slug, *Arion vulgaris* identified ten unigenes coding *in silico* for tandem-repeat galectins (Bulat *et al.,* 2016).To our knowledge, this study represents only the second investigation of galectins at the structural level in the slug species.

Materials and Methods

Screening of *I. fruhstorferi transcriptome* database: The non-redundant unigene sequences obtained from transcriptome analysis of air-breathing slug, *I. fruhstorferi* were annotated against the locally curated Protostome database (PANM DB) (BLASTx at an E-value cut-off threshold of 1.0E-5). Earlier, the raw data obtained from the sequencer were registered with NCBI sequence read archive (SRA) under SRR5936593, BioProject-PRJNA398441, and BioSample-SAMN7510659. The *I. fruhstorferi* unigenes *i.e.*, If_Uni_08465, If_Uni_08466, If_Uni_16808, If_Uni_28515 and If_Uni_38412 with lengths of 1808, 1841, 1783, 831 and 4094 bp, respectively, were found similar to homologous Galectin-4 sequences in PANM DB. If_Uni_17384 and If_Uni_28514 with lengths of 1193 and 500 bp were found to have homology to Galectin-6 sequences in PANM DB.

Bioinformatics analysis: The unigene sequences known to show homology to Galectin-4 and Galectin-6 sequences in the database were used as the query for ORF prediction analysis using the FGENESH suite (http://www.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind) and GENSCAN at http://genes.mit.edu/GENSCAN.html. The sequences were processed using Ultra Edit-32 Professional Text/HEX editor (version 12.00, IDM Computer Solutions Inc., Hamilton, OH, USA) software package. Homology analyses were performed using Basic Local Alignment Search Tool (BLASTp) from the NCBI database. Pairwise and multiple sequence alignment was processed using ClustalX2 (version 2.0) (Larkin *et*

al., 2007). The Clustal X2 aligned files were edited and presented using GeneDoc suite (Nicholas and Nicholas 1997). The percentage identity and distance matrix were scored using the 'Draw Tree' option in the ClustalX2 suite. The phylogenetic tree was constructed using MEGA software (version 7.0) (Kumar *et al.,* 2016). Molecular phylogenetics analysis was carried out by maximum-likelihood method using Poisson correction technique. The orthologous sequences of Galectin-4 and Galectin-6 from representative species of invertebrates and vertebrates for alignment and phylogenetic tree construction were retrieved from GenBank repository.

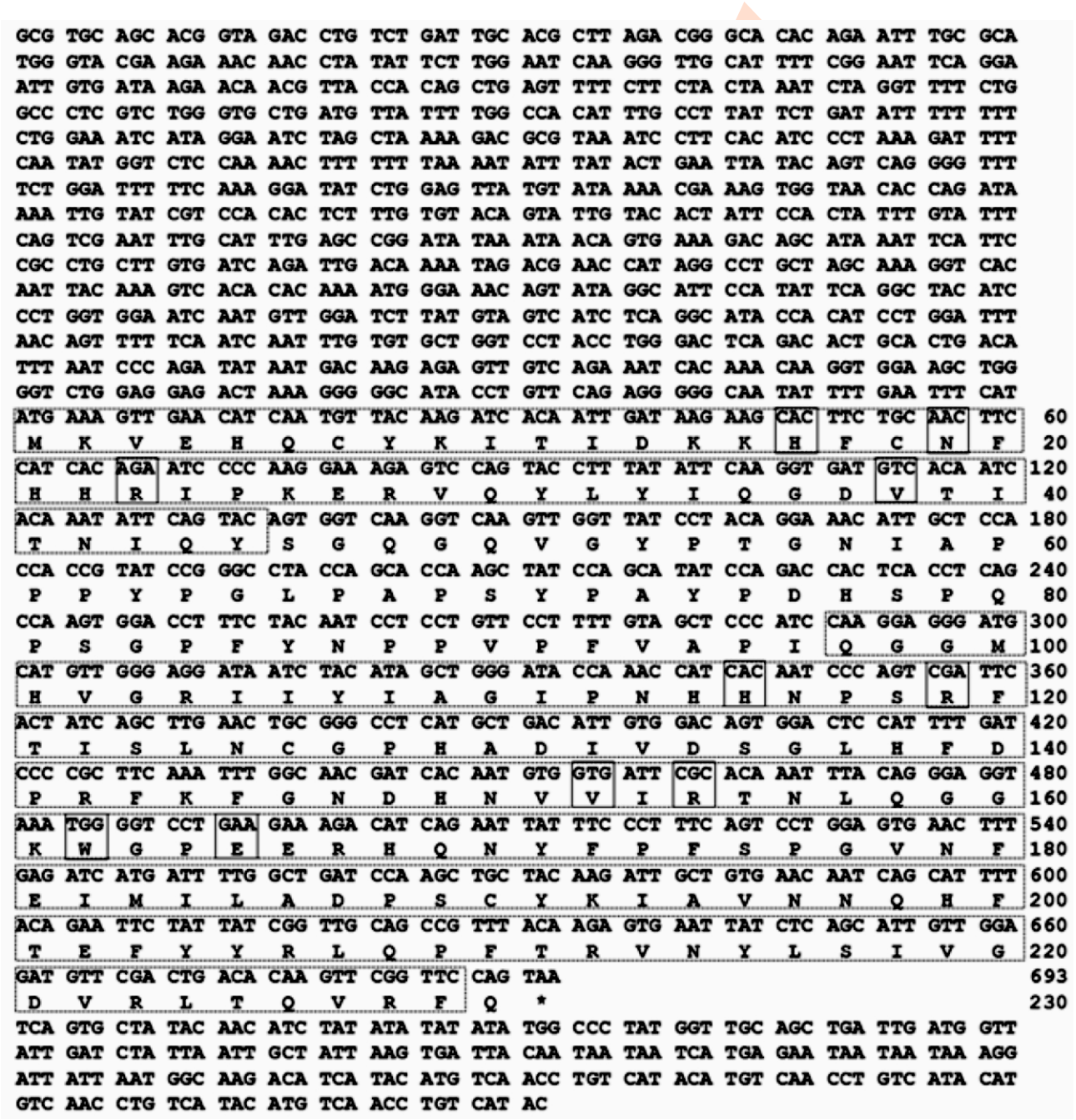


Fig. 1 : The full-length nucleotide sequence for the tandem-repeat galectin 4 from *I. fruhstorferi* (If_Gal4). The predicted ORF with the translated protein sequence is numbered. The ORF starts with ATG (initiation codon) and ends with stop codon TAA (denoted by *). Two CRDs were identified (dotted box areas for CRD1: aa1-aa135 and CRD2: aa288- aa687), separated by a 51 aa peptide linker. Boxed aa show the galactosyl-binding sites within the CRDs (CRD1: H^{16}, N^{19}, R^{23}, V^{38}; CRD2: H^{115}, R^{119}, V^{152}, R^{154}, W^{162}, E^{165}).

Results and Discussion

The unigene sequences from *I. fruhstorferi* transcriptome datasets were annotated against the locally curated database of proteins called PANM DB. The database was found more efficient in processing Next-Generation sequencing (NGS) datasets from non-model nematode, arthropod and molluscan species. Compared to the NCBI's non-redundant database (NCBInr), showed 15 times faster processing speed with greater number of annotations (Kang *et al.,* 2016). Using the annotation profile of unigenes against PANM DB (BLASTx with an E-value threshold of 1.0E-5), unigene sequences matching two members of galectin superfamily, namely galectin-4 and galectin-6 were predicted. Transcriptome characterization and *de novo* analysis has heralded a revolution in the discovery of novel candidate genes in non-model species and many earlier reports have screened galectin homologues in molluscs that have been categorized under PRRs (Castellanos-Martinez *et al.,* 2014; Ertl *et al.,* 2016; Mun *et al.,* 2017). The predicted ORF sequence for If_Gal4 showed two conserved galectin superfamily domains (CRD1 and CRD2) connected by a linker peptide (Fig. 1).

Dual CRDs were found to be the characteristic of tandem-repeat galectins, and hence If_Gal4 was considered in this category. Overall, 693 nucleotide coding sequence was predicted for If_Gal4 translating to a polypeptide of 230 aa residues. CRD1 was predicted from aa position 1 to aa position 45, while CRD2 was predicted from aa position 96 to aa position 229. CRD1 and CRD2 were separated by a less conserved 51 aa sequence called linker peptide. Four (H^{16}, N^{19}, R^{23}, V^{38} in CRD1) to six (H^{115}, R^{119}, V^{152}, R^{154}, W^{162}, E^{165} in CRD2) conserved residues responsible for galactosyl-binding activity was predicted for If_Gal4. Further, the two CRDs were different in their lengths with CRD1 smaller as compared to CRD2. The results partially confirmed the findings of a tandem-repeat galectin from the freshwater snail, *Biomphalaria glabrata*, wherein 6 aa within each CRD was responsible for galactosyl-binding activity (Yoshino *et al.,* 2008). Generally, tandem-repeat galectins of invertebrates contain 6-8 conserved aa that confers sugar-binding activity (Huang *et al.,* 2007).The transcriptome of Spanish slug, *Arion vulgaris* identified unigenes with coding sequences for tandem-repeat galectins (Bulat *et al.,* 2016).

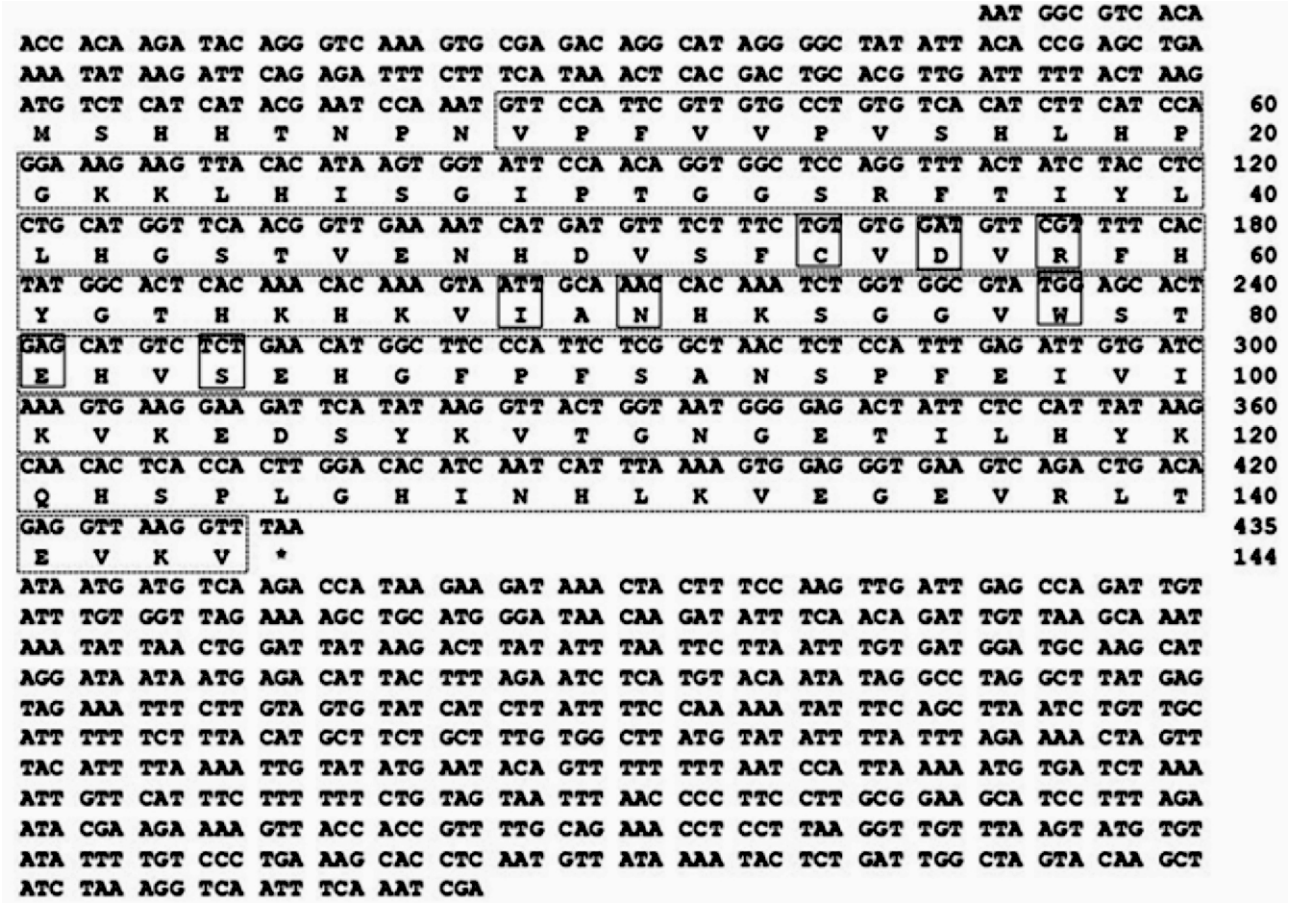


Fig. 2 : The full-length nucleotide sequence for the single-CRD galectin 6 from *I. fruhstorferi* (If_Gal6). The predicted ORF with the translated protein sequence is numbered. The ORF starts with ATG (initiation codon) and ends with stop codon TAA (denoted by *). Only one CRD was identified (dotted box area for CRD: aa9-aa144). Boxed aa show the galactosyl-binding sites within the CRD (C^{54}, D^{56}, R^{58}, I^{69}, N^{71}, W^{78}, E^{81}, S^{84}).

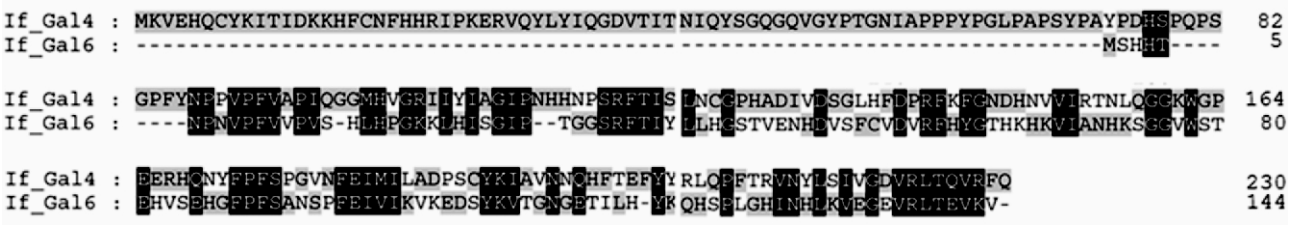


Fig. 3 : Pairwise alignment comparing the tandem-repeat If_Gal4 with the single-CRD If_Gal6. Clustal X2 was used to align the aa sequences of If_Gal4 and If_Gal6. The ClustalX2 output file is shown as a GeneDoc representation. The aa shaded black are mostly identical, in gray are similar and '-'represents terminal gaps in the alignment.

	If_Gal4	Hd_Gal4	Mg_Gal4	Cg_Gal4	Bt_Gal4	Ss_Gal4	Hs_Gal4	Cf_Gal4	Rn_Gal4	Xl_Gal4	Cs_Gal4	Aa_Gal4	Db_Gal4
If_Gal4		54	41	43	32	35	33	33	31	34	32	34	20
Hd_Gal4	0.459		46	44	35	37	36	36	34	34	35	29	22
Mg_Gal4	0.586	0.542		45	34	34	34	35	33	32	32	28	19
Cg_Gal4	0.573	0.562	0.55		40	38	39	39	39	41	36	32	22
Bt_Gal4	0.677	0.646	0.661	0.603		78	79	75	73	56	48	32	21
Ss_Gal4	0.654	0.629	0.661	0.616	0.217		80	76	72	55	50	31	21
Hs_Gal4	0.668	0.643	0.658	0.609	0.211	0.195		79	77	54	49	31	21
Cf_Gal4	0.673	0.643	0.654	0.607	0.245	0.239	0.213		72	57	49	32	22
Rn_Gal4	0.693	0.664	0.666	0.605	0.272	0.276	0.233	0.281		54	49	31	21
Xl_Gal4	0.659	0.656	0.676	0.594	0.443	0.447	0.46	0.428	0.457		51	35	19
Cs_Gal4	0.678	0.655	0.681	0.64	0.521	0.495	0.511	0.508	0.51	0.49		28	19
Aa_Gal4	0.662	0.705	0.723	0.676	0.681	0.688	0.688	0.681	0.686	0.65	0.725		23
Db_Gal4	0.802	0.783	0.808	0.778	0.787	0.794	0.784	0.781	0.788	0.812	0.812	0.768	

Fig. 4 : Percent identity and distance matrix of selected invertebrate and vertebrate galectin 4 aa sequences. Clustal X2 was used to obtain the percent identity (right side matrix) and distance matrix (left side matrix) analysis. GenBank accession nos.: Cf_Gal4: *Canis lupus familiaris*, NP_001271399.1; Hs_Gal4: *Homo sapiens*, NP_006140.1; Rn_Gal4: *Rattus norvegicus*, NP_037107.1; Bt_Gal4: *Bos taurus*, NP_001029940.1; Ss_Gal4: *Sus sucrofa*, NP_999146.1; Hd_Gal4: *Haliotis discus*, APW78605.1; Xl_Gal4: *Xenopus laevis*, NP_001079041.1; Aa_Gal4: *Aedes aegypti*, XP_001650705.1; Cg_Gal4: *Crassostrea gigas*, XM_011445428.1; Db_Gal4: *Drosophila busckii*, XP_017844125.1; Nv_Gal4: *Nasonia vitripennis*, XP_008211704.1; Cs_Gal4: *Channa striatus*, CCQ48559.1; Mg_Gal4: *Mytilus galloprovincialis*, AJQ21511.1.

The *I. fruhstorferi* unigene sequence showing matches to galectin-6 (If_Gal6) was predicted to contain a coding sequence of 435 nucleotides encoding 144 aa. Only one CRD of 136 aa was noticed containing 8 conserved aa responsible for galactosyl binding (C^{54}, D^{56}, R^{58}, I^{69}, N^{71}, W^{78}, E^{81}, S^{84}) (Fig. 2). Single CRD galectins are the prototype galectins having shorter aa sequence. Surprisingly, If_Gal6 found its homologous match with galectin-6 (a tandem-repeat galectin) and not with any of the prototype galectins. A pairwise alignment profile showed gaps at N-terminal end for If_Gal6 suggesting the absence of CRD1 (Fig. 3).

Further, comparison of If_Gal4 with its orthologs showed the highest sequence identity (54%) with the disk abalone, *Haliotis discus* dual-CRD galectin (Hd_Gal4) followed with the Pacific Oyster, *Crassostrea gigas* (43%; Cg_Gal4) and the Mediterranean mussel, *Mytilus galloprovincialis* dual-CRD galectins (41%; Mg_Gal4) (Fig. 4). The identity with vertebrate tandem-repeat galectins was in the range of 33-35%. The distance matrix also showed closer relationship of If_Gal4 and Hd_Gal4 aa sequences while *Drosophila busckii* galectin-4 (Db_Gal4) aa sequence showed maximum divergence. Phylogenetic tree analysis of vertebrate and invertebrate tandem-repeat galectin 4 grouped If_Gal4 within the metazoan molluscan cluster (Fig. 5). A separate cluster of vertebrate galectin-4 was also prominent. The *Aedes aegypti* galectin-4 (Aa_Gal4) and the Db_Gal4 were found to be the least related and formed separate lines of divergence from other invertebrate and vertebrate galectins. Further,If_Gal6 aa sequence showed the highest identity of 41% with *B. glabrata* galectin-6 (Fig. 6).

The identities with other orthologs, including the molluscs, were less than 30%. This might be because galectin-6 exists as a tandem-repeat galectin and If_Gal6 showed features of prototype-galectins. It is assumed that If_Gal6 is a prototype galectin with a single CRD containing the galactosyl-binding residues. The distance matrix also suggested similarities between If_Gal6 and Bg_Gal6. The phylogram essentially showed distribution of If_Gal6 with Bg_Gal6 confirming closer affinity of the sequences at aa level (Fig. 7).

This study has identified and structurally characterized a dual-CRD galectin (If_Gal4) and a single-CRD prototype galectin

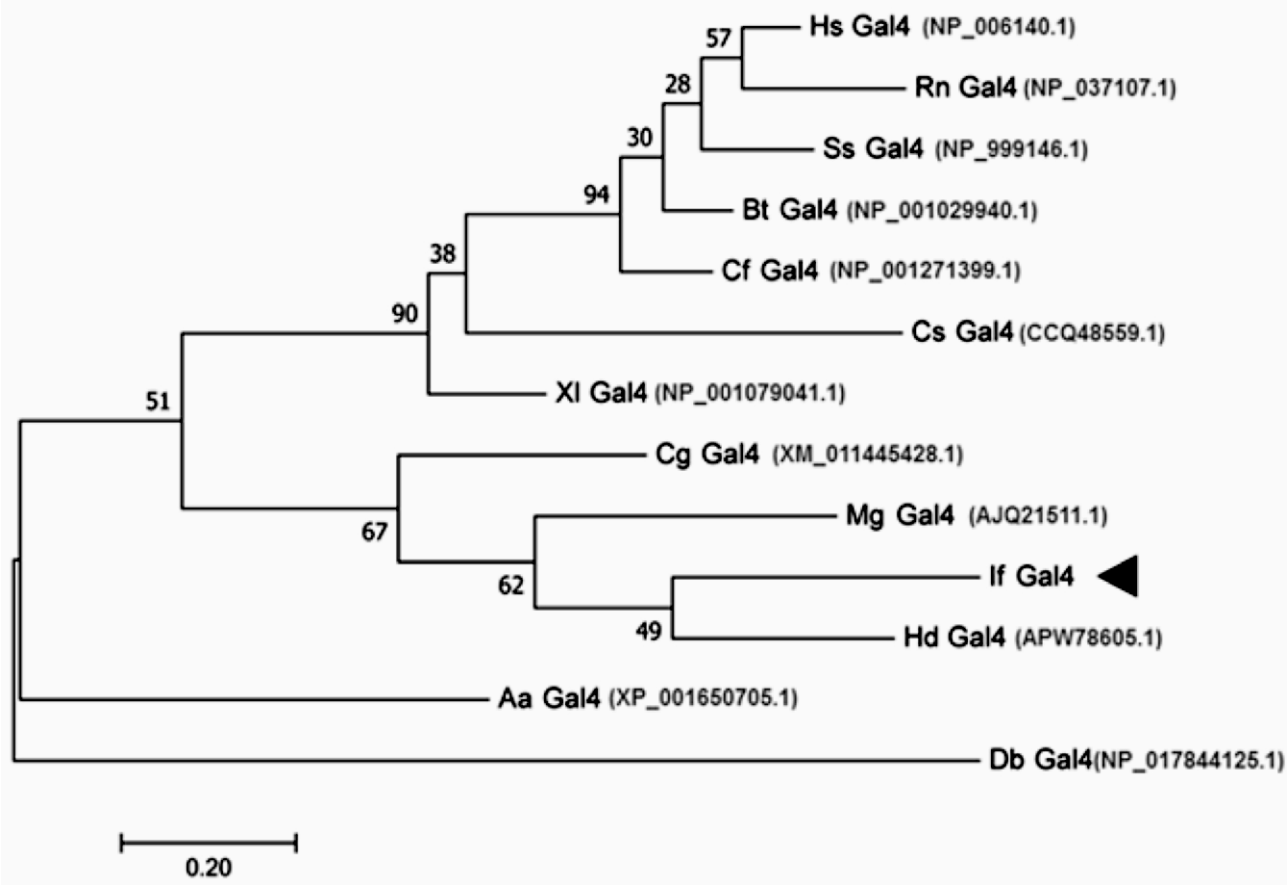


Fig. 5 : Phylogenetic tree analysis of selected invertebrate and vertebrate galectin 4 aa sequences. The analysis used the maximum likelihood method using MEGA7. Evolutionary history is based on the Poisson correction method. The tree with the highest log likelihood (-2395.11) is shown. The percentage of trees in the associated taxa clustered together right next to the branches. The GenBank accession numbers are as listed in the legend of figure 4.

	If_Gal6	Bg_Gal6	Mm_Gal6	Mz_Gal6	My_Gal6	Cg_Gal6	Bi_Gal6	Oa_Gal6	Aa_Gal6	Hs_Gal6	Ci_Gal6
If_Gal6		41	25	24	24	29	21	20	22	29	21
Bg_Gal6	0.587		36	35	23	25	25	25	24	24	26
Mm_Gal6	0.748	0.642		51	26	28	27	28	27	29	28
Mz_Gal6	0.759	0.648	0.49		24	24	25	25	25	25	29
My_Gal6	0.759	0.773	0.745	0.761		67	22	22	20	18	23
Cg_Gal6	0.712	0.753	0.725	0.745	0.333		21	23	21	19	22
Bi_Gal6	0.79	0.746	0.734	0.748	0.782	0.792		58	35	24	19
Oa_Gal6	0.797	0.746	0.716	0.751	0.782	0.771	0.423		33	22	22
Aa_Gal6	0.776	0.765	0.726	0.75	0.8	0.793	0.654	0.669		22	25
Hs_Gal6	0.712	0.761	0.709	0.749	0.821	0.812	0.759	0.775	0.784		23
Ci_Gal6	0.788	0.745	0.725	0.713	0.77	0.776	0.807	0.781	0.754	0.769	

Fig. 6 : Percent identity and distance matrix of selected invertebrate and vertebrate galectin 6 aa sequences. Clustal X2 was used to obtain the percent identity (right side matrix) and distance matrix (left side matrix) analysis. GenBank accession nos.: Mm_Gal6: *Mus musculus*, AAC04508.1; Mz_Gal6: *Mylandia zebra*, XP_023009370.2; Bi_Gal6: *Bombus impatiens*, XP_012240239.1; Oa_Gal6: *Orussius abietinus*, XP_ 012271962.1; My_Gal6: *Mizuhopecten yessoensis*, OWF37599.1; Hs_Gal6: *Hadnurux spadix*, JAV48145.1; Aa_Gal6: *Aedes albopictus*, XP_019541205.1; Ci_Gal6: *Ciona intestinalis*, XP_002131573.1; Cg_Gal6: *Crassostrea gigas*, EKC31758.1; Bg_Gal6: *Biomphalaria glabrata*, XP_013086830.1).

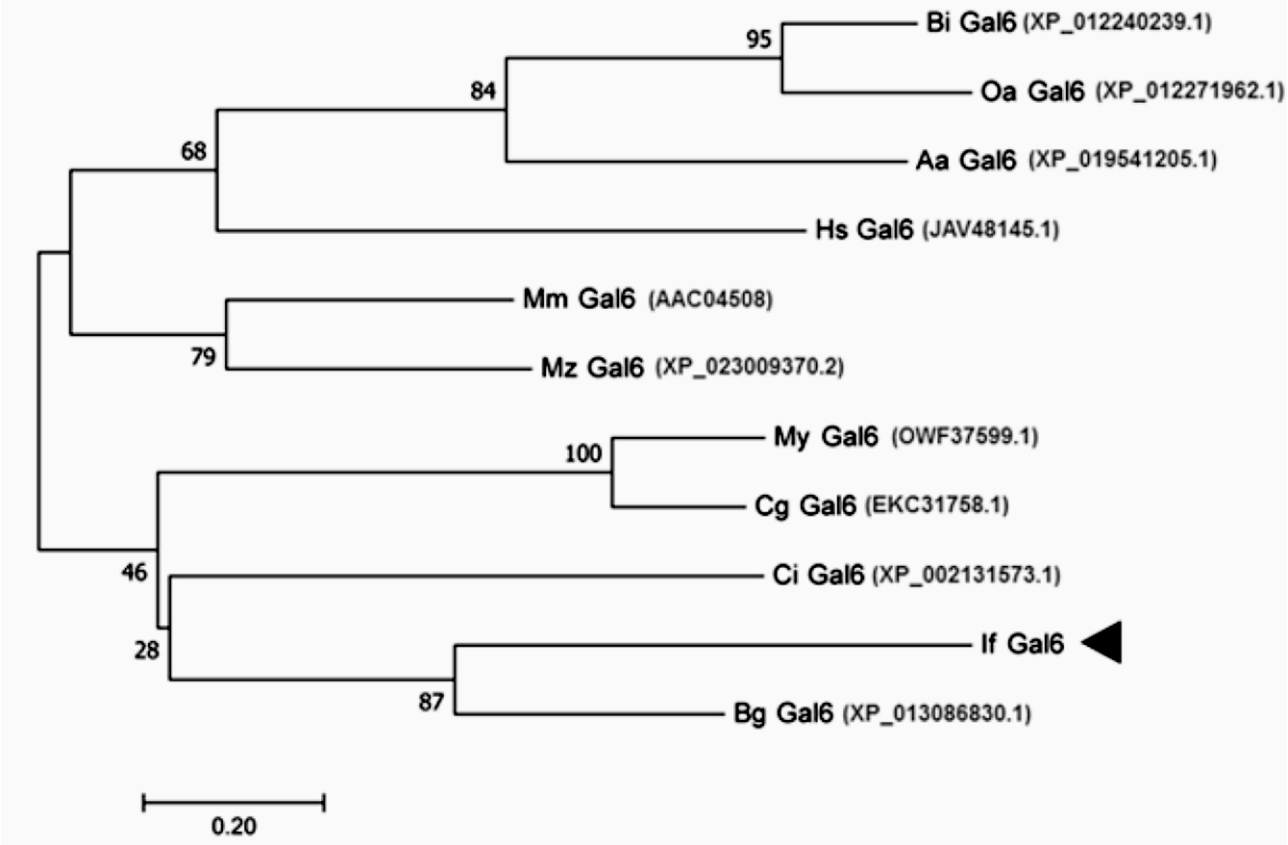


Fig. 7 : Phylogenetic tree analysis of selected invertebrate and vertebrate galectin 6 aa sequences. The analysis used the maximum likelihood method using MEGA7. Evolutionary history is based on the Poisson correction method. The tree with the highest log likelihood (-2233.21) is shown. The percentage of trees in the associated taxa clustered together is shown right next to the branches. The GenBank accession numbers are as listed in figure 6.

(If_Gal6) from the air breathing slug, *I. fruhstorferi*. Although sequence similarity algorithms showed greater identity of If_Gal6 with galectin-6 like protein of *B. glabrata,* the structural characterization indicates the presence of a single CRD unlike galectin-6 orthologs and like prototype galectin ortholog sequences.

Acknowledgments

This research was supported by "Basic Science Research Program" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A3B06034971) and "The Genetic and Genomic Evaluation of Indigenous Biological Resources" funded by the National Institute of Biological Resources (NIBR201703203) and Soonchunhyang University Research Fund.

References

Bulat, T., R. Smidak, F.J. Sialana, G. Jung, T. Rattei, M. Bilban, H. Sattmann, G. Lubec and J. Aradska: Transcriptomic and proteomic analysis of *Arion vulgaris* – Proteins for probably successful survival strategies. *PLoS ONE,* **11**, e0150614 (2016).

Castellanos-Martinez, S., D. Arteta, S. Catarino and C. Gestal: *De novo* transcriptome sequencing of the *Octopus vulgaris* hemocytes using Illumina RNA-Seq technology: Response to the infection by the gastrointestinal parasite *Aggregata octopiana*. *PLoS ONE*, **9**, e107873 (2014).

Diaz-Alvarez, L. and E. Ortega: The many roles of galectin-3, a multi-faceted molecule, in innate immune responses against pathogens. *Mediators Inflamm.,* **2017**, 9247574 (2017).

Ertl, N.G., W.A. O'Connor, A. Papanicolaou, A.N. Wiegand and A. Elizur: Transcriptome analysis of the Sydney rock oyster, *Saccostrea glomerata*: Insights into Molluscan immunity. *PLoS ONE*, **11**, e0156649 (2016).

Feng, C., A. Ghosh, M.N. Amin, T.R. Bachvaroff, S. Tasumi, M. Pasek, A. Banerjee, S. Shridhar, L.X. Wang, M.A. Bianchet and G.R. Vasta: The galectin CvGal2 from the Eastern oyster (*Crassostrea virginica*) displays unique specificity for ABH blood group oligossacharides and differentially recognizes sympatric *Perkinsus* species. *Biochemistry*, **54**, 4711-4730 (2015).

Greenhalgh, C.J., A. Loukas, D. Donald, S. Nikolaou and S.E. Newton: A family of galectins from *Haemonchus contortus*. *Mol. Biochem. Parasitol.,* **15**, 117-121 (2000).

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

Honjo, Y., P. Nangia-Makker, H. Inohara and Raza: Down-regulation of galectin-3 suppresses tumorigenicity of human breast carcinoma cells. *Clin. Cancer Res*., **7**, 661-668 (2001).

Hou, F., Y. Liu, S. He, X. Wang, A. Mao, C. Sun and X. Liu: A galectin from shrimp *Litopenaeus vannamei* is involved in immune recognition and bacteria phagocytosis. *Fish Shellfish Immunol.,* **44**, 584-591 (2015).

Huang, X., N. Tsuji, T. Miyoshi, S. Nakamura-Tsuruta, J. Hirabayashi and K. fujisaki: Molecular characterization and oligossacharide-binding properties of a galectin from the argasid tick *Ornithodorus moubata*. *Glycobiol.,* **17**, 313-323 (2007).

Hwang, E.Y., M.S. Jeong, S.K. Park, S.C. Ha, H.S. Yu and S.B. Jang: Structural basis of carbohydrate recognition and anti-inflammatory modulation by gastrointestinal nematode parasite *Toxoascaris leonina* galectin. *J. Biol. Chem.,* **291**, 25236-25338 (2016).

Kamhawi, S., M. Ramalho-Ortigao, V.M. Pham, S. Kumar, P.G. Lawyer, S.J. Turco, C. Barillas-Mury, D.L. Sacks and J.G. Valenzuela: A role of insect galectins in parasite survival. *Cell*, **119**, 329-341 (2004).

Kang, S.W., S.Y. Park, B.B. Patnaik, H.J. Hwang, J.M. Chung, D.K. Song, S.Y. Park, J.S. Lee, Y.S. Han, H.S. Park and Y.S. Lee: The Protostome Database (PANM-DB): Version 2.0 release with updated sequences. *Korean J. Malacol.,* **32**, 185-188 (2016).

Kumar, S., G. Stecher and K. Tamura: MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.,* **33**, 1870-1874 (2016).

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson and D.G. Higgins: Clustal W and Clustal X version 2.0. *Bioinformatics,* **23**, 2947-2948 (2007).

Mun, S., Y.J. Kim, K. Markkandan, W. Shin, S. Oh, J. Woo, J. Yoo, H. An and K. Han: The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*). *Genome Biol. Evol.,* **9**, 1487-1498 (2017).

Nicholas, K.B. and H.B.J. Nicholas: GeneDoc: A tool for editing and annotating multiple sequence alignments. Citeulike, 6113940 (1997).

Nio-Kobayashi, J. Tissue- and cell-specific localization of galectins, β-galactose-binding animal lectins, and their potential functions in health and disease. *Anat. Sci. Int.,* **92**, 25-36 (2017).

Pace, K.E., T. Lebestky, T. Hummel, P. Arnoux, K. Kwan and L.G. Baum: Characterization of a novel *Drosophila melanogaster* galectin. Expression in developing immune, neural and muscle tissue. *J. Biol. Chem.,* **277**, 13091-13098 (2002).

Patnaik, B.B., J.M. Chung, H.J. Hwang, M.K. Sang, J.E. Park, H.R. Min, H.C. Cho, N. Dewangan, S. Baliarsingh, S.W. Kang, S.Y. Park, Y.H. Jo, H.S. Park, W.J. Kim, Y.S. Han, J.S. Lee and Y.S. Lee: Transcriptome analysis of air-breathing land slug, *Incilaria fruhstorferi* reveals functional insights into growth, immunity and reproduction. *BMC Genomics*, **20**, 154 (2019).

Shi, X.Z., L. Wang, S. Xu, X-W. Zhang, X.F. Zhao, G.R. Vasta and J.X. Wang: A galectin from the Kuruma shrimp (*Marsupenaeus japonicus*) functions as an opsonin and promotes bacterial clearance from the hemolymph. *PLoS ONE*, **9**, e91794 (2014).

Tian, Y., T. Chen, W. Huang, P. Luo, D. Huo, L. Yun, C. Hu and Y. Cai: A new L-type lectin (LvLTLC1) from the shrimp *Litopenaeus vannamei* facilitates the clearance of *Vibrio harveyi*. *Fish Shellfish Immunol.,* **73**, 185-191 (2018).

Ueda, T., Y. Nakamura, C.M. Smith, B.A. Copits, A. Inoue, T. Ojima, S. Matsunaga, G.T. Swanson and R. Sakai: Isolation of novel prototype galectins from the marine ball sponge *Cinachyrella* sp. guided by their modulatory activity on mammalian glutamate-gated ion channels. *Glycobiol.,* **23**, 412-425 (2013).

Vasta, G.R., C. Feng, M.A. Bianchet, T.R. Bachvaroff and S. Tasumi: Structural, functional and evolutionary aspects of galectins in aquatic mollusks: From a sweet tooth to a Trojan horse. *Fish Shellfish Immunol.,* **46**, 94-106 (2015).

Wang, M., H. Wang, H. Zhang, L. Wang, M. Huang, Q. Yi, Y. Guo, Y. Gai, H. Wang, H. Zhang and L. Song: A galectin from *Eriocheir sinensis* functions as pattern recognition receptor enhancing microbe agglutination and hemocytes encapsulation. *Fish Shellfish Immunol.,* **55,** 10-20 (2016).

Wang, W., X. Song, L. Wang and L. Song: Pathogen-derived carbohydrate recognition in Molluscs immune defense. *Int. J. Mol. Sci.,* **19**, 721 (2018).

Yang, R.Y. and F.T. Liu: Galectins in cell growth and apoptosis. *Cell Mol. Life Sci.*, **60**, 267-276 (2003).

Yoshino, T.P., N. Dinguirard, J. Kunert and C.H. Hokke: Molecular and functional characterization of a tandem-repeat galectin from the freshwater snail *Biomphalaria glabrata*, intermediate host of the human blood fluke *Schistosoma mansoni*. *Gene.*, **411**, 46-58 (2008).

Zhang, D., S. Jiang, Y. Hu, S. Cui, H. Guo, K. Wu, Y. Li and T. Su: A multidomain galectin involved in innate immune response of pearl oyster *Pinctada fucata*. *Dev. Comp. Immunol.,* **35**, 1-6 (2011).