1 Molecular characterization of Galectin-8 from Nile tilapia
2 (*Oreochromis niloticus* Linn.) and its response to bacterial infection

3 Sasimanas Unajak^{1,2*}, Nutthida Pholmanee¹, Napat Songtawee³, Kornsom Srikulnath⁴,
 4 Prapansak Srisapoome⁵, Asama Kiataramkul¹, Hidehiro Kondo⁶, Ikuo Hirono⁶, and
 5 Nontawith Areechon⁵

6

7 ¹ Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam Wong
 8 Wan, Chatuchak, Bangkok 10900, Thailand

9 ² Biochemical Research Unit for Feed Utilization Assessment, Faculty of Science, Kasetsart
 10 University, Bangkok 10900, Thailand

11 ³ Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology,
 12 Mahidol University, Bangkok 10700, Thailand

13 ⁴ Laboratory of Animal Cytogenetics and Comparative Genomics, Department of Genetics,
 14 Faculty of Science, Kasetsart University, 50 Ngam Wong Wan, Chatuchak, Bangkok 10900,
 15 Thailand

16 ⁵ Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngam Wong
 17 Wan Road, Chatuchak, Bangkok 10900, Thailand

18 ⁶ Graduate School of Marine Science and Technology, Tokyo University of Marine Science
 19 and Technology, Konan 4-5-7, Minato-Ku, Tokyo 108-8477, Japan

20

21 *Corresponding author

22 Sasimanas Unajak, Ph.D. Department of Biochemistry, Faculty of Science, Kasetsart
 23 University, 50 Ngam Wong Wan, Chatuchak, Bangkok 10900, Thailand

24 Tel: +66 (2) 562 5555, Fax: + 66(2) 561-4627

25 Email: fscissmn@ku.ac.th (S. Unajak)

26 ABSTRACT

27 Galectins belong to the family of galactoside-binding proteins and play a major role
28 in the immune and inflammatory responses of vertebrates and invertebrates. The galectin
29 family is divided into three subtypes based on molecular structure; prototypes, chimera
30 types, and tandem-repeated types. We isolated and characterized the cDNA of galectin-8
31 (*OnGal-8*) in Nile tilapia (*Oreochromis niloticus*). *OnGal-8* consisted of a 966 bp open
32 reading frame (ORF) that encoded a 321 amino acid protein (43.47 kDa). Homology and
33 phylogenetic tree analysis suggested the protein was clustered with galectin-8s from other
34 animal species and shared at least 56.8 % identity with salmon galectin-8. Structurally, the
35 amino acid sequence included two distinct N- and C- terminus carbohydrate recognition
36 domains (CRDs) of 135 and 133 amino acids, respectively, that were connected by a 39
37 amino acid polypeptide linker. The N- and C-CRDs contained two conserved WG-E-I and
38 WG-E-T motifs, suggesting they have an important role in mediating the specific
39 interactions between *OnGal-8* and saccharide moieties such as β -galactoside. The structure
40 of *OnGal-8* was characterized by a two-fold symmetric pattern of 10-and 12-stranded
41 antiparallel β -sheets of both N- and C-CRDs, and the peptide linker presumably formed a
42 random coil similar to the characteristic tandem-repeat type galectin. The expression of
43 *OnGal-8* in healthy fish was highest in the skin, intestine, and brain. Experimental challenge
44 of Nile tilapia with *S. agalactiae* resulted in significant up-regulation of *OnGal-8* in the
45 spleen after 5 d. Our results suggest that *OnGal-8* is involved in the immune response to
46 bacterial infection.

47 **Keywords:** Galectin-8, Nile tilapia, *Oreochromis niloticus*, *Streptococcus agalactiae*

48

49 INTRODUCTION

50 Streptococcosis, the most common bacterial disease in cultured Nile tilapia
 51 (*Oreochromis niloticus*), causes significant losses of cultured tilapia in all regions of
 52 Thailand and in many other countries (Hernandez et al. 2009; Mian et al. 2009). Septicemia
 53 infected fish exhibit erratic swimming, hemorrhaging, exophthalmia, and
 54 meningoencephalitis (Eldar et al. 1995). This disease is caused by infection with Gram-
 55 positive cocci bacteria, primarily *Streptococcus agalactiae*. The bacterium is classified as a
 56 Group B *Streptococcus* (GBS), which are known to be infectious to human neonates,
 57 bovines, and other aquatic animals (Elliott et al. 1990; Wilkinson et al. 1973).

58 Like other vertebrates, the immune system of tilapia has both innate and adaptive
 59 components. The innate immune system serves primarily as a first line of defense against
 60 invading pathogens. During a bacterial infection, the host innate immune system recognizes
 61 the pathogen-associated molecular patterns (PAMPs) on the cell surface of potentially
 62 pathogenic bacteria using host pattern recognition receptors (PRRs), such as Toll-like
 63 receptors (TLRs), C-type lectin receptors, or galectin oligomerization. Lectin is a protein or
 64 a glycoprotein that recognizes carbohydrate moieties on the pathogen and consequently
 65 mediates agglutination, opsonization, signal transduction, cytotoxic activity, and apoptosis.

66 Galectins, a member of the S-type lectins, are a group of proteins that are involved in
 67 the innate immune response and are also linked to adaptive immunity in invertebrates and
 68 vertebrates. Galectins belong to the family of β -galactoside-binding proteins (Barondes et al.
 69 1994) of which 15 galectin members have been characterized to date (Cooper, 2002;
 70 Hirabayashi and Kasai, 1993). Generally, galectins have a conserved region of 130–150
 71 amino acid residues that form a carbohydrate recognition domain (CRD) and contain a
 72 unique WG-E motif that is specific for saccharide binding (Cooper and Barondes, 1999).
 73 Galectins can be divided into three types based on their molecular structure: prototypes,

74 chimera types, and tandem repeat types (Hirabayashi and Kasai, 1993). The prototype
75 (galectin-1, -2, -5, -7, -10, -11, -13 and -14) and chimera type (galectin-3) groups include
76 galectins that have a single CRD and form homodimers, with the exception that the chimera
77 type galectin-3 has an extended peptide on the N-terminus. The tandem-repeated types
78 (galectin-4, -6, -8, -9, and -12) include galectins that have two non-identical CRDs linked by
79 a short peptide.

80 A number of functions have been proposed for galectins, including modulation of
81 the immune system, activation of T-cell proliferation, induction of cell adhesion and
82 inhibition of many types of tumors and cancers (Elola et al. 2007). In fish, relatively few
83 types of lectins have been identified to date. These include galectin-1 in flounder (Liu et al.
84 2013), sea bass (Poisa-Beiro et al. 2009), and Atlantic cod (Rajan et al. 2013), galectin-3 in
85 tongue sole (Chen et al. 2013), and galectin-9 in rainbow trout (O'Farrell et al. 2002). These
86 galectins exhibit a range of functions, including developmental (cod galectin-1) and immune.

87 We identified a cDNA encoding gene similar to galectin-8 (Gal-8) from an EST
88 library that was generated from the spleen of *S. agalactiae* infected Nile tilapia. The full-
89 length cDNA of tilapia galectin-8 (*OnGal-8*) was subsequently identified, and we evaluated
90 the change in galectin expression in Nile tilapia following bacterial infection. Additionally,
91 we conducted a quantitative analysis of the tissue specific expression of *OnGal-8* in healthy
92 fish.

93

94

95 MATERIAL AND METHODS

96

97 1. *Sample collection*

98 Nile tilapia (*Oreochromis niloticus*; body weight ~200 g) were purchased from a
 99 local farm in Thailand. Prior to tissue sampling or the bacterial challenge experiment, the
 100 fish were acclimated for a week in aerated cement tanks (2 tons) and fed with commercial
 101 pellet feed. To analyze tissue specific levels of galectin gene expression we dissected
 102 healthy fish and removed samples of the head kidney, trunk kidney, liver, spleen, intestine,
 103 muscle, brain, and skin. The tissue samples were submerged in TRI reagent® (Molecular
 104 Research Center) and frozen at –80 °C until analysis.

105 To conduct the bacterial challenge, the fish were divided into three groups: control,
 106 PBS, and *S. agalactiae*. The fish in the PBS and *S. agalactiae* group were injected
 107 intramuscularly with 100 µl of PBS or 100 µl *S. agalactiae* suspension in PBS (1×10^8 CFU
 108 ml⁻¹), respectively. The fish in the control group did not receive and injection. We collected
 109 the spleen from three fish in each group 0 h, 6 h, 12 h, 1 d, 2 d, 3 d, 4 d, 5 d, and 6 d post-
 110 injection. The tissue samples were submerged in TRI reagent® (Molecular Research Center)
 111 and frozen at –80°C until analysis.

112

113 2. *RNA isolation and cDNA synthesis*

114 To conduct rapid amplification of cDNA ends-polymerase chain reaction (RACE-
 115 PCR) analysis, total RNA was prepared from the head kidney of *S. agalactiae* infected
 116 tilapia following the manufacturer's instructions for TRI reagent ®. The total RNA (25 µg)
 117 was purified using an Illustra mRNA Purification Kit (GE Healthcare) to isolate mRNA.
 118 Purified mRNA (1 µg) was then used as a template for first strand cDNA synthesis using a
 119 SMARTer™ RACE cDNA amplification kit (Clontech), following the manufacturer's

120 instructions.

121 To evaluate the tissue distribution of galectin expression in healthy and infected fish,
 122 total RNA was isolated from tissues in TRI Reagent® following the manufacturer's
 123 instructions. All the RNA samples were treated with DNaseI (Promega) to remove residual
 124 DNA by incubating at 37°C for 30 min followed by re-extraction with TRI reagent®. First
 125 strand cDNA was prepared from 4 µg of total RNA by RevoScript™ RT PreMix (Oligo
 126 (dT)15 Primer) (INtRON, Korea) with an oligo(dT)₁₈ primer following the manufacturer's
 127 instructions.

128

129 ***3. Cloning of full-length OnGal-8 cDNA***

130 To obtain the full-length *OnGal-8* cDNA, the 5' partial *OnGal-8* sequence was
 131 identified using a SMARTer™ RACE cDNA amplification kit (Clontech). The partial
 132 cDNA sequence of *OnGal-8* was obtained from the EST library of *S. agalactiae* infected
 133 Nile tilapia (accession no. FF281049.1). Based on the EST sequences, a gene specific
 134 primer (Gal-8-RACE-R1) was designed to amplify in the 5'direction to recover the full-
 135 length cDNA of *OnGal-8* (Table 1). Another pair of gene specific primers (Gal-8F and Gal-
 136 8R) were used to confirm the full-length of *OnGal-8* (Table 1). The PCR products were
 137 cloned using the pGEM®-T-Easy Vector System I (Promega Corporation, Madison, WI,
 138 USA). The nucleotide sequences of the DNA fragments were determined using MacroGen
 139 Sequencing Service (Korea). All nucleotide sequences were used to search for homology
 140 with available information in the National Center for Biotechnology Information (NCBI)
 141 database using the BLASTX and BLASTn programs
 142 (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and were deposited in Genbank [accession number:
 143 KP663372]

144 ***4. Bioinformatics analysis***

145 The predicted molecular mass and pI were determined using Compute pI/Mw
 146 (http://web.expasy.org/compute_pi/). A similarity search of nucleotides and proteins was
 147 performed using BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Multiple sequence
 148 alignment and similarity analyses were carried out using ClustalW 2
 149 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), BOXSHADE 3.21
 150 (http://www.ch.embnet.org/software/BOX_form.html), and MATGAT 2.0. The protein
 151 conserved domain structure was analyzed using SMART (<http://smart.embl-heidelberg.de/>)
 152 and NCBI conserved domain architecture
 153 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). We conducted a signal peptide
 154 analysis using SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Glycosylation of the
 155 protein was analyzed using NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>). We
 156 constructed a phylogenetic tree using MEGA (version 4.0) based on the neighbor-joining
 157 method with bootstrapping of 1,000 replications. Exon-intron organization was inferred
 158 using Ensembl (<http://www.ensembl.org>). The conserved synteny and gene order were
 159 compared with that of human (*Homo sapiens*), mouse (*Mus musculus*), rat (*Rattus*
 160 *norvegicus*), chicken (*Gallus gallus*), frog (*Xenopus (Silurana) tropicalis*), and Anolis lizard
 161 (*Anolis carolinensis*) using the Ensembl Genome Browser (<http://www.ensembl.org>) (Table
 162 2).

164 **5. Three-dimensional structure of OnGal-8**

165 The deduced amino acid sequence of OnGal-8 was submitted to the SWISS-MODEL
 166 automated server (Arnold et al. 2006; Schwede et al. 2003) to generate a three dimensional
 167 homology model using the crystal structure of a protease-resistant mutant form of human
 168 galectin-8 (PDB entry 4FQZ) as a template. The amino acid sequence of human galectin-8
 169 and that of *O. niloticus* share ~46% identity (~63% similarity). To improve the quality of the

170 structure generated by the server, the model was then processed using Chiron
171 (Ramachandran et al. 2011), an automated minimization server, which efficiently resolves
172 severe atomic clashes in low-resolution crystal structures or low-identity template homology
173 models with minimal perturbation in the protein backbone. The quality of both the
174 homology and refined structures was assessed based on the geometry. The stereo-chemical
175 properties were evaluated by analysis of Ramachandran plots using PROCHECK v3.5.4
176 (Laskowski et al. 1993) after submitting the atomic coordinates for SWISS-MODEL
177 structural assessment (Bordoli et al. 2008). The compatibility of the atomic models (3D)
178 with their own amino acid sequence (1D) was analysed by the Verify3D (Eisenberg et al.
179 1997; Luthy et al. 1992) on the UCLA-DOE server (<http://nihserver.mbi.ucla.edu>).

180 Molecular docking of galactose (Gal) and its derivatives, lactose (Lac) and N-
181 Acetylgalactosamine (GalNAc), to the refined model of *OnGal-8* was performed using
182 AutoDock 4.2 (Huey et al. 2007; Morris et al. 2009). The three dimensional structures of all
183 three carbohydrates were obtained using GLYCAM (<http://www.glycam.com>), after
184 optimization of the atomic coordinates using the Glycam06 parameters (Kirschner et al.
185 2008). The rotational bonds of the carbohydrates were treated as flexible whereas those of
186 the *OnGal-8* model were regarded as rigid. Grid boxes were created to cover the N-terminal
187 CRD of the protein, in which the grid spacing was 0.375 Å and the box size was set at 50 ×
188 70 × 70 Å³ and the box center was set at 53.018 Å (x), 2.612 Å (y), and 0.435 Å (z). The
189 Lamarckian Genetic Algorithm (LGA; 250 runs) was used to define the search parameters
190 and the population size was set at 150. The three dimensional structures of the protein–
191 saccharide complexes with the lowest binding energy and for the highest populations were
192 analyzed and visualized using Discovery Studio 3.5 Visualizer (Accelrys Inc., San Diego,
193 USA).

194

195 **6. Determination of *OnGal-8* mRNA expression by real-time PCR analysis**

196 To determine the level of *OnGal-8* mRNA expression in healthy fish, we performed
 197 quantitative real-time PCR analysis using the previously prepared cDNA with the specific
 198 primers, Gal-8RTF (position A⁶⁴⁶ – C⁶⁷⁰ on *OnGal-8*) and Gal-8RTR (position C⁷²⁰ – T⁷³⁹
 199 on *OnGal-8*) (Table 1). The beta-actin gene, a constitutively expressed gene, was used as an
 200 internal reference with gene specific primers (β-actinF and β-actinR) (Table 1). First strand
 201 cDNA (1 µl) was added to a reaction solution containing 1X qPCR EvaGreen Master mix
 202 E4 (GeneON), 200 nM forward primer, 200 nM reverse primer. The reaction was performed
 203 at 94°C for 1 min followed by 45 cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 30 s.

204 The level of *OnGal-8* transcript expression in the control and bacterial challenge
 205 groups were normalized with the transcript expression level of beta-actin at different time
 206 intervals. The relative expression level in the bacterial stimulated group was then
 207 normalized against the expression of transcripts in the control group. We tested for
 208 differences in expression levels between groups using one-way analysis of variance
 209 (ANOVA) followed by Duncan's new multiple range tests in SPSS. Differences were
 210 considered significant at P<0.05.

211

212

213 RESULTS

214 1. *cDNA cloning and molecular characterization of OnGal-8*

215 The 5' *OnGal-8* cDNA was amplified by a single round of 5'RACE with a
 216 specific primer. The complete sequence of *OnGal-8* consisted of 1,169 bp including an 8 bp
 217 5' untranslated region (UTR) and a 195 bp 3' UTR. The sequence also included a canonical
 218 polyadenylation signal sequence (AATAAA) and a poly (A) tail (Fig. 1). The open reading
 219 frame (ORF) of *OnGal-8* was 966 bp and encoded a 321 amino acid polypeptide with a
 220 predicted molecular mass of 43.47 kDa and a theoretical *pI* of 8.45. SignalP analysis
 221 revealed this protein has no signal peptide and no transmembrane regions. A conserved
 222 domain analysis revealed tandem[- delete]repeat carbohydrate recognition domains (CRDs)
 223 connected by a linker peptide. The N-CRD consisted of 135 amino acid residues (Ile¹⁵-
 224 Leu¹⁴⁹) and the C-CRD consisted of 133 amino acid residues (Ile¹⁸⁹-Leu³²¹) that were
 225 specific for saccharide binding activity (GLECT domain) (Fig. 1). The theoretical *pI* was
 226 7.87 for N-CRD and 9.22 for C-CRD (Fig. 2). The sequence of *OnGal-8* was deposited in
 227 GenBank (Acc. number KP663372)

228 Analysis of post-translational modification of the putative *OnGal-8* amino acid
 229 sequence revealed several putative phosphorylation sites on this protein at the linkage
 230 between the N-CRD and C-CRD. Four putative *O*-link glycosylation were predicted at
 231 Tyr¹⁵⁶, Tyr¹⁶³, Tyr¹⁶⁷, and Tyr¹⁷³. Moreover, *OnGal-8* has 6 cysteine (Cys) residues (Cys⁵⁰,
 232 Cys⁷², Cys⁷⁵, Cys⁸², Cys²¹⁹, and Cys²⁵²), 4 Cys residues on the N-terminus CRD, and 2
 233 residues on the C-terminus CRD, which we propose form three internal disulfide linkages
 234 (Cys⁵⁰ - Cys²¹⁹, Cys⁷² - Cys⁸² and Cys⁷⁵- Cys²⁵²). Moreover, sixteen predicted
 235 phosphorylation sites were observed (Ser⁵³, Ser¹³⁶, Ser¹³⁸, Ser¹⁵⁰, Ser¹⁵², Ser¹⁵³, Ser¹⁵⁴, Ser¹⁵⁷,
 236 Ser¹⁶¹, Ser¹⁷⁰, Ser¹⁸³, Ser¹⁸⁴, Ser²⁵⁰, Ser¹⁷⁹, Ser³⁰³, and Thr¹³⁴ (Fig. 1).

237

238 **2. Homology and phylogenetic tree analysis of *OnGal-8***

239 Multiple sequence alignment of *OnGal-8* and other animal galectin-8 genes revealed
 240 that *OnGal-8* shares a consensus sequence with the canonical galectin family (~HFNPR (-)
 241 ~VCN ~WG ~E ~I ~PF ~E ~V ~R and ~H ~NPRL ~WG ~E ~(-) ~PF ~G ~(-) ~R)
 242 in both the N- and C-CRD (Fig. 3) (Hadari et al. 1995). Moreover, the essential amino acids
 243 corresponding to the carbohydrate binding motif differed between the domains: WG-E-I and
 244 WG-E-V for the N- and C-CRD domains, respectively. The deduced amino acid sequence
 245 of *OnGal-8* shared 58 and 53% identity with *Salmo salar* Gal-8 (SsGal8, NP_001133778)
 246 and *Gallus gallus* Gal-8 (GgGal8, NP_00101843), respectively (Table 3).

247 The phylogenetic tree analysis of *OnGal8* suggested the protein is located in a
 248 cluster of galectin-8 that is discrete from other tandem-repeat types (Gal-4,-9 and -12),
 249 proto-type (Gal-1, -2, -7, -13, -14 and -16), and chimera-type galectins (Gal-3). With
 250 emphasis on the galectin-8 cluster, *OnGal-8* was clustered together with salmon (*Salmo*
 251 *salar*) galectin-8 (Fig 4). These results suggest that *OnGal-8* is classified as one of member
 252 in the galectin-8 gene family.

253

254 **3. Genomic structure and genetic linkage of *OnGal-8***

255 The predicted exon-intron organization of galectin-8 revealed that *OnGal-8* is
 256 composed of 9 exons interspaced by 8 introns. This arrangement is similar to that of
 257 amphibians, avians, and mammals (Fig. 5). Analysis of the genetic linkage in tilapia
 258 demonstrated that *OnGal-8* is located on SCAFFOLD GL831182.1, and is linked to the
 259 *NMBR* and *GPR126* genes. This conserved synteny was also observed in frogs, chickens,
 260 and humans. However, *Gal8*, and *NMBR* and *GPR126* are localized separately in mice
 261 (chromosomes 13 and 10, respectively) and rats (chromosomes 17 and 1, respectively) (Fig.
 262 6).

263 **3.4 Three dimensional analysis of *OnGal-8***

264 We used homology modeling followed by a rapid refinement approach to construct the
265 three dimensional structure of *OnGal-8*. The final model of the *OnGal-8* structure is shown
266 in Fig. 7A. The *OnGal-8* structure consists of two carbohydrate recognition domains (N-
267 CRD and C-CRD) connected by 33 amino acid residues in the linker region. The root mean
268 square deviation (RMSD) value between human galactin-8 and *OnGal-8* was 0.48 Å. Both
269 the N-CRD and C-CRD were comparable to each other and have a β -sandwich structure
270 formed by antiparallel β -sheets, a typical fold of the galectin CRD structure. The
271 Ramachandran plots revealed that the stereo-chemical geometries of the *OnGal-8* structure
272 were more satisfactory when the homology model was refined (an increase in the number of
273 residues in the most favored regions; data not shown). The accuracy of the overall model
274 was improved after refinement, as indicated by the 3D-1D averaged Verify3D scores.
275 However, 22 amino acid residues (S¹⁶¹-S¹⁸²) in the linker region yielded scores lower than
276 0.20 (data not shown). The inaccuracy of the model for the linker region was not unexpected
277 because the target-template alignment contained large gaps for the 33 amino acid residues in
278 the linker region (data not shown) and the SWISS-MODEL used *de novo* loop modeling
279 techniques that are less reliable than the template-based part of the model (Bordoli et al.
280 2008). However, neither the orientations of the two CRDs or the flexibility of the linker
281 region were questions of interest in our study. The results of the structural validations
282 suggested that the two CRDs of the refined *OnGal-8* model were reliable and useful for
283 further analyses.

284 We used molecular docking analyses to evaluate the binding between *OnGal-8* and
285 galactose (Gal) and its derivatives, lactose (Lac) and N-acetylgalactosamine (GalNAc). The
286 docking validation was performed by redocking the co-crystallized lactose molecule (Lac)
287 into its respective binding site on the N-CRD of human galectin-8 (pdb entry 4FQZ). The

orientations of the redocked Lac reproduced the conformation of the co-crystallized molecule with an RMSD value of 0.96 Å and 0.79 Å for the whole Lac molecule and Gal moiety, respectively (data not shown). This observation supported the reliability of our method. To perform the molecular docking analysis for *OnGal-8*, the N-CRD was used as the search region. The center coordinates of the grid box were set according to those used in the redocking of 4FQZ when *OnGal-8* was superimposed on human galectin-8 by fitting the backbones of their N-CRD. Subsequently, the Gal, Lac, and GalNac molecules were docked onto the *OnGlec-8* N-CRD with 250 genetic algorithm runs for each docked experiment. We obtained the predicted binding free energies and number of docked poses for the first three highest populations in each cluster for each docking experiment (Supplementary data Table S1). The lowest binding energy from these three clusters of the Gal-*OnGal-8* complex was comparable to that of the GalNac-*OnGal-8* complex (-3.73 ± 0.26 kcal/mol). However, these values were lower than the binding energy of the Lac-*OnGal-8* complex (-2.37 ± 0.32 kcal/mol).

Binding mode analyses revealed that all three carbohydrates were docked onto the same binding site, although they exhibited moderate diversity in their conformations (Fig. 7B-7D). The amino acid residues within 6 Å radii surrounding the *OnGal-8* binding site interacting with each of the three carbohydrates are listed in Supplementary data Table S2. For each docking system, the complex representing the highest population in a cluster was used for detection of hydrogen bonds (Fig. 7B-7D). The Gal moiety in both the Lac and GalNac complexes was located at a similar position and shared some common interactions with *OnGal-8* when compared with the Gal-*OnGal-8* complex. The residues that interact significantly via hydrogen-bonds with the Gal moiety in all three complexes include R⁴³ and E⁸⁶, of which the side chain acts as a donor or an acceptor, respectively. Further studies are needed to identify the correct binding mode of the Gal moiety against *OnGal-8* and to

313 investigate the interactions involving the protein–saccharide complexes.

314

315 ***4. Tissue expression of OnGal-8 and its response during S. agalactiae challenge***

316 We used quantitative PCR (qPCR) analysis to evaluate *OnGal-8* expression in
317 different tissues of healthy Nile tilapia. *OnGal-8* expression was detected in all tissues (Fig.
318 8A). The level of expression was highest in the skin, intestine and brain, moderate in the
319 head kidney, trunk kidney, liver, gills, muscle, gonad, heart, and PBLs, and lowest in the
320 spleen. In skin, whereas the highest expression of *OnGal-8* transcript was detected in
321 healthy fish, the expression of this gene was not altered after bacterial infection (Fig. S1).
322 However, the expression of *OnGal-8* gradually increased in the spleen following infection
323 with *S. agalactiae*, and was significantly higher (30 fold) after day 5 but decreased thereafter
324 (Fig. 8B).

325

326

327 DISCUSSION

328 The roles of galectin-8 have been elucidated in a range of animal species during the
329 last decade. Galectin-8 has several functions including the promotion of plasma cell
330 formation, T-cell and B-cell proliferation (Tribulatti et al. 2009; Tsai et al. 2011), apoptosis
331 induction (Hadari et al. 2000; Norambuena et al. 2009; Tribulatti et al. 2007), and
332 modulation of innate and adaptive immunity. Galectin-8 is also involved in the regulation of
333 many types of tumors, can be used as a biomarker of cancer, and acts as an initiating factor
334 of autophagy or xenophagy pathways following invasion of intracellular bacteria. We report
335 the isolation of the first fish galectin-8 from Nile tilapia (*O. niloticus*) and describe its
336 response to bacterial, *S. agalactiae*, infection.

337 *OnGal-8* is classified as a tandem repeat galectin in which the N- and C-terminus are
338 linked by a short peptide. The physiological properties of *OnGal-8* were inferred based on a
339 number of factors, including signal peptide determination, glycosylation, and disulfide bond
340 formation. We found no evidence of a signal peptide or transmembrane domains in *OnGal-8*.
341 In general, galectin proteins accumulate in the cytosol prior to being secreted through a
342 leader peptide-independent pathway (Houzelstein et al. 2004; Rabinovich and Toscano,
343 2009). Consistent with observations of galectin in other species, Galectin-8 is secreted into
344 the cytoplasm. We speculate that *OnGal-8* is localized primarily in the cytoplasm and uses a
345 non-classical transportation pathway to secrete target proteins.

346 Post-translational modification analysis revealed four potential *O*-linked
347 glycosylations and several putative phosphorylation sites in *OnGal-8*. Although, there is no
348 evidence of glycosylation modification in galectin-8 from other species rather than rat,
349 extensive post-translational modification as a result of glycosylation or phosphorylation has
350 been observed in rat galectin-8 (Hadari et al. 1995). The glycosylation of galectin proteins
351 can potentially play a significant role in maintaining their function. For example,

glycosylation of galectin-1 from buffalo heart tissue maintains the active conformation of galectin protein (Ashraf et al. 2010). Similarly, disulfide linking can play an important role in maintaining the function of proteins. Our data suggest that *OnGal-8* forms three internal disulfide linkages at six potential Cys residues. This differs from the arrangement of human and mouse galectin-8, both of which contain 4 Cys residues (Hadari et al. 1995). The prototype form of galectin-1 requires only 3 Cys residues to be active, with the usual disulfide bond formed by Cys¹⁶-Cys⁸⁸ and a multimeric bond formed at Cys². Disulfide bond formation has an important role in promoting the interaction between galectin-1 and a lactose moiety. Under oxidizing conditions, the protein is unable to bind to the lactose. We speculate that disulfide bond formation is critical to the function of *OnGal-8*, as with galectin-1.

Our results suggest that *OnGal-8* has diverged from other tandem-repeat galectins. It is located in a different evolutionary cluster from the avian, reptilian, and mammalian galectin-8 family. The evolutionary origin of *OnGal-8* can be inferred based on the conserved genetic linkage of *Gal8* with *NMBR* and *GPR126* in fish and frogs. A similar genome organization has also been observed in reptiles and mammals. These three genes are located on chromosome 1p in the green anole lizard (*Anolis carolinensis*) and on chromosome 3 in chickens (*Gallus gallus*) (Alfoldi et al. 2011). A comparison of the chromosome maps of squamate reptiles and chickens revealed that the genetic linkage of chicken chromosome 3 is homologous with the squamate reptile chromosome 1p (Srikulnath et al. 2009). However, *Gal8*, and *NMBR* and *GPR126* are located on separate chromosomes in mice (chromosomes 13 and 10, respectively) and rats (chromosomes 17 and 1, respectively). Taken together, these observations suggest that the genetic linkage of *Gal8*, *NMBR*, and *GPR126* is highly conserved in vertebrate evolution and that the divergence of the different linkage groups occurred when the Sauropsida (all existing

377 reptiles and birds) diverged from the Synapsida (mammals) around 320 million years ago.

378 We evaluated the role of galectin-8 in carbohydrate recognition using computational
 379 modeling (SWISS-MODEL) to explain and predict the function of *OnGal-8*. The SWISS-
 380 MODEL server has been widely used to predict the potential tertiary structure of a particular
 381 protein. We successfully obtained the predicted 3D structure of *OnGal-8* and verified its
 382 quality using RMSD, Ramachandran plot and 3D score analyses. The predicted 3D structure
 383 of *OnGal-8* consisted of a typical galectin structure characterized by antiparallel β -sheets
 384 that formed a β -sandwich structure. The carbohydrate binding site of the N-CRD predicted
 385 by our docking calculations consisted of several critical amino acid residues that are
 386 conserved among galectin families and can form hydrogen bonds with several types of
 387 saccharides via potential residues, including R⁴³, H⁶³, R⁶⁷, W⁸³, and E⁸⁶. This observation is
 388 consistent with the broad saccharide specificity of galectin-8's N-CRD (Carlsson et al.
 389 2007), whereby multiple amino acid residues within the carbohydrate binding pocket are
 390 essential for recognition of a variety of saccharides (Yoshida et al. 2012).

391 Recently, a number of members of galectin protein were predicted from genome
 392 sequence of Nile tilapia such as galectin-3, -4, and -9. In the database, there are 2 variants of
 393 predicted Nile tilapia Gal-8; transcript variant x1 (accession no. XM_003446634.2 and
 394 XM_003446206.2) and transcript variant x2 (accession no. XM_005452471.1 and
 395 XM_005477691). In this study, *OnGal-8* demonstrated sequence resembled to transcript
 396 variant x2. However, the detection of *OnGal8* expression by qRT-PCR was amplified by
 397 specific primer recognized at unique region on *OnGal-8* but not recognized Gal-8 variant 1
 398 and other galectins 9 (Fig. S2 and S3) yield only single amplicon corresponded to *OnGal-8*
 399 (data not shown).

400 We detected *OnGal-8* expression in most tissue types of Nile tilapia (e.g., skin,
 401 intestine, spleen, and liver), consistent with observations in other species (Hadari et al.

1995). In fish, several tandem-repeat galectins are expressed in a range of tissues, including the head kidney, gills, heart, and spleen (Inagawa et al. 2001; Kong et al. 2012). This is the first report demonstrating that expression of *OnGal-8* is highest in the skin of healthy fish. There are several types of galectin that are involved in cutaneous immunity, including Gal-1, Gal-3, and Gal-9. The localization of *OnGal-8* in the skin suggests it plays a role in the initial immune defense response, although, the expression of *OnGal-8* in the skin were not affected by bacterial infection. Such a role would be consistent with observations of the function of mucosal galectin in *Conger myriaster* (Brevoort), including bacterial coagulation and opsonization (Nakamura et al. 2006) which demonstrated that *OnGal-8* might serve as basal innate immunity to prevent extracellular pathogen.

A number of functions have been proposed for galectin-8, with emphasis on immunomodulation activation and pathogen recognition of both Gram-negative and Gram-positive bacteria. After *S. agalactiae* infection, *OnGal-8* was expressed in a time-dependent manner in the current study. We speculate that the immunomodulating function of *OnGal-8* involves two pathways. First, galectin-8 may act as the initiator of autophagy activation. Autophagy is a defense mechanism that targets invading pathogens such as *Salmonella typhimurium* (Huang and Brumell, 2014). During bacterial infection, galectin-8 triggers a host-vesicle harboring the invasive bacteria that acts as a pattern-recognition receptor (PRR) for the bacterial glycans. The induction of autophagy is initiated by the recruiting autophagy adaptor proteins such as NDP52 (Thurston et al. 2012), which later activate ubiquitin-mediated degradation. Second, the activation of galectin-8 may play a role in cellular defense. Galectin-8 is expressed during plasma cell differentiation and promotes plasma cell formation. Therefore, galectin-8 is essential for the proliferation of T-cell and B-cell lymphocytes. For example, galectin-8 stimulates the proliferation of human CD4 T-cells (Cattaneo et al. 2011). Moreover, galectin-8 binds to mature B cells at a higher affinity than

427 to plasma cells (Tsai et al. 2011). The delayed up-regulation of *OnGal-8* corresponds to the
 428 period of specific antibody production after immunization of Nile tilapia with recombinant
 429 cell surface proteins from *S. agalactiae* (Yi et al. 2014) or stimulation of silver catfish
 430 (*Rhamdia quelen*) with inactivated *Aeromonas hydrophila* (Kreutz et al. 2014). Taken
 431 together, the up-regulation of *OnGal-8* may contribute to the activation of the cell-specific
 432 immune response.

433 In summary, we identified and characterized *OnGal-8* isolated from *S. agalactiae*
 434 infected Nile tilapia. *OnGal-8* exhibited an interesting pattern of transcriptional expression,
 435 evidenced by the delayed response to bacterial infection. We speculate that *OnGal-8*
 436 functions to limit bacterial invasion in two ways, by detection of cytosolically exposed
 437 surface glycans that are present on the outer surface of the plasma membrane (Tam and
 438 Jacques, 2014) and by induction of autophagy or the proliferation and differentiation of cells
 439 involved in cellular defense pathways. Thus, our future research will focus on the
 440 immunomodulation function of *OnGal-8* to determine the contribution of the protein to the
 441 defense against pathogen invasion.

442

443

444 **Acknowledgments**

445 This research is supported by the Kasetsart University Research and Development Institute
 446 (KURDI), Thailand, JSPS-NRCT Asian Core University Program and Japan Science and
 447 Technology Agency/ Japan International Cooperation Agency, Science and Technology
 448 Research Partnership for Sustainable Development (JST/ JICA, SATREPS). NP is
 449 supported by Thailand Research Fund under Project: the Research and Researchers for
 450 Industries Scholarship (RRI) for Master (Grant number MRG-WI545S009).

451

452

453 **REFERENCES**

- 454 Alföldi, J., F. Di Palma, M. Grabherr, C. Williams, L. Kong, E. Mauceli, P. Russell, C.B.
 455 Lowe, R.E. Glor, J.D. Jaffe, D.A. Ray, S. Boissinot, A.M. Shedlock, C. Botka, T.A.
 456 Castoe, J.K. Colbourne, M.K. Fujita, R.G. Moreno, B.F. ten Hallers, D. Haussler, A.
 457 Heger, D. Heiman, D.E. Janes, J. Johnson, P.J. de Jong, M.Y. Koriabine, M. Lara,
 458 P.A. Novick, C.L. Organ, S.E. Peach, S. Poe, D.D. Pollock, K. de Queiroz, T.
 459 Sanger, S. Searle, J.D. Smith, Z. Smith, R. Swofford, J. Turner-Maier, J. Wade, S.
 460 Young, A. Zadissa, S.V. Edwards, T.C. Glenn, C.J. Schneider, J.B. Losos, E.S.
 461 Lander, M. Breen, C.P. Ponting and K. Lindblad-Toh. 2011. The genome of the
 462 green anole lizard and a comparative analysis with birds and mammals. **Nature**.
 463 477(7366): 587-591.
- 464 Arnold, K., L. Bordoli, J. Kopp and T. Schwede. 2006. The SWISS-MODEL workspace: a
 465 web-based environment for protein structure homology modelling. **Bioinformatics**.
 466 22(2): 195-201.
- 467 Ashraf, G.M., S. Rizvi, S. Naqvi, N. Suhail, N. Bilal, S. Hasan, M. Tabish and N. Banu.
 468 2010. Purification, characterization, structural analysis and protein chemistry of a
 469 buffalo heart galectin-1. **Amino Acids**. 39(5): 1321-1332.
- 470 Barondes, S.H., V. Castronovo, D.N. Cooper, R.D. Cummings, K. Drickamer, T. Feizi, M.A.
 471 Gitt, J. Hirabayashi, C. Hughes, K. Kasai and et al. 1994. Galectins: a family of
 472 animal beta-galactoside-binding lectins. **Cell**. 76(4): 597-598.
- 473 Bordoli, L., F. Kiefer, K. Arnold, P. Benkert, J. Battey and T. Schwede. 2008. Protein
 474 structure homology modeling using SWISS-MODEL workspace. **Nat. Protocols**.
 475 4(1): 1-13.
- 476 Carlsson, S., C.T. Oberg, M.C. Carlsson, A. Sundin, U.J. Nilsson, D. Smith, R.D.
 477 Cummings, J. Almkvist, A. Karlsson and H. Leffler. 2007. Affinity of galectin-8 and

- its carbohydrate recognition domains for ligands in solution and at the cell surface.
- Glycobiology**. 17(6): 663-676.
- Cattaneo, V., M.V. Tribulatti and O. Campetella. 2011. Galectin-8 tandem-repeat structure is essential for T-cell proliferation but not for co-stimulation. **Biochem J**. 434(1): 153-160.
- Chen, C., H. Chi, B.G. Sun and L. Sun. 2013. The galectin-3-binding protein of *Cynoglossus semilaevis* is a secreted protein of the innate immune system that binds a wide range of bacteria and is involved in host phagocytosis. **Dev Comp Immunol**. 39(4): 399-408.
- Cooper, D.N. 2002. Galectinomics: finding themes in complexity. **Biochim Biophys Acta**. 1572(2-3): 209-231.
- Cooper, D.N. and S.H. Barondes. 1999. God must love galectins; he made so many of them. **Glycobiology**. 9(10): 979-984.
- Eisenberg, D., R. Lüthy and J.U. Bowie. 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. **Journal** Volume 277(issue): 396-404.
- Eldar, A., Y. Bejerano, A. Livoff, A. Horovitz and H. Bercovier. 1995. Experimental streptococcal meningo-encephalitis in cultured fish. **Vet Microbiol**. 43(1): 33-40.
- Elliott, J.A., R.R. Facklam and C.B. Richter. 1990. Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. **J Clin Microbiol**. 28(3): 628-630.
- Elola, M.T., C. Wolfenstein-Todel, M.F. Troncoso, G.R. Vasta and G.A. Rabinovich. 2007. Galectins: matricellular glycan-binding proteins linking cell adhesion, migration, and survival. **Cell Mol Life Sci**. 64(13): 1679-1700.
- Hadari, Y.R., R. Arbel-Goren, Y. Levy, A. Amsterdam, R. Alon, R. Zakut and Y. Zick. 2000. Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis. **J**

- 503 **Cell Sci.** 113 (Pt 13): 2385-2397.
- 504 Hadari, Y.R., K. Paz, R. Dekel, T. Mestrovic, D. Accili and Y. Zick. 1995. Galectin-8. A
505 new rat lectin, related to galectin-4. **J Biol Chem.** 270(7): 3447-3453.
- 506 Hernandez, E., J. Figueroa and C. Iregui. 2009. Streptococcosis on a red tilapia,
507 Oreochromis sp., farm: a case study. **J Fish Dis.** 32(3): 247-252.
- 508 Hirabayashi, J. and K. Kasai. 1993. The family of metazoan metal-independent beta-
509 galactoside-binding lectins: structure, function and molecular evolution.
510 **Glycobiology.** 3(4): 297-304.
- 511 Houzelstein, D., I.R. Goncalves, A.J. Fadden, S.S. Sidhu, D.N. Cooper, K. Drickamer, H.
512 Leffler and F. Poirier. 2004. Phylogenetic analysis of the vertebrate galectin family.
513 **Mol Biol Evol.** 21(7): 1177-1187.
- 514 Huang, J. and J.H. Brumell. 2014. Bacteria-autophagy interplay: a battle for survival. **Nat**
515 **Rev Microbiol.** 12(2): 101-114.
- 516 Huey, R., G.M. Morris, A.J. Olson and D.S. Goodsell. 2007. A semiempirical free energy
517 force field with charge-based desolvation. **Journal of Computational Chemistry.**
518 28(6): 1145-1152.
- 519 Inagawa, H., A. Kuroda, T. Nishizawa, T. Honda, M. Ototake, U. Yokomizo, T. Nakanishi
520 and G. Soma. 2001. Cloning and characterisation of tandem-repeat type galectin in
521 rainbow trout (*Oncorhynchus mykiss*). **Fish Shellfish Immunol.** 11(3): 217-231.
- 522 Kirschner, K.N., A.B. Yongye, S.M. Tschampel, J. González-Outeiriño, C.R. Daniels, B.L.
523 Foley and R.J. Woods. 2008. GLYCAM06: A generalizable biomolecular force field.
524 Carbohydrates. **Journal of Computational Chemistry.** 29(4): 622-655.
- 525 Kong, H.J., W.J. Kim, H.S. Kim, Y.J. Lee, C.H. Kim, B.H. Nam, Y.O. Kim, D.G. Kim, S.J.
526 Lee, S.G. Lim and B.S. Kim. 2012. Molecular characterization of a tandem-repeat
527 galectin-9 (RuGlec9) from Korean rose bitterling (*Rhodeus uyekii*). **Fish Shellfish**

- 528 **Immunol.** 32(5): 939-944.
- 529 Kreutz, L.C., T.R. Pavan, A.G. Alves, A.G. Correia, B. Barriquel, E.D. dos Santos and L.J.
- 530 Barcellos. 2014. Increased immunoglobulin production in silver catfish (*Rhamdia*
- 531 quelen) exposed to agrichemicals. **Braz J Med Biol Res.** 47(6): 499-504.
- 532 Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton. 1993. PROCHECK: a
- 533 program to check the stereochemical quality of protein structures. **Journal of**
- 534 **Applied Crystallography.** 26(2): 283-291.
- 535 Liu, S., G. Hu, C. Sun and S. Zhang. 2013. Anti-viral activity of galectin-1 from flounder
- 536 *Paralichthys olivaceus*. **Fish Shellfish Immunol.** 34(6): 1463-1469.
- 537 Luthy, R., J.U. Bowie and D. Eisenberg. 1992. Assessment of protein models with three-
- 538 dimensional profiles. **Nature.** 356(6364): 83-85.
- 539 Mian, G.F., D.T. Godoy, C.A. Leal, T.Y. Yuhara, G.M. Costa and H.C. Figueiredo. 2009.
- 540 Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia.
- 541 **Vet Microbiol.** 136(1-2): 180-183.
- 542 Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell and A.J.
- 543 Olson. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective
- 544 receptor flexibility. **Journal of Computational Chemistry.** 30(16): 2785-2791.
- 545 Nakamura, O., Matsuoka, H., Ogawa, T., Muramoto, K., Kamiya, H., Watanabe T., 2006.
- 546 Opsonic effect of congerin, a mucosal galectin of the Japanese conger, *Conger*
- 547 *myriaster* (Brevoort). **Fish Shellfish Immunol.**, 20 :433–435
- 548 Norambuena, A., C. Metz, L. Vicuna, A. Silva, E. Pardo, C. Oyanadel, L. Massardo, A.
- 549 Gonzalez and A. Soza. 2009. Galectin-8 induces apoptosis in Jurkat T cells by
- 550 phosphatidic acid-mediated ERK1/2 activation supported by protein kinase A down-
- 551 regulation. **J Biol Chem.** 284(19): 12670-12679.
- 552 O'Farrell, C., N. Vaghefi, M. Cantonnet, B. Buteau, P. Boudinot and A. Benmansour. 2002.

- 553 Survey of transcript expression in rainbow trout leukocytes reveals a major
 554 contribution of interferon-responsive genes in the early response to a rhabdovirus
 555 infection. **J Virol.** 76(16): 8040-8049.
- 556 Poisa-Beiro, L., S. Dios, H. Ahmed, G.R. Vasta, A. Martinez-Lopez, A. Estepa, J. Alonso-
 557 Gutierrez, A. Figueras and B. Novoa. 2009. Nodavirus infection of sea bass
 558 (*Dicentrarchus labrax*) induces up-regulation of galectin-1 expression with potential
 559 anti-inflammatory activity. **J Immunol.** 183(10): 6600-6611.
- 560 Rabinovich, G.A. and M.A. Toscano. 2009. Turning 'sweet' on immunity: galectin-glycan
 561 interactions in immune tolerance and inflammation. **Nat Rev Immunol.** 9(5): 338-
 562 352.
- 563 Rajan, B., V. Kiron, J.M. Fernandes and M.F. Brinchmann. 2013. Localization and
 564 functional properties of two galectin-1 proteins in Atlantic cod (*Gadus morhua*)
 565 mucosal tissues. **Dev Comp Immunol.** 40(2): 83-93.
- 566 Ramachandran, S., P. Kota, F. Ding and N.V. Dokholyan. 2011. Automated minimization of
 567 steric clashes in protein structures. **Proteins: Structure, Function, and**
 568 **Bioinformatics.** 79(1): 261-270.
- 569 Schwede, T., J. Kopp, N. Guex and M.C. Peitsch. 2003. SWISS-MODEL: an automated
 570 protein homology-modeling server. **Nucleic Acids Research.** 31(13): 3381-3385.
- 571 Srikulnath, K., K. Matsubara, Y. Uno, A. Thongpan, S. Suputtitada, S. Apisitwanich, Y.
 572 Matsuda and C. Nishida. 2009. Karyological characterization of the butterfly lizard
 573 (*Leiolepis reevesii rubritaeniata*, Agamidae, Squamata) by molecular cytogenetic
 574 approach. **Cytogenet Genome Res.** 125(3): 213-223.
- 575 Tam, J.C. and D.A. Jacques. 2014. Intracellular immunity: finding the enemy within--how
 576 cells recognize and respond to intracellular pathogens. **J Leukoc Biol.** 96(2): 233-
 577 244.

- 578 Thurston, T.L., M.P. Wandel, N. von Muhlinen, A. Foeglein and F. Randow. 2012. Galectin
579 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion.
580 **Nature**. 482(7385): 414-418.
- 581 Tribulatti, M.V., V. Cattaneo, U. Hellman, J. Mucci and O. Campetella. 2009. Galectin-8
582 provides costimulatory and proliferative signals to T lymphocytes. **J Leukoc Biol**.
583 86(2): 371-380.
- 584 Tribulatti, M.V., J. Mucci, V. Cattaneo, F. Agüero, T. Gilmartin, S.R. Head and O.
585 Campetella. 2007. Galectin-8 induces apoptosis in the CD4(high)CD8(high)
586 thymocyte subpopulation. **Glycobiology**. 17(12): 1404-1412.
- 587 Tsai, C.M., C.H. Guan, H.W. Hsieh, T.L. Hsu, Z. Tu, K.J. Wu, C.H. Lin and K.I. Lin. 2011.
588 Galectin-1 and galectin-8 have redundant roles in promoting plasma cell formation.
589 **J Immunol**. 187(4): 1643-1652.
- 590 Wilkinson, H.W., L.G. Thacker and R.R. Facklam. 1973. Nonhemolytic group B
591 streptococci of human, bovine, and ichthyic origin. **Infect Immun**. 7(3): 496-498.
- 592 Yi, T., Y.-W. Li, L. Liu, X.-X. Xiao and A.-X. Li. 2014. Protection of Nile tilapia
593 (*Oreochromis niloticus* L.) against *Streptococcus agalactiae* following immunization
594 with recombinant FbsA and α -enolase. **Aquaculture**. 428–429(0): 35-40.
- 595 Yoshida, H., S. Yamashita, M. Teraoka, A. Itoh, S. Nakakita, N. Nishi and S. Kamitori.
596 2012. X-ray structure of a protease-resistant mutant form of human galectin-8 with
597 two carbohydrate recognition domains. **FEBS J**. 279(20): 3937-3951.
- 598

599 **Table and Figure legends**

600 **Table 1** Sequences of primers used in this study

601 **Table 2** Amino acid sequences used for phylogenetic, genome organization, and
602 genetic linkage analysis

603 **Table 3** Similarity and identity of *OnGal-8* with other organisms.

604

605 **Figure 1** The nucleotide and deduced amino acid sequences of galectin-8 are given in
606 the top and bottom lines, respectively. The number indicates the position of the nucleotides
607 and amino acids. The square boxes represent the putative site for Cys-disulfide bond
608 formation. The circle represents the putative site for phosphorylation. The triangle
609 represents the site for *O*-linked glycosylation. The polyadenylation signal sequence is
610 underlined. The asterisk represents the stop codon. The N- and C- terminus carbohydrate
611 recognition domains are shown on the left.

612

613 **Figure 2** Multiple sequence alignment of *OnGal-8* with other galectin-8s. The grey boxes
614 represent amino acids with similar properties and black boxes represent identical amino
615 acids. The amino acid on the alignment demonstrates the signature consensus amino acids
616 residues in the galectin-8 family.

617

618 **Figure 3** Comparison of pI and molecular weight properties of the N-terminus and C-
619 terminus carbohydrate recognition domains of *OnGal-8*.

620

621 **Figure 4** Phylogenetic tree of the amino sequences of galectin genes. The phylogeny
622 analysis was calculated using the neighbor-joining method (MEGA 6.0)

623

624 **Figure 5** Comparison of the genome organization of *OnGal-8* (Gal8) on the tilapia
 625 chromosome with galectin-8s from frog, chickens, and human,. The boxes indicate the
 626 exons. The number on the chromosome demonstrates the length of the nucleotide sequence
 627 on the genome.

628

629 **Figure 6** Genetic linkage of *OnGal-8* in tilapia (Gal8). The genes that surround Gal-8 are
 630 indicated.

631

632 **Figure 7** Putative 3D structure of *OnGal-8* predicted by homology modeling (A) and the
 633 predicted binding site between *OnGal-8* and galactose (Gal, B), lactose (Lac, C) and N-
 634 acetylgalactosamine (GalNAc, D). The grey-and red-wired structure represent the
 635 saccharide moieties. The yellow-red and yellow-blue wire structures represent the amino
 636 acid residues represented by a single-code amino acid and the number indicates the position
 637 of the amino acids. The green dashed line represents the hydrogen bond interactions
 638 between amino acid residues lining the active site of *OnGal-8* and the saccharide moiety.

639

640 **Figure 8** Expression analysis of *OnGal-8* by relative quantitative real-time PCR. (A)
 641 Tissue distribution analysis; head kidney, trunk kidney, spleen, liver, skin, gills, muscle,
 642 intestine, gonad, brain, heart, and PBLs of healthy Nile tilapia. (B) Relative quantitative
 643 real-time PCR of *OnGal-8* from the spleen of Nile tilapia experimentally challenged with *S.*
 644 *agalactiae* at 0 h, 6 h, 12 h, 1 d, 2 d, 3 d, 4 d, 5 d, and 6 d. The bars represent the standard
 645 error of three fish samples at each time point. The asterisks represent a significant difference
 646 ($P < 0.05$).

647

Table 1

Primer name	Sequence (5'→ 3')
Gal-8-RACE-R1	CTCACGGCGAGCAGCGTCACATC
Gal-8F	CCGAATTCATGTCGATTTCGAACCCG
Gal-8R	ATGCGGCCGCTTATTATTTGAAAACACTCGACA
Gal-8RTF	AGCTGGAGGGTCTTGACACC
Gal-8RTR	AGAGGTGGGGCTTGAGCTTG
β-actinF	CCAGCCTTCCTTCCTTGGTAT
β-actinR	GACGTCGCACTTCATGATGCT

651 **Table 2**

652

Galectin type	Name	Organisms	Accession number
Galectin-8	<i>OnGal-8</i>	<i>Oreochromis niloticus</i>	This study
	<i>IpGal-8</i>	<i>Ictalurus punctatus</i>	AHH40047.1
	<i>HsGal-8</i>	<i>Homo sapiens</i>	NP_963839.1
	<i>SsGal-8</i>	<i>Salmo salar</i>	NP_001133778.1
	<i>MmGal-8</i>	<i>Mus musculus</i>	NP_001277984.1
	<i>GgGal-8</i>	<i>Gallus gallus</i>	NP_001010843.1
	<i>ClGal-8</i>	<i>Columba livia</i>	EMC86069.1
	<i>XlGal-8</i>	<i>Xenopus laevis</i>	NP_001082394.1
	<i>PaGal-8</i>	<i>Pteropus alecto</i>	ELK05107.1
	<i>CgGal-8</i>	<i>Crassostrea gigas</i>	EKC40503.1
	<i>MmGal-8</i>	<i>Macaca mulatta</i>	NP_001253674.1
	<i>CmGal-8</i>	<i>Callorhinchus milii</i>	AFP08612.1
	<i>XrGal-8</i>	<i>Xenopus (Silurana) tropicalis</i>	NP_001135558.1
	<i>AmGal-8</i>	<i>Alligator mississippiensis</i>	XP_006258249.1
	<i>MfGal-8</i>	<i>Micrurus fulvius</i>	JAB54192.1
	<i>OhGal-8</i>	<i>Ophiophagus hannah</i>	ETE69412.1
Galectin-3	<i>RnGal-3</i>	<i>Rattus norvegicus</i>	NP_114020.1
	<i>MmGal-3</i>	<i>Mus musculus</i>	NP_001139425.1
	<i>GgGal-3</i>	<i>Gallus gallus</i>	NP_999756.1
	<i>SsGal-3</i>	<i>Salmo salar</i>	NP_001134305.1
	<i>IpGal-3</i>	<i>Ictalurus punctatus</i>	AHH38491.1
	<i>HsGal-3</i>	<i>Homo sapiens</i>	NP_002297.2
Galectin-1	<i>MmGal-1</i>	<i>Mus musculus</i>	NP_032521.1
	<i>DlGal-1</i>	<i>Dicentrarchus labrax</i>	ACF77003.1
	<i>HsGal-1</i>	<i>Homo sapiens</i>	NP_002296.1

Galectin type	Name	Organisms	Accession number
Galectin-2	<i>MmGal-2</i>	<i>Mus musculus</i>	NP_079898.2
	<i>SsGal-2</i>	<i>Salmo salar</i>	ACN12482.1
	<i>HsGal-2</i>	<i>Homo sapiens</i>	NP_006489.1
Galectin-7	<i>MmGal-7</i>	<i>Mus musculus</i>	NP_032522.2
	<i>HsGal-7</i>	<i>Homo sapiens</i>	NP_002298.1
galectin-4	<i>HsGal-4</i>	<i>Homo sapiens</i>	NP_006140.1
galectin-6	<i>HsGal-6</i>	<i>Homo sapiens</i>	AAI60275.1
galectin-9	<i>HsGal-9</i>	<i>Homo sapiens</i>	NP_001035167.2
galectin-10	<i>HsGal-10</i>	<i>Homo sapiens</i>	NP_001819.2
galectin-12	<i>HsGal-12</i>	<i>Homo sapiens</i>	NP_001136007.1
galectin-13	<i>HsGal-13</i>	<i>Homo sapiens</i>	NP_037400.1
galectin-14	<i>HsGal-14</i>	<i>Homo sapiens</i>	ACR09644.1
galectin-15	<i>OaGal-15</i>	<i>Ovis aries</i>	NP_001009238.1
galectin-16	<i>HsGal-16</i>	<i>Homo sapiens</i>	NP_001177370.2

653

654

FIGURE 1

658

FIGURE 2

Organisms	N-terminal		C-terminal	
	pI	MW	pI	MW
<i>Salmo salar</i>	7.00	14.447	6.65	15.194
<i>Gallus gallus</i>	7.91	14.889	8.54	15.213
<i>Rattus norvegicus</i>	9.55	14.961	6.33	15.084
<i>Homo sapiens</i>	8.96	14.636	7.10	15.139
<i>Bos taurus</i>	8.85	14.845	7.09	14.160
<i>Xenopus laevis</i>	8.85	14.845	7.09	15.160
<i>Nile tilapia</i>	7.87	14.387	9.22	15.042

659

660

661

TABLE 3

	1	2	3	4	5	6	7	8	9	10	11	12
IDENTITY												
OnGal-8		57.3	46.2	58.1	49.7	52.2	51.2	48.2	48.1	50.6	47.1	36.2
IpGal-8	74.8		47.1	69.3	47.9	51.9	50.8	48.9	50.2	51.9	48.1	36.3
HsGal-8	63.0	61.0		44.3	68.5	56.5	55.4	49.0	49.0	58.5	49.0	38.8
SsGal-8	75.7	82.1	59.6		48.9	52.4	51.7	48.6	51.1	49.5	52.2	39.9
MmGal-8	68.9	64.6	79.7	64.6		61.2	58.6	53.7	52.0	62.5	54.2	41.5
GgGal-8	71.7	73.0	69.9	72.1	77.5		82.8	60.8	62.7	79.1	61.0	43.5
ClGal-8	72.3	71.8	68.5	71.8	73.8	90.3		59.2	60.9	69.0	59.9	44.4
XlGal-8	69.5	69.5	66.0	68.6	74.2	77.5	75.5		85.1	58.1	54.1	42.0
XtGal-8	68.8	69.5	65.5	69.8	72.0	77.1	75.9	94.6		58.4	55.4	42.0
AmGal-8	69.5	73.2	69.4	70.3	76.3	89.2	82.1	74.6	75.2		61.1	45.4
MfGal-8	65.1	67.9	63.8	72.7	70.5	77.5	77.1	71.7	72.1	78.3		68.5
OhGal-8	57.9	57.4	54.3	61.8	59.7	63.5	63.3	61.6	61.6	65.5	79.8	
SIMILARITY												

662

663

FIGURE 3

GgGal-8 1 --MMSLDGPKKTSNPPIPYVGTILGGLVPGELIVLHGSVPDDADRFQVDLQCGSSSIKPR
 ClGal-8 1 --MMSLDGPKKTSNPPIPYVGTILGGLVPGELIVLHGAVEDDADRFQVDLQCGSSSIKPR
 XlGal-8 1 --MAQTGLCRTMDPVVPYVGTIFGGLPQGQIVIHGTVHPDADRFOIDFORGNVOPRS
 HsGal-8 1 MMSLDNNLQNIITNPIIPYVGTIPDQDLPCTLIVIRGHVPSDADRFOVDLQCGSSMKPR
 MmGal-8 1 --MMSLDNNLQNIITNPIIPYVGTITEQLKPGSLIVIRGHVPEKDSERFQVDLQCGSSSLKPR
 IpGal-8 1 --MSVKOPKQIVLKPEIPIAVTILGGLHPGEMVLIQGSVPADADRFOVDLQCGSSSKPR
 SsGal-8 1 --MSTINPKQTVLNPIIPACTILGGLQPGEMVLIQCTVPSDCERFQVDLQCGSSSKPR
 OnGal-8 1 --MSTINPROVFHNPSIIPAGTILGGLPGEVLIQGSVPSGADRFOVDLQCGSSVKPR

HFNPR VCN WGEI PF E V

GgGal-8 60 DVAFHFNPRFKSGCVVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 ClGal-8 60 DVAFHFNPRFRSGCIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 XlGal-8 59 DVAFHFNPRFKSGCIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 HsGal-8 61 DVAFHFNPRFKRAGCIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 MmGal-8 60 DVAFHFNPRFKRSGCIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 IpGal-8 59 DVAFHFNPRFKRD-HIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 SsGal-8 59 DVAFHFNPRFKRSGCIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK
 OnGal-8 59 DVAFHFNPRFKKS-CIVCNTLEKEKWGEEITYEMPFQKGRPFELIIMILKDKFQVSVNK

R

GgGal-8 120 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----SVQGAQPS-----SVGV
 ClGal-8 120 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 XlGal-8 119 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 HsGal-8 121 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 MmGal-8 120 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 IpGal-8 118 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 SsGal-8 119 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM
 OnGal-8 118 KHLLEYNHRISLERIDTLGIYKGVQIKSIEFVSN-----VTGTAFAHN-----PVFPTKM

HLNPR VRN

GgGal-8 166 TKTNTENGEMPDGLQFG-----V
 ClGal-8 170 TKTNTENGEMPDGLQFG-----V
 XlGal-8 166 TKTNTENGEMPDGLQFG-----V
 HsGal-8 168 TKTNTENGEMPDGLQFG-----V
 MmGal-8 176 TKTNTENGEMPDGLQFG-----V
 IpGal-8 163 TKTNTENGEMPDGLQFG-----V
 SsGal-8 158 TKTNTENGEMPDGLQFG-----V
 OnGal-8 172 TKTNTENGEMPDGLQFG-----V

HFNPR VCN

GgGal-8 184 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 ClGal-8 188 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 XlGal-8 184 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 HsGal-8 228 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 MmGal-8 194 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 IpGal-8 181 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 SsGal-8 165 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL
 OnGal-8 190 PVVGLVLSALHPGCTVAIKGEVKNPKSFTINLKSSDSKDIALHLNPRLNKKVFNRSYL

WGEI PF G V R

GgGal-8 244 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 ClGal-8 248 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 XlGal-8 244 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 HsGal-8 288 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 MmGal-8 254 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 IpGal-8 241 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 SsGal-8 225 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV
 OnGal-8 250 HDSWGEEERENFPFSPGMYFELIIFCDARQFKVAVNGVHTLEYKHRFKOLEKINILEV

GgGal-8 304 TGDVOLLVRSW
 ClGal-8 308 TGDVOLLVRSW
 XlGal-8 304 TGDVOLLVRSW
 HsGal-8 348 TGDVOLLVRSW
 MmGal-8 314 TGDVOLLVRSW
 IpGal-8 301 TGDVOLLVRSW
 SsGal-8 285 TGDVOLLVRSW
 OnGal-8 310 TGDVOLLVRSW

FIGURE 4

668

669

FIGURE 5

FIGURE 6

FIGURE 7

FIGURE 8

(A)

(B)