Identification and characterization of a novel legume-like lectin cDNA sequence from the red marine algae *Gracilaria fisheri*

SUKANYA SUTTISRISUNG[1], SAENGCHAN SENAPIN[1,2], BOONSIRM WITHYACHUMNARNKUL[1,3,4] and KANOKPAN WONGPRASERT[4,*]

[1]*Center of Excellence for Shrimp Molecular Biology and Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand*

[2]*National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani, Thailand*

[3]*Shrimp Genetic Improvement Center, National Center for Genetic Engineering and Biotechnology, Science and Technology Development Agency, Surat Thani, Thailand*

[4]*Department of Anatomy, Faculty of Science, Mahidol University, Bangkok, Thailand*

**Corresponding author (Email, sckbp@mahidol.ac.th)*

A legume-type lectin (L-lectin) gene of the red algae *Gracilaria fisheri* (GFL) was cloned by rapid amplification of cDNA ends (RACE). The full-length cDNA of GFL was 1714 bp and contained a 1542 bp open reading frame encoding 513 amino acids with a predicted molecular mass of 56.5 kDa. Analysis of the putative amino acid sequence with NCBI-BLAST revealed a high homology (30–68%) with legume-type lectins (L-lectin) from *Griffithsia japonica*, *Clavispora lusitaniae, Acyrthosiphon pisum*, *Tetraodon nigroviridis* and *Xenopus tropicalis*. Phylogenetic relationship analysis showed the highest sequence identity to a glycoprotein of the red algae *Griffithsia japonica* (68%) (GenBank number AAM93989). Conserved Domain Database analysis detected an N-terminal carbohydrate recognition domain (CRD), the characteristic of L-lectins, which contained two sugar binding sites and a metal binding site. The secondary structure prediction of GFL showed a β-sheet structure, connected with turn and coil. The most abundant structural element of GFL was the random coil, while the α-helixes were distributed at the N- and C-termini, and 21 β-sheets were distributed in the CRD. Computer analysis of three-dimensional structure showed a common feature of L-lectins of GFL, which included an overall globular shape that was composed of a β-sandwich of two anti-parallel β-sheets, monosaccharide binding sites, were on the top of the structure and in proximity with a metal binding site. Northern blot analysis using a DIG-labelled probe derived from a partial GFL sequence revealed a hybridization signal of ~1.7 kb consistent with the length of the full-length GFL cDNA identified by RACE. No detectable band was observed from control total RNA extracted from filamentous green algae.

[Suttisrisung S, Senapin S, Withyachumnarnkul B and Wongprasert K 2011 Identification and characterization of a novel legume-like lectin cDNA sequence from the red marine algae *Gracilaria fisheri*. *J. Biosci.* **36** 833–843] **DOI** 10.1007/s12038-011-9144-8

1. Introduction

Lectins are carbohydrate-binding proteins that are widely present in all living organisms (Sharon 2008). They have been reported to play important roles in defence mechanism by immobilizing particles through agglutination (Espinosa *et al.* 2009) and encapsulation (Koizumi *et al.* 1999), to limit pathogen infection and cancer metastasis through their binding to carbohydrates on cell surfaces (Lu *et al.* 1994). Since they possess the ability to agglutinate cells or

Keywords. cDNA cloning; 3D structure; *Gracilaria fisheri*; L-lectin; RACE-PCR

Abbreviations used: CRD, carbohydrate recognition domain; GFL, *Gracilaria fisheri* lectin; L-lectin, legume-type lectin

The GenBank accession number for the GFL sequence reported in this paper is ACY56710.

precipitate polysaccharides and glycoconjugates, they are used as tools for blood screening, cell identification and immunological research (Pinto *et al.* 2009). Many lectins are useful for viral inhibitors, such as *N*-acetylglucosamine-binding lectin extracted from the stinging nettle root of *Urtica dioica*, which has displayed pronounced antiviral properties (Shibuya *et al.* 1986). Also, mannose-binding lectins, such as cyanovirin-*N* derived from the blue-green algae *Nostoc ellipsosporum*, have shown high-affinity interactions with the envelope glycoprotein gp120 of HIV, enhancing its potential as an anti-HIV microbicide (Boyd *et al.* 1997).

Recent studies have shown that algae are a good source for novel lectins and have unique molecular structures, carbohydrate-binding specificities and biological activities, which relate to beneficial roles of algal lectins in biochemical and biomedical applications. Marine algal lectins are of interest because their small molecules can induce minor immunogenicity, and their several disulfide bridges possess great stability and also high specificity for complex carbohydrates and glyco-conjugates (Nagano *et al.* 2005b). The presence of lectins in marine macroalgae was first established by Boyd *et al.* (1966). Thereafter, a number of reports on distribution, isolation, characterization and biochemical properties of marine algae lectins have been published (Hori *et al.* 1981). It has been shown that lectins from marine algae exhibit protein properties similar to those of higher plants, but are different in some aspects. Algal lectins are generally of low molecular weight (about 10–30 kDa), and have a high content of acidic amino acids, giving isoelectric points ranging from 4 to 6 (Melo *et al.* 2004). They also possess thermostability and metal-independent hemagglutination, and have an affinity for glycoproteins but not for monosaccharides (Rogers and Hori 1993). These properties suggest that the algal lectins possess molecular structures and carbohydrate-binding specificities distinct from known lectins derived from other sources. Marine algal lectins have been reported both in monomeric and dimeric forms. Lectin isolated from *Gracilaria ornate* showed a monomeric protein with molecular weight of approximately 17 kDa (Leite *et al.* 2005) and lectin from *Caulerpa cupressoides* had a dimeric protein of 45 kDa, consisting of 23 kDa subunits (Benevides *et al.* 2001).

Red marine algae *Gracilaria* spp. (Gracilariaceae) are important sources of high-quality agar worldwide (Oliveira *et al.* 2000), and also are good sources of novel lectins. Several biomedical applications of lectin from *Gracilaria* spp. have been demonstrated. Hori *et al.* (1981) demonstrated that lower concentrations of lectin from *G. bursa-pastoris* showed stronger agglutination activity on mouse tumour cells FM3A than lectins from land plants. Lectins extract from *G. tikvahae* HBOI strains G-3 and G-5 have been shown to induce high mitogenic activity on human lymphocytes and murine splenocytes, respectively (Bird *et al.* 1993). Recently, the lectin Griffithsin, isolated from a red algae *Griffithsia* sp., showed a highly potent HIV entry inhibitory effect and is currently being investigated as a potential microbicide for use in the prevention of the transmission of HIV (Mori *et al.* 2005). Although several studies on marine algal lectins (an amino acid sequences obtained directly from purified protein) have been published, few genes encoding lectins from algae have been isolated and characterized. The first amino acid sequence of marine algal lectin published was obtained from the red algae *Bryothamnion triquetrum* in the year 2000 (Calvete *et al.* 2000). Others have been reported for *Griffithsia japonica* (GenBank number AAM93989), *Ulva pertusa* (Wang *et al.* 2004) and three species of *Hypnea*, i.e. *H. japonica* (Hori *et al.* 2000), *H. cervicornis* and *H. musciformis* (Nagano *et al.* 2005a). However, cDNA sequences of lectins from the family Gracilariaceae, including *Gracilaria fisheri*, have not been reported. Therefore, the present study was the first to use expressed sequence tags (ESTs) from a public database and a set of molecular biology tools to clone and sequence a full-length lectin cDNA of *G. fisheri*, and to use bioinformatics tools to analyse the structure.

2. Materials and methods

2.1 *Algal samples*

Vegetative thalli of the red algae *G. fisheri* and an unidentified species of filamentous green algae were collected from the culture ponds at the Genetic Shrimp Improvement Center, Surat Thani, Thailand. They were cleaned of the epiphytes, rinsed with distilled water, frozen in liquid nitrogen and stored at −80°C until used.

2.2 *RNA extraction*

Total RNA was extracted from healthy thalli of *G. fisheri* using TRIZOL reagent (RMC, USA), following the protocol as previously described (Dos Reis Falcão *et al.* 2008). Briefly, 1 g of frozen *G. fisheri* was powdered in liquid nitrogen using a mortar and pestle and then put into a 15 ml Falcon tube containing 3 ml TRIZOL reagent and 0.3 ml 5 M KOAc. The sample was centrifuged at 8000*g* at 4°C for 20 min. The supernatant was transferred to a new tube and 0.9 ml of cold ethanol, 1/3 volume of phenol:chloroform were added sequentially and then centrifuged at 8000*g* at 4°C for 20 min. RNA in the supernatant was precipitated by addition of LiCl to a final concentration of 3 M, and incubated at −20°C for 3 h. The pellet was washed twice with 70% ethanol and air-dried. RNA quality was assessed on 1% agarose gel electrophoresis. Total RNA was treated with RQ1 RNase-Free DNase (Promega, USA) to remove contaminating DNA. The

amount of RNA was quantitated using a Beckman Coulter DU 530 U (Beckman Coulter, USA).

2.3 *Cloning of the partial fragment Gracilaria fisheri L-lectin gene*

A partial sequence of GFL was obtained by RT-PCR using primers PGF1 and PGR1 (table 1), designed from available partial lectin sequences of *G. changii* (GenBank numbers DV963305 and DV962191). A 25 µl reaction solution contained 120 ng of RNA template, 100 µM of each primer, 0.5 µl of SuperScript One-Step RT/Platinum Taq mix (Invitrogen, USA), and 1× reaction buffer. The protocol comprised one initial step at 50°C for 30 min and 94°C for 2 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min and final extension at 72°C for 5 min. PCR products were analysed on 1.5% agarose gel electrophoresis. An amplified amplicon of 1208 bp was excised and purified using QIAquick PCR purification kit (QIAGEN, Germany), and subsequently cloned into a pDrive Cloning Vector (QIAGEN, Germany). Recombinant plasmids were sequenced by 1st BASE Ltd (Malaysia).

2.4 *Cloning of full-length GFL cDNA by rapid amplification of cDNA ends*

The 3′ and 5′ RACE methods were employed to amplify the 3′ and 5′ regions of the GFL, respectively, using a GeneRacer kit (Invitrogen, USA) and gene-specific primers, following the instruction manual. Briefly, for 5′ end amplification, total RNA of *G. fisheri* was treated with calf intestinal phosphatase to remove 5′ phosphates and then treated with tobacco acid pyrophosphates to remove the 5′ cap from full-length mRNA while leaving a 5′ phosphate required for ligation. The 5′ end of the treated mRNA was then ligated with the GeneRacer RNA oligomer by T4 RNA ligase, and then reverse-transcribed using the GeneRacer oligo (dT) primer and avian myeloblastosis virus reverse transcriptase. A touch-down PCR was performed as suggested in the instruction manual using the Platinum *Taq* DNA polymerase, GeneRacer 5′ primer G1 (specific to the GeneRacer RNA oligomer) and gene-specific primer PGR2 (table 1). A nested PCR was then carried out using the primers G2 (GeneRacer 5′ nested primer, table 1) and PGRN, following the PCR conditions in the instruction manual. The amplified PCR products were cloned into a plasmid vector pCR4-TOPO, which was transformed into competent One Shot TOP10 cells with a TOPO TA Cloning kit. The recombinant plasmids were extracted and analysed by DNA sequencing. For 3′ end amplification, RNA was reversed-transcribed using the GeneRacer oligo (dT) primer. The first round PCR was performed with the gene-specific primers PGF2 and G3 (GeneRacer 3′ primer, table 1). Nested PCR was then conducted with primers PGFN and G4 (GeneRacer 3′ nested primer, table 1). Cloning of amplified fragments was performed as mentioned above. Sequence information obtained from the above clones was used to design a pair of specific primers, FGF-F and FGF-R (table 1), to amplify the full-length of GFL cDNA in a RT-PCR reaction using similar conditions as above, except that 1.5 min was used in the extension step of 72°C. Amplified products were cloned and sent for sequencing.

2.5 *Sequence and structure analysis of GFL*

A homology search of the GFL cDNA sequence was submitted to BLAST (*http://www.ncbi.nlm.nih.gov/BLAST/*) and EMBL (European Molecular Biology Laboratory, *www.ebi.ac.uk/embl/*) databases. An analysis for protein domains was carried out using Conserved Domain Database software

Table 1. List of oligonucleotide primers

Name	Sequence from 5′ to 3′	Remarks
PGF1	GCGTCCCACAATGTCCGTGCTGTT	RT-PCR
PGR1	CGCGGATCAGCTGGCCTGTGATAT	RT-PCR
PGF2	ACCAATGCTGCGTCGTCTTCCAGTAGTG	3′RACE
PGFN	GCACACATCGCTTTCAGGGTCGTTGGAT	3′RACE Nested PCR, Probe
PGR2	GGAGGTGAGGCGCACGACGTCTCTAC	5′RACE
PGRN	TACGGTTGGCGAAAGGTGTGGTGCACAT	5′RACE Nested PCR
G1	GCACGAGGACACUGACAUGGACUGA	5′RACE (kit)
G2	GGACACTGACATGGACTGAAGGAGTA	5′RACE Nested PCR (kit)
G3	GCTGTCAACGATACGCTACGTAACG	3′RACE (kit)
G4	CGCTACGTAACGGCATGACAGTG	3′RACE Nested PCR (kit)
FGF-F	ACTAACACACCTCCTTCTCTGTCC	RT-PCR
FGF-R	GCAGTGGCTGAGAGCGTTTTGTTC	RT-PCR, Probe

and InterPro database (*http://www.ebi.ac.uk/InterProScan/*). Prediction of the secondary structure of GFL was analysed with SOPMA program (*http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html*) and a tertiary structure analysis was performed using Swiss-Model (*http://swiss-model.expasy.org*) with the crystal structure of mammalian VIP 36 chain A (PDB 2DUR) (Satoh *et al.* 2007) as a template. A phylogenetic tree of GFL sequence was analysed using the Molecular Evolutionary Genetics Analysis (MEGA 4) (Tamura *et al.* 2007).

2.6 *Northern blot analysis*

To confirm the expression of GFL in *G. fisheri*, northern blot analysis was undertaken. Labelled DNA probe was prepared using a commercial PCR DIG-labelling mix (Roche Molecular Biochemicals, USA), according to the manufacturer's instructions. Primers PGFN and FGF-R (table 1), expected to produce a 0.5 kb amplicon and recombinant plasmid containing a full-length GFL cDNA insert, were used in a labelling reaction. Total RNA samples extracted from *G. fisheri* and filamentous green algae (10 µg each) were denatured and separated in a 1.2% formaldehyde-denatured (w/v) agarose gel. The RNA was transferred to Nytran supercharge nylon membranes (Schleicher & Schull BioScience, USA). Hybridization was performed in Church's buffer (Church and Gilbert 1984) for 16 h at 55°C, then washed in 0.1× SSC with 0.1% SDS at 55°C. The hybridized probe-anti-DIG-AP complex was visualized on Kodak X-ray film after incubation of the membrane with CDP-Star chemiluminescent substrate (Roche, Germany).

3. Results

3.1 *cDNA cloning and sequence analysis of G. fisheri lectin*

A 1208 bp partial sequence (clone A, figure 1) of *G. fisheri* lectin (GFL) was initially obtained by RT-PCR, using PGF1 and PFR1 primers designed from the EST database of *G. changii* (accession number DV963305 and DV962191). Gene specific primers were subsequently designed from the known sequence for 3′ and 5′ RACE reactions. In a 5′ RACE assay, a 232 bp fragment (clone B) was generated, which contained 168 bp overlapping sequences with the initially isolated fragment. In a 3′ RACE, a 594 bp fragment (clone C) was obtained with a 152 bp overlapping the known sequence. The continuous GFL cDNA of 1680 bp (Clone D, figure 1) was amplified and confirmed that the combined sequences were correct. DNA sequence analysis revealed that the combined sequences of 1714 bp included 74 nucleotides of 5'-untranslated region (UTR), a 1542-nucleotide open reading frame, and a 92-nucleotide 3'-UTR including a stop codon (TAA) and poly (A) tail (figures 1 and 2). The sequence was deposited at GenBank under accession number ACY56710. The entire 1542 bp open reading frame encoded a putative protein of 513 deduced amino acids with a postulated molecular weight of 56.5 kDa, and an isoelectric point of 5.8. BLASTP analysis of the entire deduced protein sequence showed high homology to L-lectin or legume-like lectin proteins of many organisms, such as *Griffithsia japonica* (algae), *Clavispora lusitaniae* (fungi), *Acyrthosiphon pisum* (pea aphid), *Tetraodon nigroviridis* (pufferfish) and *Xenopus tropicalis* (Western clawed frog), with ranges of 30–68 % identity and 47–77% similarity. According to the rules of predict signal peptide (Nielsen *et al.* 1997), a putative signal peptide cleavage site between A19 and E20 was identified from the GFL sequence (figure 2). Conserved Domain Database (CDD) analysis detected an N-terminal carbohydrate recognition domain (CRD), characteristic of L-lectins, extending from residues 41 to residue 271. In addition, two sugar binding sites (D126 and G255), and a metal binding site (D157), were present in the putative CRD. A C-terminal transmembrane domain, predicted by InterPro, was located from amino acids 479 to 499 (figure 2).

3.2 *The predicted structure of GFL*

The overall schematic structure of GFL and a comparison of L- lectins from variety of organisms are displayed in

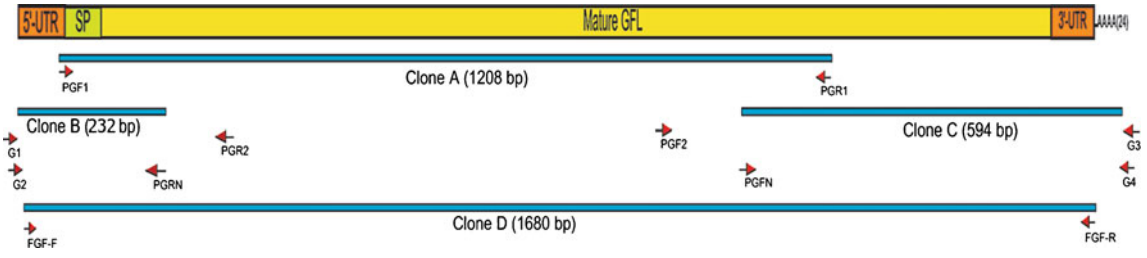

Figure 1. Schematic diagram of the cloning steps for GFL cDNA. Clones A–C represent partial GFL sequences, and clone D is a continuously amplified GFL cDNA. The arrows indicate primers. The entire ORF including poly (A) tail of *GFL* (1714 bp) is shown above the diagram. SP and UTR indicate the signal peptide and untranslated regions, respectively.

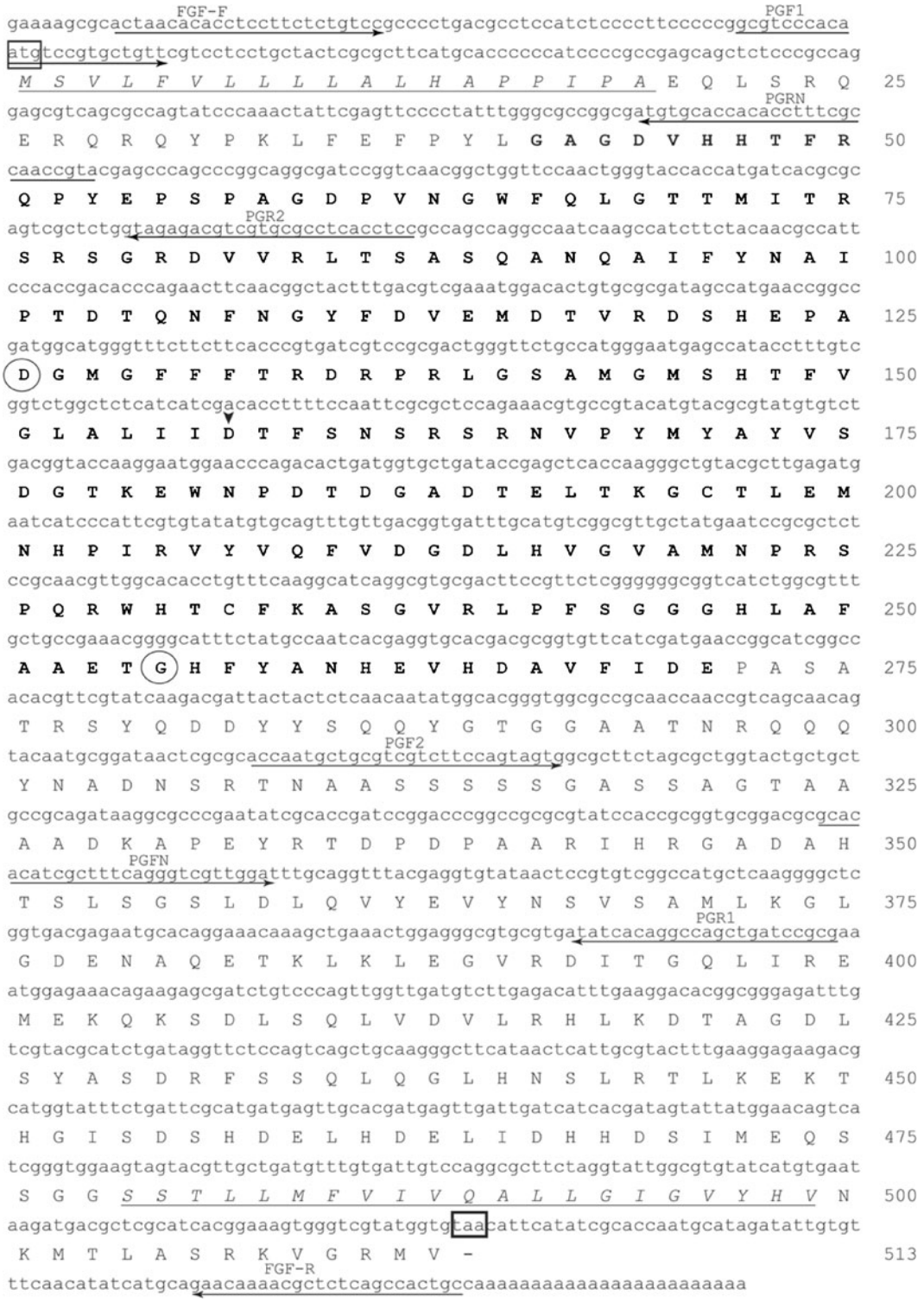

Figure 2. Nucleotide and deduced amino acid sequences of the gene encoding GFL from *Gracilaria fisheri*. Names of primers are indicated above the sequences, and 5′–3′ is represented with arrows. The letters in boxes indicate the start codon (ATG) and stop codon (TAA). The bold letters indicate a putative L-type lectin domain. Italics with underline represent putative N-terminal signal peptide sequence and C-terminal transmembrane region. The predicted sugar binding sites are marked by circles. The arrowhead indicates a metal binding site.

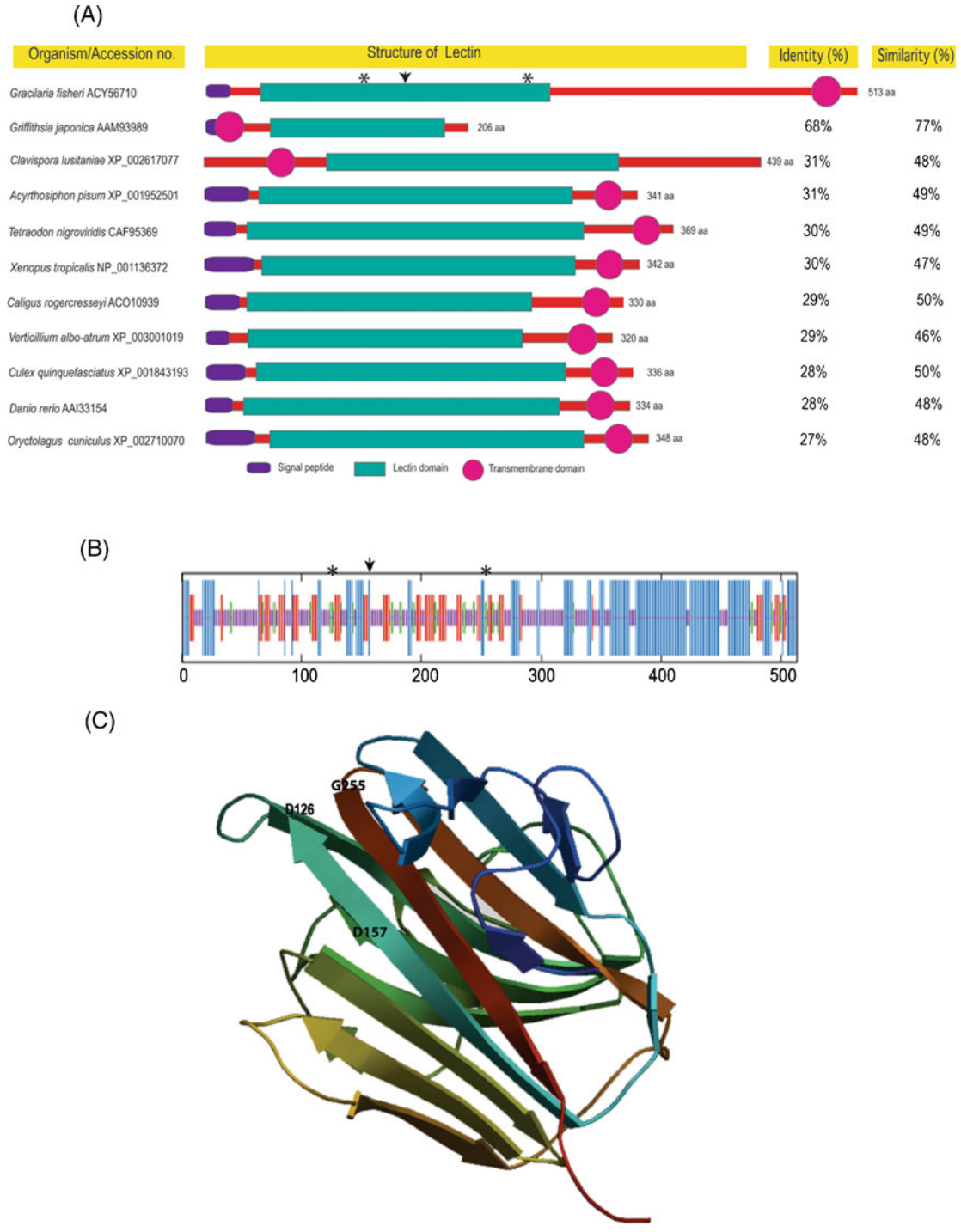

Figure 3. The predicted structures of GFL. (**A**) Comparison of overall structure of GFL with related L-lectins from other organisms. The red lines represent the putative amino acid sequences where the length is indicated on the right end. The purple and green boxes indicate putative signal peptide and CRD domain, respectively. Pink circles mark predicted transmembrane regions. Asterisks show predicted sugar binding sites. The percentage of identity and similarity of GFL to other L-type lectins derived from BLASTP analysis is indicated on the right margin. (**B**) The two-dimensional structure of the predicted GFL polypeptide. The helices, sheets, turns and coils are indicated as blue, red, green and purple vertical lines, respectively. Asterisks show predicted sugar binding sites. The arrow indicates a metal binding site. (**C**) The computational three-dimensional structure of GFL. β-sheets are indicated as patches. Turns and loops are indicated as lines. Amino acids constituting sugar binding sites are indicated as D126 and G255, and the metal binding site is indicated as D157.

figure 3A. It revealed that the predicted lectin domain and signal peptide region were found in all sequences. The deduced amino acid sequence of GFL showed overall 68% identity with the 23 kDa leg-like lectin of the red algae *Griffithsia japonica* (accession number AAM93989), 28–31% identities with the vesicular integral-membrane protein VIP36, *Verticillium albo-atrum* (accession number XP_003001019), and leg-like lectin *Clavispora lusitaniae* (accession number XP_002617077) from fungi. Sequence comparison of GFL with the animal lectins from the phyla Arthropoda and Vertebrata showed 30% identity with the lectin_VIP36_VIPL from *Tetraodon nigroviridis* (accession number CAF95369), 29% identity with vesicular integral-membrane protein VIP36 precursor from *Caligus rogercresseyi,* and 27–31% identities with the mannose-binding lectins from *Acyrthosiphon pisum* (accession number XP_001952501), *Culex quinquefasciatus* (accession number XP_001843193), *Xenopus tropicalis,* (accession number NP_001136372), *Danio rerio* (accession number AAI33154) and *Oryctolagus cuniculus,* (accession number XP_002710070) (figure 3A).

Analysis of the secondary structure of GFL using Hierarchical Neural Network (Combet *et al.* 2000) demonstrated β-sheets structure connected with turn and coil (figure 3B). The putative GFL peptide contained 36% of α-helix, 17% of β-sheet, 7% of β-turn and 40% of random coil. GFL comprised 21 β-sheets distributed in the CRD, whereas the random coil was abundant in the structural element, and the α-helixes were distributed in the N- and C-termini (figure 3B). Swiss-Model structure prediction showed a similar folding mode and spatial configuration of GFL to PDB 2DUR Chain A (Satoh *et al.* 2007) (figure 3C). The CRD of GFL had an overall globular shape composed of a β-sandwich of two anti-parallel β-sheets, and it is noteworthy that the β-sheets occurred predominantly in the structure of this lectin (figure 3C). The overall folding of GFL, which were typically built from β-sheets, connected by turns and loops, created a very tight structural scaffold. The carbohydrate binding sites were on the top of the structure and a metal binding site was on the upward surface of a β-strand, in close proximity with the carbohydrate binding sites (figure 3C).

3.3 *Phylogenetic tree of the GFL*

A phylogenetic tree was constructed by Molecular Evolutionary Genetics Analysis (MEGA 4) software based on the L-type lectins from Arthropoda, Vertebrata, Fungi and Rhodophyta (red algae). The result showed that GFL was closely related with L-lectin from the red algae *Griffithsia japonica* glycoprotein (accession number AAM93989) with 100% bootstrap supported (500 bootstrap replicas), as expected, but had far relationship with other L-type lectin which idenified as mannose binding lectins in Arthropoda and Vertebrata (figure 4).

3.4 *Transcriptional expression of GFL*

To determine if the putative ORF of GFL identified by sequence analysis is transcriptionally active, total RNA from *G. fisheri* was subjected to northern blot analysis using DIG-labelled probe derived from a partial GFL sequence, and the results are shown in figure 5. RNA extracted from filamentous green algae was used as a control. Equivalent loading and integrity of the RNA in each lane was confirmed by ethidium bromide staining of the gel (figure 5A). The northern hybridization revealed the presence of GFL transcripts from only the *G. fisheri* RNA sample (figure 5B). The size of the positive band estimated to be ~1.7 kb, which was consistent with the transcript size predicted from the GFL cDNA.

4. Discussion

The present study describes a full-length cDNA of a novel legume-like lectin cloned from the red algae *G. fisheri*, namely GFL by an EST search and RACE techniques. The full-length cDNA of GFL was 1714 bp and contained 1542 bp open reading frame encoding a putative protein of 513 deduced amino acids. A search from BLASTP database demonstrated that GFL had high homology (30–68%) with legume-type lectins (L-lectin) from *Griffithsia japonica*, *Clavispora lusitaniae, Acyrthosiphon pisum*, *Tetraodon. nigroviridis* and *Xenopus tropicalis*, indicating GFL was a member of the L-lectin family. A phylogenetic tree analysis showed a close evolutionary relationship (68% identity) of GFL with a glycoprotein of the red algae *Griffithsia japonica*. Moreover, they were not closely linked to any clusters of the other known L-lectins from other species. This suggested that the L-lectin genes from these two red algae might constitute a novel lectin. Although they are closely related, the primary structure of GFL was different from that of *G. japonica*, and that of other algal lectin sequences (Nagano *et al.* 2005a). Comparing the deduced amino acid sequences of other L-lectins, GFL showed a longer size (513 aa) with a long carboxyl terminal chain. This suggested a highly variable molecular evolution of L-lectin in different species and it is noteworthy that a lectin in a given algae does not necessarily imply a similar lectin present in closely related species.

Typically, most of the L-lectins consist of two or four subunits, each with a single, small carbohydrate combining site with the same specificity, and a metal binding site (for calcium and manganese). Each subunit has a size of 25–30 kDa and is commonly composed of a single polypeptide chain of 200–300 aa (Sharon and Lis 2002). In the present study, the full-length cDNA sequence of GFL encoded one

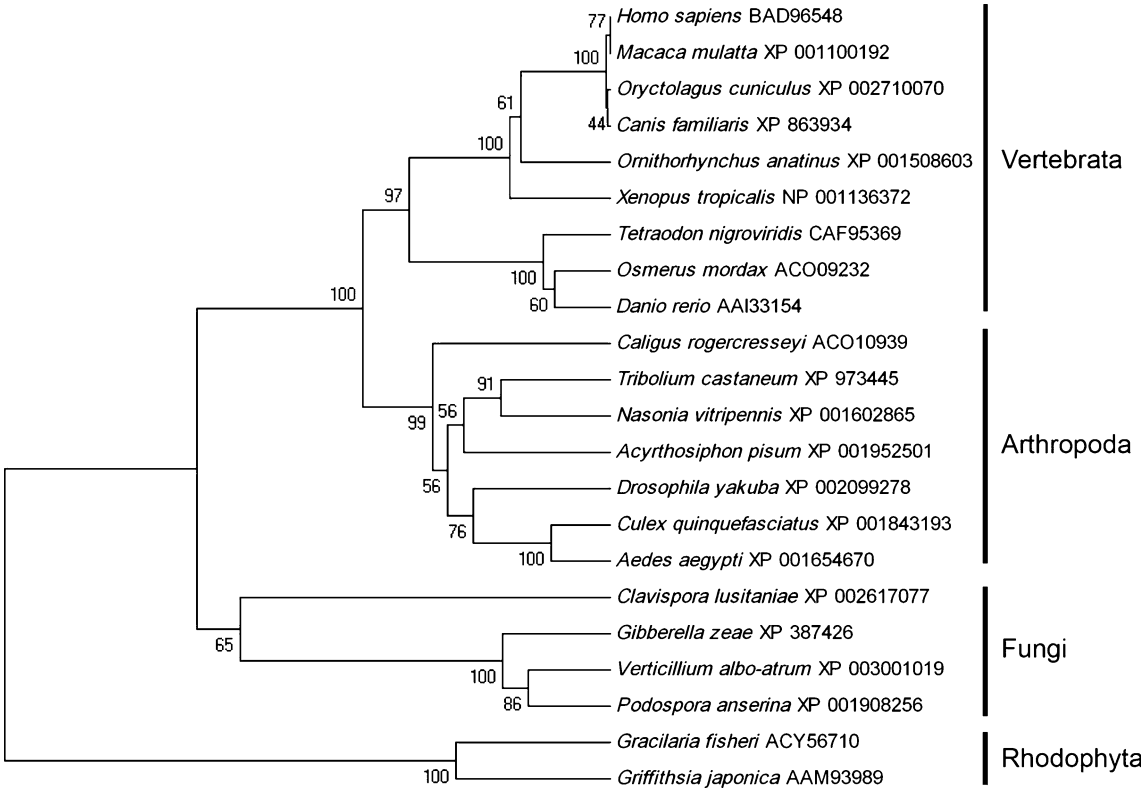

Figure 4. Phylogenetic analysis of L-lectin proteins. Accession numbers of the protein sequences from the GenBank database are indicated. Percentage bootstrap values (500 replicas) are given at the nodes of the tree.

subunit of 513 aa containing an N-terminal signal peptide of 19 aa and C-terminal transmenbrane domain of 21 aa. These lengths were similar to those of other reported L-type lectins. However, transmembrane domain of L-lectins from *G. japonica* and *Clavispora lusitaniae* is located at the N-terminal. Excluding the signal peptide (19 aa), GFL encoded a lectin protein of 494 aa with a molecular weight of 54.6 kDa and a pI of 5.7. In addition, many process sites such as glycosylation and phosphorylation were found in the C-terminal of GFL, suggesting possible sites of posttranslational processing. Indeed, L-lectins are generally synthesized as a precursor and undergo an initial removal of an N-terminal signal peptide. They are subsequently post-translationally processed, and that may consist of cleavage proteolysis, C-terminal trimming, sequence removal and re-ligation of the original C- and N-termini (Loris *et al.* 1998). Thus, the predicted molecular weight of mature protein of GFL may be smaller or larger than 54.6 kDa, depending on

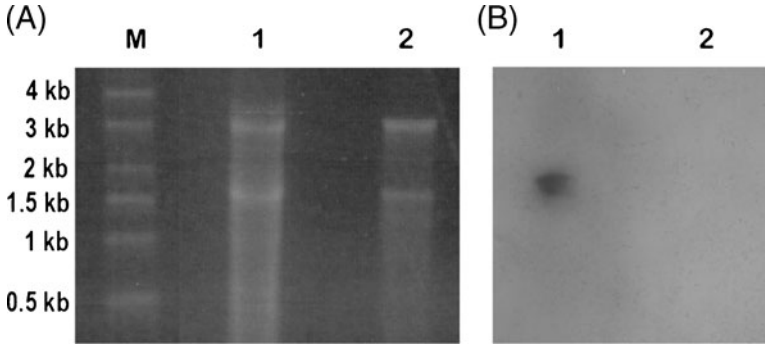

Figure 5. Northern blot analysis of the expression of GFL. (**A**) Ethidium bromide-stained denaturing agarose gel of total RNA (10 μg) from *Gracilaria fisheri* (lane 1), and a filamentous green alga (lane 2). (**B**) The blot hybridized with DIG-labelled probe from GFL cDNA (0.5 kb fragment) 35 M, Perfect RNA marker 0.2–10 kb.

the type of post-translational processing. X-ray crystallography and further analysis of natural or heterologously produced GFL by mass spectrometry or SDS-PAGE of a purified GFL can determine its exact MW and help to elucidate its molecular basis.

The prediction of carbohydrate binding from the structural analysis of GFL demonstrated a potential binding with monosaccharides as normally shown in L-lectins. L-lectins typically use a conserved set of amino acid residues, Asp and Gly (or Arg), as a binding site for monosaccharides, while variable parts of polypeptide are used to achieve specificity (Sharon and Lis 2002). The highly conserved key amino acids residues Asp (D126) and Gly (G255) were displayed in GFL and were assumed to be essential for carbohydrate binding (Sharma and Surolia 1997). However, the mannose binding site (QDNY) was not present, suggesting that GFL would not be a mannose binding lectin. Additionally, the GFL contained a metal binding site at Asp (D157), which was essential for correct folding and internal arrangements of the carbohydrate binding site (Loris *et al.* 1998). The metal binding site is also crucial for its biological activity as shown in all L-lectins that divalent cations (Mn^{2+} and Ca^{2+}) are required for hemagglutination activity (Sharon and Lis 2002). GFL also contained the characteristic site of Ala (A125) preceding to Asp (D126) suggesting the location of a *cis*-peptide bond between Ala-Asp, which is known to be stabilized by metal ions (Sharon and Lis 2002). This arrangement is a key event in the metal-induced activation of GFL. It is important to note that the above findings of GFL were in contrast to those of other marine red algae lectins. It has been shown that most red algal lectins have low molecular weight, bind to large molecules such as glycoprotein or yeast mannan bearing high-mannose N-glycans, and they do not require divalent cations for hemagglutination (Rogers and Hori 1993). Nevertheless, these characteristics are extended to a high-molecular-mass lectin (57 kDa) from *Gracilaria Cornea* (Lima *et al*. 2005). However, it should be mentioned that some high-molecular-weight lectins (MW>60000) from the red algae *Ptilota serrata* (Rogers *et al.* 1990) and *P. filicina* (Sampaio *et al*. 1998) are able to bind to monosaccharides such as D-galactose, D-fucose, *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc) and also require metals for their biological activity. These data indicate that red algae have a diverse family of lectins, and each has a distinct structure and role. Computer analysis of the secondary and tertiary structures also showed that GFL had the secondary and three-dimensional structures similarly to other L-lectins (Satoh *et al.* 2006). These included the carbohydrate binding sites which located at the top of the protein structure, and was in the close proximity with the metal binding site.

Over 100 L-lectins have been isolated from seeds and others from shoot, root, bark and nodules of land plants (Van Damme *et al.* 1998). They have been reported to have diverse function, such as creating a symbiosis with soil bacteria, of the genus *Rhizobium*, fixing atmospheric nitrogen, rendering plants independent of supply of external nitrogen fertilizer (Fountain *et al.* 1977), and defense against pathogens (biotic stress) by way of lectin–pathogen interaction via recognizing and immobilizing the pathogens, to prevent subsequent growth and multiplication (Etzler 1986). This kind of defensive role of lectins has not been clearly demonstrated in marine algae. However, they have been reported to have a role in recognition and adhesion of sexual gametes during reproduction (Kim and Kim 1999). Previous studies have also shown biological activities of lectins from marine algae in aggregation of fungi, bacteria, blue-green algae, diatoms, dinoflagellates and erythrocytes (Hori *et al.* 1996). Additionally, lectin-like proteins from the red algae *Eucheuma serra* and *Pterocladia capillacea* have been shown to inhibit the growth of the marine bacterium *Vibrio vulnificus*, and it was suggested that the lectin mediated bacterial–algal recognition and binding via the specific carbohydrates on either the algal cell walls or plasma membranes of the bacteria, providing defence for the algae (Liao *et al.* 2003). However, in our study, the function of GFL in defence of the algae, or against biotic stress, has not been elucidated. From northern blot analyses, it was demonstrated that GFL was expressed in vegetative thalli of *G. fisheri*, indicating that it is a constitutively expressed protein (Yao *et al.* 2003). *G. fisheri*, in the present study, was collected from the culture ponds where the seaweed do not develop the reproductive stages, and most of the algae were vegetative thalli throughout the year. Thus, we could not determine the change of GFL expression in different stages of growth of the seaweed. It has been reported that plant lectin genes express differentially under various abiotic stresses such as temperature shock, drought and high salinity stresses (Jiang *et al*. 2010). In this study, *G. fisheri* was cultured in the ponds with controlled conditions of nutrition and salinity; however, there was a seasonal variation of temperature and irradiance. We found that vegetative thalli of *G. fisheri* collected from the culture ponds did not express GFL differently from season to season (data not shown), and that might be due to the temperature change was not large enough to alter the GFL expression. Hence, further study of the differential expression of GFL under biotic/abiotic conditions is of interest to gain more knowledge about GFL's functions.

In conclusion, the present study is the first identification of an L-lectin cDNA sequence from the red marine algae *G. fisheri*. L-lectins constitute a large family of homologous proteins that have diversity and specificity of protein–carbohydrate interactions (Sharon and Lis 1995). Therefore,

they are widely studied as useful tools in immunology and applications for the understanding of cell–cell recognition, adhesion, tumour spread, bacterial and viral infection, and inflammation (Rudiger *et al.* 2000). Hence, expression and purification of the GFL protein, and functional analyses will further elucidate its structure–function relationship. Understanding the carbohydrate recognition of GFL from *G. fisheri* and its biological activities could possibly draw a considerable economic attention to exploit novel and useful lectins from *G. fisheri,* for many applications, particularly as an anti-pathogen agent.

Acknowledgements

This work was supported by Mahidol University Postdoctoral Fellowship Program, the Thailand Research Fund and the Commission on Higher Education. We would like to thank the Shrimp Genetic Improvement Center, the National Center for Genetic Engineering and Biotechnology, the Ministry of Science and Technology at Surat Thani for algae samples, and Professor Peter Hanna for his critical comments on the manuscript.

References

Benevides NMB, Holanda ML, Melo FR, Pereira MG, Monteiro ACO and Freitas ALP 2001 Purification and partial characterization of the lectin from the marine green alga *Caulerpa cupressoides* (Vahl) C. Agardh. *Bot. Mar*. **44** 17–22

Bird K, Chiles T, Longley R, Kendrick A and Kinkema M 1993 Agglutinins from marine macroalgae of the southeastern United States. *J. Appl. Phycol.* **5** 213–218

Boyd W, Almodóvar L and Boyd L 1966 Agglutinins in marine algae for human erythrocytes. *Transfusion* **6** 82–83

Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, *et al.* 1997 Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob. Agents Chemother*. **41** 1521–1530

Calvete JJ, Costa FH, Saker-Sampaio S, Murciano MP, Nagano CS, Cavada BS, *et al.* 2000 The amino acid sequence of the agglutinin isolated from the red marine alga *Bryothamnion triquetrum* defines a novel lectin structure. *Cell Mol. Life Sci.* **57** 343–350

Church G and Gilbert W 1984 Genomic sequencing. *Proc. Natl. Acad. Sci. USA*. **81** 1991–1995

Combet C, Blanchet C, Geourjon C and Deleage G 2000 NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25** 147–150

Dos Reis Falcão V, Pedroso Tonon A, Cabral Oliveira M and Colepicolo P 2008 RNAIsolation method for polysaccharide rich algae: agar producing *Gracilaria tenuistipitata* (Rhodophyta). *J. Appl. Phycol.* **20** 9–12

Espinosa EP, Perrigault M, Ward JE, Shumway SE and Allam B 2009 Lectins associated with the feeding organs of the oyster *Crassostrea virginica* can mediate particle selection. *Biol. Bull.* **217** 130–141

Etzler ME 1986 Distribution and function of plant lectins; in *The lectin: properties, functions and applications in biology and medicine* (eds) IE Liener, Goldstein and NSharon (New York: Academic Press) pp 371–425

Fountain DW, Foard DE, Replogle WD and Yang WK 1977 Lectin release by soybean seeds. *Science* **197** 1185–1187

Hori K, Miyazawa K and Ito K 1981 Hemagglutinins in marine algae. *Bull. Jpn. Soc. Sci. Fish* **47** 793–798

Hori K, Ogata T, Kamiya H, Mimuro M 1996 Lectin-like compounds and lectin receptors in marine microalgae: hemagglutination and reactivity with purified lectins. *J. phycol.* **32** 783–790

Hori K, Matsubara K and Miyazawa K 2000 Primary structures of two hemagglutinins from the marine red alga, *Hypnea japonica*. *Biochim. Biophys. Acta*. **1474** 226–236

Jiang SY, Ma Z and Ramachandran S 2010 Evolutionary history and stress regulation of the lectin superfamily in higher plants. *BMCEvol. Biol.* **10** 79

Kim S-H and Kim GH 1999 Cell–cell recognition during fertilization in the red alga, *Aglaothamnion oosumiense* (Ceramiaceae, Rhodophyta). *Hydrobiologia* **398–399** 81–89

Koizumi N, Imamura M, Kadotani T, Yaoi K, Iwahana H and Sato R 1999 The lipopolysaccharide-binding protein participating in hemocyte nodule formation in the silkworm *Bombyx mori* is a novel member of the C-type lectin superfamily with two different tandem carbohydrate-recognition domains. *FEBS Lett.* **443** 139–143

Leite YF, Silva LM, Amorim RC, Freire EA, de Melo Jorge DM, Grangeiro TB and Benevides NM 2005 Purification of a lectin from the marine red alga *Gracilaria ornata* and its effect on the development of the cowpea weevil *Callosobruchus maculatus* (Coleoptera: Bruchidae). *Biochim. Biophys. Acta.* **1724** 137–145

Liao WR, Lin JY, Shieh WY, Jeng WL and Huang R 2003 Antibiotic activity of lectins from marine algae against marine vibrios. *J. Ind. Microbiol. Biotechnol.* **30** 433–439

Lima ME, Carneiro ME, Nascimento AE, Grangeiro TB, Holanda ML, Amorim RC and Benevides NM 2005 Purification of a lectin from the marine red alga *Gracilaria cornea* and its effects on the cattle tick *Boophilus microplus* (Acari: Ixodidae). *J. Agric. Food Chem.* **53** 6414–6419

Loris R, Hamelryck T, Bouckaert J and Wyns L 1998 Legume lectin structure. *Biochim. Biophys. Acta.* **1383** 9–36

Lu Y, Pelling JC and Chaney WG 1994 Tumor cell surface beta 1–6 branched oligosaccharides and lung metastasis. *Clin. Exp. Metastasis* **12** 47–54

Melo FR, Benevides NMB, Pereira MG, Holanda ML, Mendes FNP, Oliveira SRM, Freitas ALP and Silva LMCM 2004 Purification and partial characterisation of a lectin from the red marine alga *Vidalia obtusiloba* C. Agardh. *Revista. Brasil. Bot.* **27** 263–269

Mori T, O'Keefe BR, Sowder RC 2nd, Bringans S, Gardella R, Berg S, Cochran P, *et al.* 2005 Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga *Griffithsia* sp. *J. Biol. Chem.* **280** 9345–9353

Nagano CS, Debray H, Nascimento KS, Pinto VP, Cavada BS, Saker-Sampaio S, Farias WR, Sampaio AH and Calvete JJ

2005a HCA and HML isolated from the red marine algae *Hypnea cervicornis* and *Hypnea musciformis* define a novel lectin family. *Protein Sci.* **14** 2167–2176

Nagano CS, Gallego del Sol F, Cavada BS, Nascimento KS, Nunes EV, Sampaio AH and Calvete JJ 2005b Crystallization and preliminary X-ray diffraction analysis of HML, a lectin from the red marine alga *Hypnea musciformis*. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **61** 997–999

Nielsen H, Engelbrecht J, Brunak S and von Heijne G 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10** 1–6

Oliveira EC, Alveal K and Anderson R 2000 Mariculture of the agar-producing Gracilarioid red algae. *Res. Fish. Sci.* **8** 345–378

Pinto VP, Debray H, Dus D, Teixeira EH, de Oliveira TM, Carneiro VA, Teixeira AH, Filho GC, *et al.* 2009 Lectins from the red marine algal species *Bryothamnion seaforthii* and *Bryothamnion triquetrum* as tools to differentiate human colon carcinoma cells. *Adv. Pharmacol. Sci.* doi: 10.1155/2009/862162

Rogers DJ and Hori K 1993 Marine algal lectins: new developments. *Hydrobiologia.* **260–261** 589–593

Rogers DJ, Fish B and Barwell CJ 1990 Isolation and properties of lectins from two red marine algae: *Plumaria elegans* and *Ptilota serrata*; in *Lectins: biology, biochemistry, clinical biochemistry* (eds) TC Bog-Hansen and DL JFreed (St Louis: Sigma Chemical Company) pp 49–52

Rudiger H, Siebert HC, Solis D, Jimenez-Barbero J, Romero A, von der Lieth CW, Diaz-Marino T and Gabius HJ 2000 Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets. *Curr. Med. Chem.* **7** 389–416

Sampaio AH, Rogers DJ and Barwell CJ 1998 A galactose-specific lectin from the red marine alga *Ptilota filicina*. *Phytochemistry.* **48** 765–769

Satoh T, Sato K, Kanoh A, Yamashita K, Yamada Y, Igarashi N, Kato R, Nakano A and Wakatsuki S 2006 Structures of the carbohydrate recognition domain of Ca^{2+} -independent cargo receptors Emp46p and Emp47p. *J. Biol. Chem.* **281** 10410–10419

Satoh T, Cowieson NP, Hakamata W, Ideo H, Fukushima K, Kurihara M, Kato R, Yamashita K and Wakatsuki S 2007 Structural basis for recognition of high mannose type glycoproteins by mammalian transport lectin VIP36. *J. Biol. Chem.* **282** 28246–28255

Sharma V and Surolia A 1997 Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J. Mol. Biol.* **267** 433–445

Sharon N 2008 Lectins: past, present and future. *Biochem. Soc. Trans.* **36** 1457–1460

Sharon N and Lis H 1995 Lectins proteins with a sweet tooth: functions in cell recognition. *Essays Biochem.* **30** 59–75

Sharon N and Lis H 2002 How proteins bind carbohydrates: lessons from legume lectins. *J. Agric. Food Chem.* **50** 6586–6591

Shibuya N, Goldstein IJ, Shafer JA, Peumans WJ and Broekaert WF 1986 Carbohydrate binding properties of the stinging nettle (*Urtica dioica*) rhizome lectin. *Arch. Biochem. Biophys.* **249** 215–224

Tamura K, Dudley J, Nei M and Kumar S 2007 MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24** 1596–1599

Van Damme EJM, Peumans J, Pusztai A and Bardocz S 1998 Handbook of plant lectins: Properties and biomedical applications (Chichester: Wiley) pp 452

Wang S, Zhong FD, Zhang YJ, Wu ZJ, Lin QY and Xie LH 2004 Molecular characterization of a new lectin from the marine alga *Ulva pertusa*. *Acta Biochim. Biophys. Sin.* **36** 111–117

Yao JH, Zhao XY, Liao ZH, Lin J, Chen ZH, Chen F, Song J, Sun XF and Tang KX 2003 Cloning and molecular characterization of a novel lectin gene from *Pinellia ternata*. *Cell Res.* **13** 301–308

MS received 21 March 2011; accepted 22 July 2011

*e*Publication: 30 September 2011

Corresponding editor: B JAGADEESHWAR RAO