<table>
<thead>
<tr>
<th>Title</th>
<th>cDNA cloning and characterization of a rhamnose-binding lectin SUL-I from the toxopneustid sea urchin Toxopneustes pileolus venom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Hatakeyama, Tomomitsu; Ichise, Ayaka; Yonekura, Tomokazu; Unno, Hideaki; Goda, Shuichiro; Nakagawa, Hideyuki</td>
</tr>
<tr>
<td>Citation</td>
<td>Toxicon, 94, pp.8-15; 2015</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2015-02</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10069/35011">http://hdl.handle.net/10069/35011</a></td>
</tr>
</tbody>
</table>

© 2015 Elsevier Ltd.; NOTICE: this is the author's version of a work that was accepted for publication in Toxicon. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Toxicon, 94, (2015)
cDNA cloning and characterization of a rhamnose-binding lectin SUL-I from the toxopneustid sea urchin *Toxopneustes pileolus* venom

*Tomomitsu Hatakeyama*¹, *Ayaka Ichise*¹, *Tomokazu Yonekura*¹, *Hideaki Unno*¹, *Shuichiro Goda*¹, and *Hideyuki Nakagawa*²

1Biomolecular Chemistry Laboratory, Graduate School of Engineering, Nagasaki University, Bunkyo-machi 1-14, Nagasaki 852-8521, Japan, ²Division of Environmental Symbiosis, Graduate School of Integrated Arts and Sciences, Tokushima University, Tokushima 770-8502, Japan

Running title: L-rhamnose-binding lectin from the toxopneustid sea urchin

Address correspondence to: Tomomitsu Hatakeyama, Biomolecular Chemistry Laboratory, Graduate School of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan. Tel: +81-95-819-2686; Fax: +81-95-819-2684; E-mail: thata@nagasaki-u.ac.jp

Abbreviations: CRD, carbohydrate-recognition domain; SUEL, sea urchin egg lectin; SUL-I, sea urchin lectin-I; PD, polyamidoamine dendrimer; RACE, rapid amplification of cDNA ends; RBL, rhamnose-binding lectin; TBS, Tris-buffered saline
Abstract

The globiferous pedicellariae of the venomous sea urchin *Toxopneustes pileolus* contain several biologically active proteins. Among these, a galactose-binding lectin SUL-I isolated from the venom in the large globiferous pedicellariae shows several activities such as mitogenic, chemotactic, and cytotoxic activities through binding to the carbohydrate chains on the cells. We cloned cDNA encoding SUL-I by reverse transcription-PCR using the degenerate primers designed on the basis of the N-terminal amino acid sequence of the protein and expressed the recombinant SUL-I (rSUL-I) in *Escherichia coli* cells. The SUL-I gene contains an open reading frame of 927 nucleotides corresponding to 308 amino acid residues, including 24 residues of a putative signal sequence. The mature protein with 284 residues is composed of three homologous regions, each showing similarity with the carbohydrate-recognition domains of the rhamnose-binding lectins, which have been mostly found in fish eggs. While rSUL-I exhibited binding activity for several galactose-related sugars, the highest affinity was found for L-rhamnose among carbohydrates tested, confirming that SUL-I is a rhamnose-binding lectin. rSUL-I also showed hemagglutinating activity toward rabbit erythrocytes, indicating the existence of more than one carbohydrate-binding site to cross-link the carbohydrate chains on the cell surface, which may be closely related to its biological activities.

Keywords: sea urchin, *Toxopneustes pileolus*, cDNA cloning, rhamnose, lectin, carbohydrate-binding

1. Introduction

The venom obtained from the globiferous pedicellariae of the sea urchin *Toxopneustes pileolus* contains several biologically active proteins (Kimura et al. 1975; Nakagawa et al. 2003; Nakagawa et al. 1991). Among these, the galactose-specific lectin SUL-I isolated from the venom of the large globiferous pedicellariae shows various activities such as chemotactic activity on guinea pig neutrophils and mitogenic activity on murine splenocytes by binding to the carbohydrate chains on
target cells (Nakagawa et al. 1996; Takei and Nakagawa 2006). N-terminal sequence analysis of
SUL-I suggested that this lectin has some similarity with rhamnose-binding lectins (RBLs), majority
of which have been isolated from fish eggs (Tateno 2010). RBLs are also referred to as the sea urchin
egg lectin (SUEL ) family because of their homology with those found in the eggs of the sea urchin
Anthocidaris crassispina. Till date, the lectins having homology with SUEL have mostly been
found in fish eggs, while some homologous proteins have also been found in mammals, e.g., mouse
latrophilin-1, a putative G-protein-coupled receptor involved in synaptic function (Vakonakis et al.
2008).

In various organisms, lectins are known to play important roles in molecular and cellular
recognition processes because of the vast diversity of the carbohydrate chain structures on their
surface. Lectins are categorized into several families (Kilpatrick 2002). Among these, two major
groups are C-type lectins and galectins (S-type lectins) (Drickamer 1988). C-type lectins were named
owing to their Ca^{2+}-dependent carbohydrate-binding activity, and these lectins contain common
carbohydrate-recognition domains (CRDs) composed of 110–130 amino acid residues. They are
distributed in various organisms and are known to play important roles in various biological
molecular recognition systems, including the immune system and cell adhesion processes
(Drickamer 1999). Some C-type lectins and C-type lectin-like proteins (Zelensky and Gready 2005)
have also been found abundantly in snake venoms (Igci and Demiralp 2011). They contribute to the
toxicity of the venom by binding to the carbohydrate chains on target cells of the victims. Therefore,
it may be important to elucidate the structure and function of the lectins in animal venoms to
understand their implications in toxicity. However, there is very limited knowledge concerning the
lectins in animal venoms, particularly those from marine organisms.

In the present study, we have cloned SUL-I cDNA from the large globiferous pedicellariae of T.
pileolus and expressed it in Escherichia coli cells to elucidate its structure and carbohydrate-binding
properties. The results reveal its structural relationship with RBLs. The putative three-dimensional
structure constructed by homology modeling using SUEL domain provides insights into its carbohydrate-recognition mechanism.

2. Materials and Methods

2.1. Materials

Oligonucleotides and polyamidoamine dendrimers (PDs) with 64 amino surface groups (ethylenediamine core, generation 4.0, M.W. 14,214) (amino-PD) were purchased from Sigma-Aldrich. The plasmid vectors used in this study were as follows: pTAC2 vector was obtained from DynaExpress and pET-3a expression vector was obtained from Novagen. Melibiose, lactose, and mannose were obtained from Wako Pure Chemicals (Osaka, Japan). Rabbit blood was obtained from Nippon Bio-Test Laboratories (Tokyo, Japan). The lactose-immobilized Cellulofine (lactose-Celluloine) column was prepared by attaching lactose to Cellulofine gels (JNC Corp., Tokyo, Japan) using the cross-linking reagent divinyl sulfone (Sigma-Aldrich), as described in previously (Hatakeyama et al. 1994). All other chemicals were of analytical grade for biochemical use.

2.2. cDNA cloning of SUL-I

The globiferous pedicellariae of T. pileolus were flash frozen in liquid nitrogen and ground to a powder form. Total RNA was extracted using Isogen solution (Nippon Gene, Tokyo, Japan). Poly(A) RNA was collected using the Oligotex-dT30 mRNA Purification Kit (Takara, Otsu, Japan), and cDNA was synthesized using the SMARTer cDNA Cloning Kit (Clontech). A DNA fragment corresponding to the N-terminal region of SUL-I cDNA was amplified by polymerase chain reaction (PCR) using two degenerate primers, DF1: 5'-GTIGGIMGIACITGYGARGGIAA-3' and DR1: 5'-CCIGGISWRTTICKICCRTARTT-3'. This DNA fragment was cloned into pTAC-2 vector using E. coli JM109 cells (Clontech) and sequenced using ABI PRISM 3130 Genetic Analyzer (Applied Biosystems). The amino acid sequence deduced from this DNA fragment was in accordance with
that determined from the purified protein. Therefore, 3′- and 5′-rapid amplification of cDNA ends
(3′-RACE and 5′-RACE) were then performed using the primer (IF1) designed from this region
using the SMARTer cDNA Cloning Kit. Primers used for the degenerate PCR, 3′-RACE, and
5′-RACE are indicated by arrows in Fig. 1. The amino acid sequence of SUL-I was compared with
the Uniprot database (www.uniprot.org) by BLAST search (Altschul et al. 1990). Multiple
alignments of the sequences were performed using Clustal Omega (Sievers et al. 2011). The
determined nucleotide sequence was deposited in DDBJ/EMBL/GenBank (accession number:
LC003233). Chemical and physical parameters for SUL-I were calculated from the deduced
sequence using the ProtPram tool in ExPASy Bioinformatics Resource Portal (www.expasy.org)

2.3. Expression and purification of the recombinant SUL-I in E. coli cells
The coding region of the mature SUL-I protein with the initiator methionine residue was
amplified by PCR using two primers (forward:
5′-AAGGAGATATACATATGGCTGTGGGAAGAACTTGTGA-3′ and reverse:
5′-GTTAGCAGCCGATCCATCAGCTGGATTTCCCAGCCAT-3′) and inserted into a pET-3a vector
at NdeI and BamHI restriction sites using the In-Fusion HD Cloning Kit (Clontech). The plasmid was
amplified in E. coli JM109 cells, and the protein was expressed in E. coli BL21(DE3)pLysS cells
(Novagen). The recombinant SUL-I (rSUL-I) was induced with 0.4 mM isopropyl thiogalactoside,
and the cells were incubated for an additional 18 h at 37°C. Because the recombinant proteins were
obtained as inclusion bodies after the induction and disruption of the cells, they were solubilized in
the solubilization buffer (50 mM Tris–HCl pH 8.0, 0.2 M NaCl, 1 mM ethylenediamine tetraacetate,
6 M guanidine hydrochloride), and the protein was refolded in the refolding buffer (0.1 M Tris–HCl
pH 8.0, 0.4 M L-arginine, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione,
0.1 mM phenylmethylsulfonyl fluoride). After dialysis of the refolded proteins in Tris-buffered saline
124 (TBS; 10 mM Tris–HCl pH 7.5, 0.15 M NaCl), the protein was purified by affinity chromatography
125 using the lactose-Cellulofine column (1.4 × 4 cm). Protein concentrations were determined from the
126 molar absorption coefficients at 280 nm calculated from the amino acid compositions of the proteins.
127
128 2.4. N-terminal amino acid sequence analysis
129
130 The N-terminal amino acid sequence of the expressed protein was determined using a protein
131 sequencer PPSQ-21 (Shimadzu, Kyoto, Japan).
132
133 2.5. Hemagglutination assay
134
135 The hemagglutination assay was performed by mixing serial twofold dilutions of sample
136 proteins in TBS (30 μl) with the same volume of a 5% (v/v) suspension of rabbit erythrocytes in
137 round-bottomed microtiter plate wells (96 wells). The extent of agglutination was visually
138 determined after incubation for 1 h at room temperature.
139
140 2.6. Measurements of carbohydrate-binding activity using sugar-PD
141
142 Sugar-PDs containing disaccharides (lactose, melibiose, maltose) were prepared by the reductive
143 amination reaction between an aldehyde group of the reducing sugars and primary amino groups of
144 amino-PDs. Each reducing sugar (110 μM) was incubated with amino-PD (0.17 μM) in 1 ml of 0.2
145 M sodium phosphate buffer (pH 8.0) in the presence of 110 μM NaBH₃CN for 24 h at 45°C. The
146 solution was then dialyzed against water to remove residual reagents, and the resulting sugar-PDs
147 were collected after freeze-drying. The carbohydrate-binding activity was evaluated by the increase
148 in Rayleigh scattering of the lectin solution after the addition of different sugar-PD solutions on the
149 basis of their complex-formation abilities. After recording the initial scattering intensity at 420 nm of
150 the lectin solution (20 μg/ml, 1 ml) in TBS using the Model F-3010 Fluorescence Spectrophotometer
151 (Hitachi) at 25°C, small volumes of sugar-PD solution in the same buffer were serially added, and
the changes in the scattering intensity were recorded. Values were corrected for dilution by the
addition of the sugar-PD solution.

2.7. Homology modeling

Homology modeling of SUL-I was performed using SWISS-MODEL server (Arnold et al.
2006) by the automatic modeling mode. For the construction of the model, the crystal structure of
CSL3 (PDB code 2ZX2) (Shirai et al. 2009) was used. The figures for the protein models were
drawn using the program PyMOL (DeLano 2002).

3. Results

3.1. cDNA cloning and sequence analysis of SUL-I

mRNA purified from the large globiferous pedicellariae of *T. pileolus* was used for cDNA
synthesis. Amplification of a DNA fragment was performed by PCR using the degenerate primers
DF1 and DR1 (Fig. 1), designed on the basis of the N-terminal amino acid sequence of SUL-I
(AVGRTXEGKSLDLEXPEGYIISVNYANYGRNSPGY) reported previously (Nakagawa et al.
1999). As a result, a fragment of about 100 bp was amplified and its nucleotide sequence was
determined. The amino acid sequence deduced from the resulting fragment
(SLDLECPEGYIISVNYA) was found to correspond to residues 10–26 of SUL-I, in which
unknown residues designated as X were identified as cysteine (or half-cystine) residues. A primer
(IF1) was then designed on the basis of this sequence and used for 3'-RACE to determine the
3'-terminal sequence of cDNA. Further amplification of cDNA was performed by 3'-RACE as well as
5'-RACE using the primers F1, F2, R1, and R2, leading to the total cDNA sequence of SUL-I, as
shown in Fig. 1. The open reading frame of SUL-I consists of 927 bp, corresponding to 308 amino
acid residues. The 24 N-terminal amino acid residues were assumed to be the signal sequence, and
the mature protein contains 284 amino acid residues with a molecular mass of 30,489 Da.
3.2. Comparison of the amino acid sequence of SUL-I

BLAST search for the homologous proteins of SUL-I revealed sequence similarities with several RBLs (Fig. 2) (Tateno 2010). While the highest similarity was found with the putative RBL from the sea urchin (*Strongylocentrotus purpuratus*) egg (65% identity), SUL-I showed apparent similarities with RBLs distributed among diverse species, ranging from invertebrate to vertebrate organisms. Many RBLs have been found in fish eggs, and they are mostly composed of two or three domains with approximately 90 amino acid residues, which are referred to as SUEL domains on the basis of the similarities with the sea urchin (*Anthocidaris crassispina*) egg lectin (Ozeki et al. 1991). SUL-I was also found to contain three SUEL-like domains. When these sequences were aligned with those of two SUEL domains of the chum salmon (*Oncorhynchus keta*) egg lectin CSL3, apparent similarities, including cysteine (or half-cystine) residues, except for Cys97 and Cys123, were observed (Fig. 3). Because these cysteine residues are known to form intradomain disulfide bonds (Shirai et al. 2009), the similarity of the positions of cysteine residues strongly suggests that they adopt similar tertiary structures. Identical residues are relatively abundant around the C-terminal part of the domains, which reflects that the C-terminal portion is important to construct carbohydrate-binding sites, as revealed by X-ray crystallographic analysis of CSL3 (Shirai et al. 2009).

3.3. Expression and purification of SUL-I

To characterize SUL-I, including carbohydrate-binding ability, the gene encoding mature SUL-I was inserted into a pET-3a vector and the recombinant protein (rSUL-I) was expressed using *E. coli* cells. The expressed protein was exclusively recovered from inclusion bodies after the disruption of the induced cells. Therefore, they were once solubilized using 6 M guanidine hydrochloride and then refolded in the buffer containing arginine to promote refolding in a soluble form. The resulting
solubilized protein was then subjected to affinity chromatography using the lactose-Cellulofine column. As shown in Fig. 4, after washing the unadsorbed proteins from the column, the adsorbed proteins were eluted with 0.2 M galactose-containing buffer. SDS-PAGE of these fractions showed a band around 30 kDa (Fig. 4B), indicating that rSUL-I was successfully refolded and exhibited a galactose-binding ability. N-terminal amino acid sequence analysis confirmed the sequence up to 20 residues (AVGRTXEGKSLDLEXPEGYI), in which X was assumed to be cysteine (half-cysteine) that cannot be detected with the protein sequencer. The initiator methionine residue was found to be cleaved off after synthesis. The final yield of active rSUL-I was 1.2 mg from the culture of 1 l.

3.4. Carbohydrate-binding ability of rSUL-I

The hemagglutinating activity of rSUL-I was examined using rabbit erythrocytes. As shown in Fig. 5, after serially diluted rSUL-I solution was mixed with erythrocyte suspension, agglutination of the cells was observed at as low as 3.1 µg/mL. This revealed that rSUL-I can bind to the carbohydrate chains on rabbit erythrocytes and has more than one carbohydrate-binding site per protein molecule, which is necessary to hemagglutinate the cells. To evaluate the relative affinity of rSUL-I for various carbohydrates, a binding assay using sugar-PD (Hatakeyama et al. 2012) was performed. As shown in Fig. 6, when rSUL-I was mixed with sugar-PDs containing lactose-, melibiose-, or maltose-PD, the formation of the complexes was observed by the increase in light scattering at 420 nm. The highest increase was observed for lactose-PD, followed by melibiose-PD, although the latter showed a gradual decrease in light scattering with its increasing concentration. Based on these results, a competitive binding assay was performed using lactose-PD. After the incubation of rSUL-I with lactose-PD to pre-form their complex, several competitive carbohydrates were serially added, and the changes in light scattering were measured. As shown in Fig. 7, a decrease in the light scattering indicated that L-rhamnose shows the highest affinity for rSUL-I, followed by lactose, whereas galactose and glucose induce very low inhibition. These results
revealed that rSUL-I can bind L-rhamnose more strongly than other galactose-related carbohydrates, as could be predicted from the similarities with RBLs as mentioned above.

3.5. Homology model of the tertiary structure of SUL-I

Among SUEL-family proteins, tertiary structures of CSL3 from chum salmon eggs (Shirai et al. 2009) and mouse latrophilin (Vakonakis et al. 2008) have only been reported at present. CSL3 is composed of a homodimer of the subunit with a molecular mass of 20 kDa. The monomer subunit contains two domains with two carbohydrate-binding sites, which further form a dimer with four carbohydrate-binding sites (Shirai et al. 2009). Because individual domains of SUL-I show a significant sequence similarity with the domains of CSL3 as mentioned above, homology modeling of SUL-I was constructed using the Swiss-Model server (Arnold et al. 2006). The highest structural similarity was found between domain 3 of SUL-I and the C-terminal domain of CSL3, and their superposed models are shown in Fig. 8. As shown in Fig. 8, the highly conserved region in domain 3 of SUL-I (Asn263–Leu277) corresponds to the loop in the carbohydrate-binding site of CSL3, which surrounds the bound carbohydrate. In the case of CSL3, bound L-rhamnose is hydrogen-bonded with Asn174 and Asp179 located in this region, in addition to Glu107 that is another conserved residue in the N-terminal portion of the domain (Fig. 3). The loop composed of Cys181–Thr184 also appears to be important for maintaining appropriate orientation of the sugar at the binding site. On the other hand, Gln143 of CSL3, which forms van der Waals contact with the methyl group at the 6th position of rhamnose, is missing in SUL-I.

4. Discussion

Several biologically active proteins, including lectins, have been isolated from the venomous sea urchins, e.g., galactose-specific lectins SUL-I, II, and III from *T. pileolus*, and heparin-binding
lectin TGL-1 from *Tripneustes gratilla* (Edo et al. 2012; Nakagawa et al. 2003; Sakai et al. 2013), whereas very limited structural information has been obtained till date. Among these, SUL-I was characterized as a galactose-specific lectin, which exerts several biological activities such as chemotactic, mitogenic, and cytotoxic activities (Nakagawa et al. 2003; Nakagawa et al. 1999; Satoh et al. 2002; Takei and Nakagawa 2006). These activities were assumed to be induced via binding of lectin to the cell surface carbohydrate chains. The various biological activities of SUL-I suggest that SUL-I as well as other *T. pileolus* lectins play significant roles in the toxic action of the *T. pileolus* venom in cooperation with the other substances by disturbing normal cellular functions. To elucidate a detailed mechanism of the actions of SUL-I, it is very important to determine the structure of SUL-I, along with the characterization of the carbohydrate-binding properties. In the present study, we performed cDNA cloning and expression of SUL-I to obtain its structural and functional information.

The complete amino acid sequence deduced from cDNA revealed that SUL-I has apparent homology with RBLs from various organisms. RBLs are known to contain SUEL-domains and categorized on the basis of their domain structures (Tateno 2010). Many RBLs have been found in fish eggs, and they contain two or three SUEL domains tandemly repeated in a single polypeptide chain. From the deduced amino acid sequence, SUL-I was found to contain three SUEL domains, in which several amino acid residues are conserved compared with other SUEL domains, suggesting that SUL-I can also basically recognize carbohydrates in a similar manner as the other related lectins. rSUL-I was successfully expressed in *E. coli* cells, although its yield was relatively low. Because the expressed protein was recovered from the inclusion bodies, they were solubilized via a refolding process. The resulting soluble rSUL-I exhibited a binding ability toward the lactose-immobilized affinity column. The hemagglutination assay using rabbit erythrocytes indicated that rSUL-I has more than one carbohydrate-binding site per molecule. Therefore, it seems likely that all three SUEL domains are functional in terms of the carbohydrate-binding ability because of their
similarity, particularly around the C-terminal conserved region. The binding specificity of rSUL-I toward several different carbohydrates was examined by the binding assay using sugar-PD, which we have developed as a simple and sensitive assay method to examine carbohydrate-binding activity of lectins (Hatakeyama et al. 2012). When rSUL-I was incubated with sugar-PDs containing lactose, melibiose, or maltose, the highest increase in the complex between lectin and the sugar-PD was observed for lactose-PD, suggesting that SUL-I preferably binds β-galactoside rather than α-galactoside (melibiose-PD). On the other hand, the competitive inhibition experiments revealed that rSUL-I can bind to L-rhamnose with the highest affinity among the sugars tested, while lactose also showed a comparable affinity. These results confirmed that SUL-I is RBL, as predicted from its primary structure. Among RBLs known till date, the tertiary structure was only determined for CSL3 from chum salmon (*O. keta*) eggs (Shirai et al. 2009). CSL3 is a lectin composed of two identical subunits, each of which contains two CRDs tandemly repeated in a single polypeptide chain. As shown in Fig. 3, amino acid residues involved in the interaction with carbohydrates in SCL3 are highly conserved with SUL-I. The homology model constructed using CSL3 suggested that the structure of the carbohydrate-binding site and the rhamnose-recognition mechanism may be basically similar between these proteins. However, there is a conspicuous difference that the residue corresponding to Gln143 of CSL3, which makes van der Waals contact with bound rhamnose, is missing in SUL-I (Fig. 8). It seems possible that such a structural difference in the carbohydrate-binding site may be closely related to their target molecules. On the other hand, it is also a significant difference that SUL-I has three domains in its polypeptide chain, although CSL3 has two domains. Elucidation of the natural target molecules would provide important clues to clarify the physiological role of SUL-I. Further studies regarding molecular-recognition mechanisms as well as the three-dimensional structure of SUL-I would provide valuable insights into the role of this lectin as a toxic component in the venom of *T. pileolus.*
Acknowledgements

This work was supported by a Grants-in-Aid for Scientific Research (C) to TH (26450128) and Grants-in-Aid for Young Scientists (B) to HU (10452872) and SG (00346587) from the Japan Society for the Promotion of Science (JSPS).

Conflict of interest

The authors declare that there are no conflicts of interest.

References


Immunomodulatory activity of a pedicellarial venom lectin from the toxopneustid sea urchin, Toxopneustes pileolus. Toxin Rev. 31, 54-60.


Igci, N., Demiralp, D.O., 2011. A preliminary investigation into the venom proteome of *Macrovipera lebetina obtusa* (Dwigubsky, 1832) from Southeastern Anatolia by MALDI-TOF mass spectrometry and comparison of venom protein profiles with *Macrovipera lebetina lebetina* (Linnaeus, 1758) from Cyprus by 2D-PAGE. Arch. Toxicol.


Nakagawa, H., Tu, A.T., Kimura, A., 1991. Purification and characterization of Contractin A from the


Satoh, F., Nakagawa, H., Yamada, H., Nagasaka, K., Nagasaka, T., Araki, Y., Tomihara, Y., Nozaki,


**Figure Legends**

Fig. 1. The nucleotide and deduced amino acid sequences of SUL-I. The N-terminal amino acid sequence determined from the purified protein (Nakagawa et al. 2003) is indicated by a broken line. The N-terminal amino acid of mature protein is numbered as “+1.” The primers used for PCR are indicated by horizontal arrows.

Fig. 2. Comparison of the amino acid sequence of SUL-I and other RBLs. Alignment was conducted by Clustal Omega program (Sievers et al. 2011). The sequences are from the following species: S.purpuratus, a predicted lectin from purple sea urchin (*S. purpuratus*) (UniProt accession number W4Y3M7) (identity: 65%); CSL3, RBL from chum salmon (*O. keta*) eggs (UniProt accession number P86179) (identity: 40%) (Shiina et al. 2002); SAL, RBL from amur catfish (*Silurus asotus*) eggs (UniProt accession number Q9PVW8) (identity: 36%) (Hosono et al. 1999); C.gigas, a predicted lectin from pacific oyster (*Crassostrea gigas*) (UniProt accession number K1QXA7) (identity: 42%) (Zhang et al. 2012); B.floridae, a predicted lectin from Branchiostoma floridae (*Florida lancelet*) (UniProt accession number C3YYD1) (identity: 46%) (Putnam et al. 2008); N.vectensis, a predicted lectin from starlet sea anemone (*Nematostella vectensis*) (UniProt accession number A7T1R6) (identity: 41%) (Putnam et al. 2007); SUEL, RBL from the sea urchin (*A. crassispina*) eggs (UniProt accession number P22031) (identity: 46%) (Sasaki and Aketa 1981). Asterisks, colons, and periods indicate the positions of identical, strongly similar, and weakly similar residues, respectively. Identities were calculated by BLAST on the Uniprot website (http://www.uniprot.org/).

Fig. 3. Internal sequence similarity and comparison with those of SCL3. Conserved cysteine residues are marked by short vertical arrows. Two additional cysteine residues present in SUL-I are enclosed in circles. The residues involved in carbohydrate binding in CSL3 and their corresponding residues are involved in carbohydrate binding in CSL3 and their corresponding residues...
in SUL-I are enclosed in boxes.

Fig. 4. Expression and purification of SUL-I. A. Affinity chromatography using lactose-Cellulofine column. After elution of unadsorbed proteins with TBS, adsorbed SUL-I was eluted with TBS containing 0.2 M galactose at the position indicated by an arrow. B. SDS-PAGE of the bound protein.

Fig. 5. Hemagglutinating activity of SUL-I. Rabbit erythrocyte suspension (5% v/v) in TBS was mixed with indicated concentrations of rSUL-I in the same buffer. After 1 h, hemagglutination was visually examined.

Fig. 6. Increase in light scattering at 420 nm caused by complex formation between rSUL-I and sugar-PDs. rSUL-I (12.4 µg/mL) in TBS was incubated with lactose-PD (▲), melibiose-PD (■), or maltose-PD (●) of indicated concentrations for 10 min at 25°C, and light scattering at 420 nm was measured using a fluorescence spectrophotometer.

Fig. 7. Carbohydrate-binding specificity of rSUL-I examined by competitive inhibition of complex formation between rSUL-I and lactose-PD by various carbohydrates. A, rSUL-I/lactose-PD complex was pre-formed by mixing rSUL-I (12 µg/mL) and lactose-PD (2.1 µg/mL), and the indicated carbohydrate solutions in the same buffer were serially added. Rhamnose (●), lactose (■), methyl-α-D-galactoside (▲), methyl β-D-galactoside (♦), melibiose (○), mannose (□), galactose (△), and glucose (◇). The initial light scattering intensity of the pre-formed complex was taken as 100%. The curves for rhamnose, lactose, methyl β-galactoside, methyl α-galactoside, and melibiose were drawn by fitting the data to the logistic function using ImageJ (Schneider et al. 2012). B,
Comparison of the carbohydrate concentrations required for 50% inhibition of the complex formation (IC₅₀) between rSUL-I and sugar-PDs. The IC₅₀ values were calculated using ImageJ.

Fig. 8. Comparison between the homology model of domain 3 of SUL-I and the C-terminal domain of CSL3. The homology model of SUL-I domain 3 (red) was constructed by the Swiss-Model server (http://swissmodel.expasy.org/) and superposed with the C-terminal domain of CSL3 (PDB code 2ZX2) (blue) using the program PyMOL. A, Overall structure. B, Closeup view of the carbohydrate-binding sites.
Fig. 1
**Fig. 2**

SUL-I        ------MAMITGKLVLCCFLMASSIGMSSAAVGRTCEGKSLDLECPEGYIISVNYANYGR 54
S.purpuratus ------MAVIGKLVLCCFLVAASISSSAAVRSCGGALLSFCSGT1AIRHANYGR 53
CSL3  
SAL        ------MMLILKLSLLALATPGGLLVSANNITCYGDBQKLHCETGLII--VKSLLYGR 52
C.gigas       S-NFLVCPAGG-----QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    EDVSQEVQD5IFDVK-----EVMSALEGTSGPRVCEHQRILSINCPAQQINIVSALYGR 176
N.vectensis  ----------------------------------------VCYLCIAANTCLGRPSVFRLCENROGTLRCPKGKVIVVAYANYGR 45
SUEL  

SUL-I        N-SPGICPHKS----SNAPPCSASSSLRINHECDGRSSCSWATNDVFGDCPGRVYKYL 109
S.purpuratus TAGNIGCPHKS--------IKDTCSPAASSSFINNCRGRTCSVSATNVFGDCPETYKYL 109
CSL3  
SAL        T-DSSTCTSNRPDAQVAVTTCT-----SLPITITDGRCNLPCDCEKTLDDLQNTDPCQTYKYY 109
C.gigas       S-NLFVCPAGG-----QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    T-TRVCPSGP-----IRTRTCSRPSLARVRTSCHGKSSCSVAAASNVFGDCPETYKYL 231
N.vectensis  T-ARKVCRHNS--------IKRTTCSRKSIIKLARHCAGNALNARSNVFGDCPETYKYL 100
SUEL  

SUL-I        EVDYSCEPK-DCQRELDECGNSINMLCPYAEPAIHYAMYQOTSE--PCVPS-----163
S.purpuratus QVDSCEPK-SSQVERPEDGFIENCP-EPPAIHYAMLYVRQGQGPDSSVLCHP-----163
CSL3  
SAL        NTSFDCING-----NYAVIENHYSTLDGND-----AILIVNANVGRSS---QISCNLPGFNG 160
C.gigas       EVDYSCEPK--QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    EVSSCTC-----RPSGPRVCQEHQRLSINCPAQQ---QINIVSALGRRT---TVCSG-----280
N.vectensis  EVLYHCSYL-SSALVRFCNENQRLTCPKGK----VIVVAYANYGRTAK-GVCRNH-----151
SUEL  

SUL-I        SISTTNCAASSSLSTARSVCEGRSECSIAASNDVFGDCPITKEYLIVDVICAR-------R 218
S.purpuratus KITGRNCC-IAGNGAGLCSGCSVAAASNVFGDCPITKEYLIVDVICAR-------R 218
CSL3  
SAL        QLTDNCLOSQSTSKMAERCKGKSCIEVIPASFNVFGDCPITKEYLIVDVICAR-------R 218
C.gigas       EVDYSCEPK--QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    PIRTNCPSLPRAVRTRCQEGSSCSCSVAASNSVFGDCPYKLYEVSTCVCTPRV-----S 339
N.vectensis  SIKTTRCSRKSIIKLRACHGCNIALNARSNVFGDCPITKEYLIVHICRNSSV 211
SUEL  

SUL-I        GRSCUGSLSSLQCSSQTI5VDAFYGRTAGPEIC-------KGN-AQQOONCRASSSLNIVQG 273
S.purpuratus KITGRNCC-IAGNGAGLCSGCSVAAASNVFGDCPITKEYLIVDVICAR-------R 218
CSL3  
SAL        SIIIECSSDQLCDRG-ERIQARNYGRQH--DCSISGPRHQOLKNTLSQSTTSMNSE 160
C.gigas       EVDYSCEPK--QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    ASACEHQTVTLCRSTGQRLNVISALGRRT-AFC-----PSGPITRNTCRANSLARVRT 394
N.vectensis  FHLCENRQGQRLTCPKGKVIVVAYANYGRTAK-GVCRH-----RHMSMKTRCRSKILIRK 266
SUEL  

SUL-I        ACGRSSCCSVANNNVFDPDCPITKEYLIVDVICAR-------R 218
S.purpuratus KITGRNCC-IAGNGAGLCSGCSVAAASNVFGDCPITKEYLIVDVICAR-------R 218
CSL3  
SAL        SIIIECSSDQLCDRG-ERIQARNYGRQH--DCSISGPRHQOLKNTLSQSTTSMNSE 160
C.gigas       EVDYSCEPK--QWNTNCYSIGSSSTQVRNTCQONQWQCSISASDALFGDCPETYKYL 102
B.floridae    SCQKSCCSSVAAASNVFGDCPITKEYLIVDVICAR-------R 218
N.vectensis  ACHGENKCALNARSNVFGDCPITKEYLIVDVICAR-------R 218
SUEL  

**Fig. 2**
Fig. 3

SUL-I domain1  ----AVGRTCEGKSLDLECPEGY--IISVNYANYGRN-SPGIC----PHKSSNAPPCSAS  49
SUL-I domain2  ----QRELDCEGMSINMLCPYAETPAIHCVAMYGRQTSEPVC----PSKSISTTNCAAS 149
SUL-I domain3  ----ARRGRSCGSSLTLSGSSQ--TISVLDAYGRTAGPEIC------KGNAQDQNCRAE 241
CSL3 N-domain  ----AISITCEGSDALLQCDGAK---IHKRYNGRRQ-HDVCSIGRPDN---LDTNCLSQ 52
CSL3 C-domain  QQETISSIICSSDSQLLCDRGE---IRIQRANYGRRQ-HDVCSIGRPQHLKNTNCLSQ 152

SUL-I domain1  SSLRIINEHCDGRSSCSVHATADVPCPCRGVYKYLEVVDSCRDPDQ  97
SUL-I domain2  SSLSTARSVCEGRSECSIAASADVPCPCIGTYKYLEIDIC------- 191
SUL-I domain3  SSLNIVQSCACNGRSCSVNANINVPCVCVGYKYLEVYKCA------- 284
CSL3 N-domain  STSKMAERCQKSECIVPASEFVPQDCVGYKLDTKQSCVQ-----  96
CSL3 C-domain  STITKMAERCDGKRQCIVKVSASVPGQPCVGYKLTDVAYTCD----- 195

*:     . * *: .* : ..* ****** *.****:  * *

Fig. 3
Fig. 4

Panel A: Absorbance at 280 nm vs. Fraction No. (2 ml)

Panel B: SDS-PAGE gel showing marker bands and bound protein. Arrow indicates 30 kDa.
### Fig. 5

<table>
<thead>
<tr>
<th>rSUL-I conc. (μg/ml)</th>
<th>12.4</th>
<th>6.2</th>
<th>3.1</th>
<th>1.6</th>
<th>0.8</th>
<th>0.4</th>
<th>0.2</th>
<th>0.1</th>
<th>0.05</th>
<th>0.025</th>
<th>control</th>
</tr>
</thead>
</table>

![Image of petri dishes with varying concentrations of rSUL-I](image-url)
Fig. 6

Sugar-PD conc. (mg/ml)

D Light scattering at 420 nm (arb. units)

0 100 200 300 400

0 20 40 60 80 100

ΔLight scattering at 420 nm (arb. units)

Sugar-PD conc. (μg/ml)

0 20 40 60 80 100

Fig. 6
A

Light scattering at 420 nm (%)

Carbohydrate conc. (μg/ml)

B

IC_{50} (μg/ml)

Rhamnose  Lactose  Me-β-Gal  Me-α-Gal  Melibiose

Fig. 7
CSL3: Tyr185
SUL-I: Tyr274
CSL3: Glu107
SUL-I: Glu199
CSL3: Asn174
SUL-I: Asn263
CSL3: Asp179
SUL-I: Asp268

α-L-rhamnose

CSL3: Gln143

Fig. 8