1 **cDNA cloning and characterization of a rhamnose-binding lectin SUL-I from the toxopneustid**

2 **sea urchin *Toxopneustes pileolus* venom**

3

4 **Tomomitsu Hatakeyama[1], Ayaka Ichise[1], Tomokazu Yonekura[1], Hideaki Unno[1], Shuichiro**

5 **Goda[1], and Hideyuki Nakagawa[2]**

6

7 [1]Biomolecular Chemistry Laboratory, Graduate School of Engineering, Nagasaki University,

8 Bunkyo-machi 1-14, Nagasaki 852-8521, Japan, [2]Division of Environmental Symbiosis, Graduate

9 School of Integrated Arts and Sciences, Tokushima University, Tokushima 770-8502, Japan

10

11

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

13

14

15 Address correspondence to: Tomomitsu Hatakeyama, Biomolecular Chemistry Laboratory, Graduate

16 School of Engineering, Nagasaki University, 1-14 Bunkyo-machi, Nagasaki 852-8521, Japan. Tel:

17 +81-95-819-2686; Fax: +81-95-819-2684; E-mail: thata@nagasaki-u.ac.jp

18

19 Abbreviations: CRD, carbohydrate-recognition domain; SUEL, sea urchin egg lectin; SUL-I, sea

20 urchin lectin-I; PD, polyamidoamine dendrimer; RACE, rapid amplification of cDNA ends; RBL,

21 rhamnose-binding lectin; TBS, Tris-buffered saline

22

23

24 **Abstract**

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

40

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

42

43 **1. Introduction**

44 The venom obtained from the globiferous pedicellariae of the sea urchin *Toxopneustes pileolus*
45 contains several biologically active proteins (Kimura et al. 1975; Nakagawa et al. 2003; Nakagawa et
46 al. 1991). Among these, the galactose-specific lectin SUL-I isolated from the venom of the large
47 globiferous pedicellariae shows various activities such as chemotactic activity on guinea pig
48 neutrophils and mitogenic activity on murine splenocytes by binding to the carbohydrate chains on

49 target cells (Nakagawa et al. 1996; Takei and Nakagawa 2006). N-terminal sequence analysis of

50 SUL-I suggested that this lectin has some similarity with rhamnose-binding lectins (RBLs), majority

51 of which have been isolated from fish eggs (Tateno 2010). RBLs are also referred to as the sea urchin

52 egg lectin (SUEL) family because of their homology with those found in the eggs of the sea urchin

53 A*nthocidaris crasssispina*. Till date, the lectins having homology with SUEL have mostly been

54 found in fish eggs, while some homologous proteins have also been found in mammals, e.g., mouse

55 latrophilin-1, a putative G-protein-coupled receptor involved in synaptic function (Vakonakis et al.

56 2008).

57 In various organisms, lectins are known to play important roles in molecular and cellular

58 recognition processes because of the vast diversity of the carbohydrate chain structures on their

59 surface. Lectins are categorized into several families (Kilpatrick 2002). Among these, two major

60 groups are C-type lectins and galectins (S-type lectins) (Drickamer 1988). C-type lectins were named

61 owing to their Ca^{2+}-dependent carbohydrate-binding activity, and these lectins contain common

62 carbohydrate-recognition domains (CRDs) composed of 110–130 amino acid residues. They are

63 distributed in various organisms and are known to play important roles in various biological

64 molecular recognition systems, including the immune system and cell adhesion processes

65 (Drickamer 1999). Some C-type lectins and C-type lectin-like proteins (Zelensky and Gready 2005)

66 have also been found abundantly in snake venoms (Igci and Demiralp 2011). They contribute to the

67 toxicity of the venom by binding to the carbohydrate chains on target cells of the victims. Therefore,

68 it may be important to elucidate the structure and function of the lectins in animal venoms to

69 understand their implications in toxicity. However, there is very limited knowledge concerning the

70 lectins in animal venoms, particularly those from marine organisms.

71 In the present study, we have cloned SUL-I cDNA from the large globiferous pedicellariae of *T.*

72 *pileolus* and expressed it in *Escherichia coli* cells to elucidate its structure and carbohydrate-binding

73 properties. The results reveal its structural relationship with RBLs. The putative three-dimensional

74 structure constructed by homology modeling using SUEL domain provides insights into its

75 carbohydrate-recognition mechanism.

76

77 **2. Materials and Methods**

78 *2.1. Materials*

79 Oligonucleotides and polyamidoamine dendrimers (PDs) with 64 amino surface groups

80 (ethylenediamine core, generation 4.0, M.W. 14,214) (amino-PD) were purchased from

81 Sigma-Aldrich. The plasmid vectors used in this study were as follows: pTAC2 vector was obtained

82 from DynaExpress and pET-3a expression vector was obtained from Novagen. Melibiose, lactose,

83 and mannose were obtained from Wako Pure Chemicals (Osaka, Japan). Rabbit blood was obtained

84 from Nippon Bio-Test Laboratories (Tokyo, Japan). The lactose-immobilized Cellulofine

85 (lactose-Celluloine) column was prepared by attaching lactose to Cellulofine gels (JNC Corp., Tokyo,

86 Japan) using the cross-linking reagent divinyl sulfone (Sigma-Aldrich), as described in previously

87 (Hatakeyama et al. 1994). All other chemicals were of analytical grade for biochemical use.

88

89 *2.2. cDNA cloning of SUL-I*

90 The globiferous pedicellariae of *T. pileolus* were flash frozen in liquid nitrogen and ground to a

91 powder form. Total RNA was extracted using Isogen solution (Nippon Gene, Tokyo, Japan). Poly(A)

92 RNA was collected using the Oligotex-dT30 mRNA Purification Kit (Takara, Otsu, Japan), and

93 cDNA was synthesized using the SMARTer cDNA Cloning Kit (Clontech). A DNA fragment

94 corresponding to the N-terminal region of SUL-I cDNA was amplified by polymerase chain reaction

95 (PCR) using two degenerate primers, DF1: 5′-GTIGGIMGIACITGYGARGGIAA-3′ and DR1:

96 5′-CCIGGISWRTTICKICCRTARTT-3′. This DNA fragment was cloned into pTAC-2 vector using *E.*

97 *coli* JM109 cells (Clontech) and sequenced using ABI PRISM 3130 Genetic Analyzer (Applied

98 Biosystems). The amino acid sequence deduced from this DNA fragment was in accordance with

99 that determined from the purified protein. Therefore, 3′- and 5′-rapid amplification of cDNA ends
100 (3′-RACE and 5′-RACE) were then performed using the primer (IF1) designed from this region
101 using the SMARTer cDNA Cloning Kit. Primers used for the degenerate PCR, 3′-RACE, and
102 5′-RACE are indicated by arrows in Fig. 1. The amino acid sequence of SUL-I was compared with
103 the Uniprot database (www.uniprot.org) by BLAST search (Altschul et al. 1990). Multiple
104 alignments of the sequences were performed using Clustal Omega (Sievers et al. 2011). The
105 determined nucleotide sequence was deposited in DDBJ/EMBL/GenBank (accession number:
106 LC003233). Chemical and physical parameters for SUL-I were calculated from the deduced
107 sequence using the ProtPram tool in ExPASy Bioinformatics Resource Portal (www.expasy.org)
108 (Artimo et al. 2012; Gasteiger et al. 2005).
109
110 *2.3. Expression and purification of the recombinant SUL-I in* E. coli *cells*
111 The coding region of the mature SUL-I protein with the initiator methionine residue was
112 amplified by PCR using two primers (forward:
113 5′-AAGGAGATATACATATGGCTGTGGGAAGAACTTGTGA-3′ and reverse:
114 5′-GTTAGCAGCCGGATCCATCAGCTGATTCCCAGCCAT-3′) and inserted into a pET-3a vector
115 at *Nde*I and *Bam*HI restriction sites using the In-Fusion HD Cloning Kit (Clontech). The plasmid was
116 amplified in *E. coli* JM109 cells, and the protein was expressed in *E. coli* BL21(DE3)pLysS cells
117 (Novagen). The recombinant SUL-I (rSUL-I) was induced with 0.4 mM isopropylthiogalactoside,
118 and the cells were incubated for an additional 18 h at 37°C. Because the recombinant proteins were
119 obtained as inclusion bodies after the induction and disruption of the cells, they were solubilized in
120 the solubilization buffer (50 mM Tris–HCl pH 8.0, 0.2 M NaCl, 1 mM ethylenediamine tetraacetate,
121 6 M guanidine hydrochloride), and the protein was refolded in the refolding buffer (0.1 M Tris–HCl
122 pH 8.0, 0.4 M L-arginine, 2 mM EDTA, 5 mM reduced glutathione, 0.5 mM oxidized glutathione,
123 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 The N-terminal amino acid sequence of the expressed protein was determined using a protein

130 sequencer PPSQ-21 (Shimadzu, Kyoto, Japan).

131

132 *2.5. Hemagglutination assay*

133 The hemagglutination assay was performed by mixing serial twofold dilutions of sample

134 proteins in TBS (30 μl) with the same volume of a 5% (v/v) suspension of rabbit erythrocytes in

135 round-bottomed microtiter plate wells (96 wells). The extent of agglutination was visually

136 determined after incubation for 1 h at room temperature.

137

138 *2.6. Measurements of carbohydrate-binding activity using sugar-PD*

139 Sugar-PDs containing disaccharides (lactose, melibiose, maltose) were prepared by the reductive

140 amination reaction between an aldehyde group of the reducing sugars and primary amino groups of

141 amino-PDs. Each reducing sugar (110 μM) was incubated with amino-PD (0.17 μM) in 1 ml of 0.2

142 M sodium phosphate buffer (pH 8.0) in the presence of 110 μM $NaBH_3CN$ for 24 h at 45°C. The

143 solution was then dialyzed against water to remove residual reagents, and the resulting sugar-PDs

144 were collected after freeze-drying. The carbohydrate-binding activity was evaluated by the increase

145 in Rayleigh scattering of the lectin solution after the addition of different sugar-PD solutions on the

146 basis of their complex-formation abilities. After recording the initial scattering intensity at 420 nm of

147 the lectin solution (20 μg/ml, 1 ml) in TBS using the Model F-3010 Fluorescence Spectrophotometer

148 (Hitachi) at 25°C, small volumes of sugar-PD solution in the same buffer were serially added, and

149 the changes in the scattering intensity were recorded. Values were corrected for dilution by the
150 addition of the sugar-PD solution.
151
152 *2.7. Homology modeling*
153 Homology modeling of SUL-I was performed using SWISS-MODEL server (Arnold et al.
154 2006) by the automatic modeling mode. For the construction of the model, the crystal structure of
155 CSL3 (PDB code 2ZX2) (Shirai et al. 2009) was used. The figures for the protein models were
156 drawn using the program PyMOL (DeLano 2002).
157
158 **3. Results**
159 *3.1. cDNA cloning and sequence analysis of SUL-I*
160 mRNA purified from the large globiferous pedicellariae of *T. pileolus* was used for cDNA
161 synthesis. Amplification of a DNA fragment was performed by PCR using the degenerate primers
162 DF1 and DR1 (Fig. 1), designed on the basis of the N-terminal amino acid sequence of SUL-I
163 (AVGRTXEGKSLDLEXPEGYIISVNYANYGRNSPGY) reported previously (Nakagawa et al.
164 1999). As a result, a fragment of about 100 bp was amplified and its nucleotide sequence was
165 determined. The amino acid sequence deduced from the resulting fragment
166 (SLDLECPEGYIISVNYA) was found to correspond to residues 10–26 of SUL-I, in which
167 unknown residues designated as X were identified as cysteine (or half-cystine) residues. A primer
168 (IF1) was then designed on the basis of this sequence and used for 3′-RACE to determine the
169 3′-terminal sequence of cDNA. Further amplification of cDNA was peformed by 3′-RACE as well as
170 5′-RACE using the primers F1, F2, R1, and R2, leading to the total cDNA sequence of SUL-I, as
171 shown in Fig. 1. The open reading frame of SUL-I consists of 927 bp, corresponding to 308 amino
172 acid residues. The 24 N-terminal amino acid residues were assumed to be the signal sequence, and
173 the mature protein contains 284 amino acid residues with a molecular mass of 30,489 Da.

174

175 *3.2. Comparison of the amino acid sequence of SUL-I*

176 BLAST search for the homologous proteins of SUL-I revealed sequence similarities with

177 several RBLs (Fig. 2) (Tateno 2010). While the highest similarity was found with the putative RBL

178 from the sea urchin (*Strongylocentrotus purpuratus*) egg (65% identity), SUL-I showed apparent

179 similarities with RBLs distributed among diverse species, ranging from invertebrate to vertebrate

180 organisms. Many RBLs have been found in fish eggs, and they are mostly composed of two or three

181 domains with approximately 90 amino acid residues, which are referred to as SUEL domains on the

182 basis of the similarities with the sea urchin (*Anthocidaris crassispina*) egg lectin (Ozeki et al. 1991).

183 SUL-I was also found to contain three SUEL-like domains. When these sequences were aligned with

184 those of two SUEL domains of the chum salmon (*Oncorhynchus keta*) egg lectin CSL3, apparent

185 similarities, including cysteine (or half-cystine) residues, except for Cys97 and Cys123, were

186 observed (Fig. 3). Because these cysteine residues are known to form intradomain disulfide bonds

187 (Shirai et al. 2009), the similarity of the positions of cysteine residues strongly suggests that they

188 adopt similar tertiary structures. Identical residues are relatively abundant around the C-terminal part

189 of the domains, which reflects that the C-terminal portion is important to construct

190 carbohydrate-binding sites, as revealed by X-ray crystallographic analysis of CSL3 (Shirai et al.

191 2009).

192

193 *3.3. Expression and purification of SUL-I*

194 To characterize SUL-I, including carbohydrate-binding ability, the gene encoding mature SUL-I was

195 inserted into a pET-3a vector and the recombinant protein (rSUL-I) was expressed using *E. coli* cells.

196 The expressed protein was exclusively recovered from inclusion bodies after the disruption of the

197 induced cells. Therefore, they were once solubilized using 6 M guanidine hydrochloride and then

198 refolded in the buffer containing arginine to promote refolding in a soluble form. The resulting

199 solubilized protein was then subjected to affinity chromatography using the lactose-Cellulofine
200 column. As shown in Fig. 4, after washing the unadsorbed proteins from the column, the adsorbed
201 proteins were eluted with 0.2 M galactose-containing buffer. SDS-PAGE of these fractions showed a
202 band around 30 kDa (Fig. 4B), indicating that rSUL-I was successfully refolded and exhibited a
203 galactose-binding ability. N-terminal amino acid sequence analysis confirmed the sequence up to 20
204 residues (AVGRTXEGKSLDLEXPEGYI), in which X was assumed to be cysteine (half-cysteine)
205 that cannot be detected with the protein sequencer. The initiator methionine residue was found to be
206 cleaved off after synthesis. The final yield of active rSUL-I was 1.2 mg from the culture of 1 l.
207

208 *3.4. Carbohydrate-binding ability of rSUL-I*

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

224 revealed that rSUL-I can bind L-rhamnose more strongly than other galactose-related carbohydrates,
225 as could be predicted from the similarities with RBLs as mentioned above.
226

227 *3.5. Homology model of the tertiary structure of SUL-I*

228 Among SUEL-family proteins, tertiary structures of CSL3 from chum salmon eggs (Shirai et
229 al. 2009) and mouse latrophilin (Vakonakis et al. 2008) have only been reported at present. CSL3 is
230 composed of a homodimer of the subunit with a molecular mass of 20 kDa. The monomer subunit
231 contains two domains with two carbohydrate-binding sites, which further form a dimer with four
232 carbohydrate-binding sites (Shirai et al. 2009). Because individual domains of SUL-I show a
233 significant sequence similarity with the domains of CSL3 as mentioned above, homology modeling
234 of SUL-I was constructed using the Swiss-Model server (Arnold et al. 2006). The highest structural
235 similarity was found between domain 3 of SUL-I and the C-terminal domain of CSL3, and their
236 superposed models are shown in Fig. 8. As shown in Fig. 8, the highly conserved region in domain 3
237 of SUL-I (Asn263–Leu277) corresponds to the loop in the carbohydrate-binding site of CSL3, which
238 surrounds the bound carbohydrate. In the case of CSL3, bound L-rhamnose is hydrogen-bonded with
239 Asn174 and Asp179 located in this region, in addition to Glu107 that is another conserved residue in
240 the N-terminal portion of the domain (Fig. 3). The loop composed of Cys181–Thr184 also appears to
241 be important for maintaining appropriate orientation of the sugar at the binding site. On the other
242 hand, Gln143 of CSL3, which forms van der Waals contact with the methyl group at the 6th position
243 of rhamnose, is missing in SUL-I.
244
245

246 **4. Discussion**

247 Several biologically active proteins, including lectins, have been isolated from the venomous
248 sea urchins, e.g., galactose-specific lectins SUL-I, II, and III from *T. pileolus*, and heparin-binding

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

261 The complete amino acid sequence deduced from cDNA revealed that SUL-I has apparent
262 homology with ;RBLs from various organisms. RBLs are known to contain SUEL-domains and
263 categorized on the basis of their domain structures (Tateno 2010). Many RBLs have been found in
264 fish eggs, and they contain two or three SUEL domains tandemly repeated in a single polypeptide
265 chain. From the deduced amino acid sequence, SUL-I was found to contain three SUEL domains, in
266 which several amino acid residues are conserved compared with other SUEL domains, suggesting
267 that SUL-I can also basically recognize carbohydrates in a similar manner as the other related lectins.

268 rSUL-I was successfully expressed in *E. coli* cells, although its yield was relatively low.
269 Because the expressed protein was recovered from the inclusion bodies, they were solubilized via a
270 refolding process. The resulting soluble rSUL-I exhibited a binding ability toward the
271 lactose-immobilized affinity column. The hemagglutination assay using rabbit erythrocytes indicated
272 that rSUL-I has more than one carbohydrate-binding site per molecule. Therefore, it seems likely that
273 all three SUEL domains are functional in terms of the carbohydrate-binding ability because of their

274 similarity, particularly around the C-terminal conserved region. The binding specificity of rSUL-I
275 toward several different carbohydrates was examined by the binding assay using sugar-PD, which we
276 have developed as a simple and sensitive assay method to examine carbohydrate-binding activity of
277 lectins (Hatakeyama et al. 2012). When rSUL-I was incubated with sugar-PDs containing lactose,
278 melibiose, or maltose, the highest increase in the complex between lectin and the sugar-PD was
279 observed for lactose-PD, suggesting that SUL-I preferably binds β-galactoside rather than
280 α-galactoside (melibiose-PD). On the other hand, the competitive inhibition experiments revealed
281 that rSUL-I can bind to L- rhamnose with the highest affinity among the sugars tested, while lactose
282 also showed a comparable affinity. These results confirmed that SUL-I is RBL, as predicted from its
283 primary structure. Among RBLs known till date, the tertiary structure was only determined for CSL3
284 from chum salmon (*O. keta*) eggs (Shirai et al. 2009). CSL3 is a lectin composed of two identical
285 subunits, each of which contains two CRDs tandemly repeated in a single polypeptide chain. As
286 shown in Fig. 3, amino acid residues involved in the interaction with carbohydrates in SCL3 are
287 highly conserved with SUL-I. The homology model constructed using CSL3 suggested that the
288 structure of the carbohydrate-binding site and the rhamnose-recognition mechanism may be basically
289 similar between these proteins. However, there is a conspicuous difference that the residue
290 corresponding to Gln143 of CSL3, which makes van der Waals contact with bound rhamnose, is
291 missing in SUL-I (Fig. 8). It seems possible that such a structural difference in the
292 carbohydrate-binding site may be closely related to their target molecules. On the other hand, it is
293 also a significant difference that SUL-I has three domains in its polypeptide chain, although CSL3
294 has two domains. Elucidation of the natural target molecules would provide important clues to
295 clarify the physiological role of SUL-I. Further studies regarding molecular-recognition mechanisms
296 as well as the three-dimensional structure of SUL-I would provide valuable insights into the role of
297 this lectin as a toxic component in the venom of *T. pileolus*.
298

299 **Acknoledgements**

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

303

304 **Conflict of interest**

305 The authors declare that there are no conflicts of interest.

306

307 **References**

308 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search
309 tool. J. Mol. Biol. 215, 403-410.

310 Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based
311 environment for protein structure homology modelling. Bioinformatics 22, 195-201.

312 Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel,
313 V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti,
314 R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012.
315 ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, W597-603.

316 DeLano, W.L. 2002. The PyMOL Molecular Graphics Systems. DeLano Scientific, San Carlos, CA.

317 Drickamer, K., 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. J.
318 Biol. Chem. 263, 9557-9560.

319 Drickamer, K., 1999. C-type lectin-like domains. Curr. Opin. Struct. Biol. 9, 585-590.

320 Edo, K., Sakai, H., Nakagawa, H., Hashimoto, T., Shinohara, M., Ohura, K., 2012.
321 Immunomodulatory activity of a pedicellarial venom lectin from the toxopneustid sea urchin,
322 *Toxopneustes pileolus*. Toxin Rev. 31, 54-60.

323 Gasteiger, E., Hoogland, C., Gattiker, A., S., D., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005.

324 *Protein Identification and Analysis Tools on the ExPASy Server*, Humana Press.

325 Hatakeyama, T., Karino, R., Terai, Y., Kimura, M., Unno, H., Goda, S., 2012. An assay for
326 carbohydrate-binding activity of lectins using polyamidoamine dendrimer conjugated with
327 carbohydrates. Biosci. Biotechnol. Biochem. 76, 1999-2001.

328 Hatakeyama, T., Kohzaki, H., Nagatomo, H., Yamasaki, N., 1994. Purification and characterization
329 of four Ca^{2+}-dependent lectins from the marine invertebrate, *Cucumaria echinata*. J. Biochem.
330 116, 209-214.

331 Hosono, M., Ishikawa, K., Mineki, R., Murayama, K., Numata, C., Ogawa, Y., Takayanagi, Y., Nitta,
332 K., 1999. Tandem repeat structure of rhamnose-binding lectin from catfish (*Silurus asotus*) eggs.
333 Biochim. Biophys. Acta 1472, 668-675.

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

338 Kilpatrick, D.C., 2002. Animal lectins: a historical introduction and overview. Biochim. Biophys.
339 Acta 1572, 187-197.

340 Kimura, A., Hayashi, H., Kuramoto, M., 1975. Studies of urchi-toxins: Separation, purification and
341 pharmacological actions of toxinic substances. Japan. J. Pharmacol. 25, 109-120.

342 Nakagawa, H., Hashimoto, T., Hayashi, H., Shinohara, M., Ohura, K., Tachikawa, E., Kashimoto, T.,
343 1996. Isolation of a Novel Lectin from the Globiferous Pedicellariae of the Sea Urchin
344 *Toxopneustes Pileolus*. Adv. Exp. Med. Biol. 391, 213-223.

345 Nakagawa, H., Tanigawa, T., Tomita, K., Tomihara, Y., Araki, Y., Tachikawa, E., 2003. Recent
346 studies on the phathological effects of purified sea urchin toxins. Journal of Toxicology 22,
347 633-649.

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

349 pedicellarial venom of sea urchin, *Toxopneustes pileolus*. Arch. Biochem. Biophys. 284,
350 279-284.
351 Nakagawa, H., Yamaguchi, C., Tomoyoshi, F., Hayashi, H., 1999. novel motogenic lectin from the
352 globiferous pcdicellariac of sea urchin, *Toxopneustes pileolus*. J. Chem. Soc. Pak. 21, 305-310.
353 Ozeki, Y., Matsui, T., Suzuki, M., Titani, K., 1991. Amino acid sequence and molecular
354 characterization of a D-galactoside-specific lectin purified from sea urchin (*Anthocidaris*
355 *crassispina*) eggs. Biochemistry 30, 2391-2394.
356 Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi,
357 M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutiérrez, E.L., Dubchak, I., Garcia-Fernàndez, J.,
358 Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V.,
359 Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov,
360 A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M.,
361 Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N., Rokhsar, D.S., 2008. The amphioxus
362 genome and the evolution of the chordate karyotype. Nature 453, 1064-1071.
363 Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro,
364 H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M.,
365 Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q., Rokhsar, D.S., 2007. Sea anemone
366 genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317,
367 86-94.
368 Sakai, H., Edo1, K., Nakagawa, H., Shinohara, M., Nishiitsutsuji, R., Ohura, K., 2013. Isolation and
369 partial characterization of a L-rhamnose-binding lectin from the globiferous pedicellariae of the
370 toxopneustid sea urchin, *Toxopneustes pileolus*. Int. Aquat. Res. 5, 1-10.
371 Sasaki, H., Aketa, K., 1981. Purification and distribution of a lectin in sea urchin (*Anthocidaris*
372 *crassispina*) egg before and after fertilization. Exp Cell Res 135, 15-19.
373 Satoh, F., Nakagawa, H., Yamada, H., Nagasaka, K., Nagasaka, T., Araki, Y., Tomihara, Y., Nozaki,

374 M., Sakuraba, H., Ohshima, T., Hatakeyama, T., Aoyagi, H., 2002. Fishing for bioactive
375 substances from scorpionfish and some sea urchins. J. Nat. Toxins 11, 297-304.
376 Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image
377 analysis. Nat. Methods 9, 671-675.
378 Shiina, N., Tateno, H., Ogawa, T., Muramoto, K., Saneyoshi, M., Kamiya, H., 2002. Isolation and
379 characterization of L-rhamnose-binding lectins from chum salmon (*Oncorhynchus keta*) eggs.
380 Fish. Sci. 68, 1352-1366.
381 Shirai, T., Watanabe, Y., Lee, M.S., Ogawa, T., Muramoto, K., 2009. Structure of rhamnose-binding
382 lectin CSL3: unique pseudo-tetrameric architecture of a pattern recognition protein. J. Mol. Biol.
383 391, 390-403.
384 Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H.,
385 Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of
386 high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.
387 Takei, M., Nakagawa, H., 2006. A sea urchin lectin, SUL-1, from the toxopneustid sea urchin
388 induces DC maturation from hyman monocyte and drives Th1 polirization in vitro. Toxicol.
389 Appl. Pharmacol. 213, 27-36.
390 Tateno, H., 2010. SUEL-related lectins, a lectin family widely distributed throughout organisms.
391 Biosci. Biotechnol. Biochem. 74, 1141-1144.
392 Vakonakis, I., Langenhan, T., Prömel, S., Russ, A., Campbell, I.D., 2008. Solution structure and
393 sugar-binding mechanism of mouse latrophilin-1 RBL: a 7TM receptor-attached lectin-like
394 domain. Structure 16, 944-953.
395 Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. FEBS J. 272,
396 6179-6217.
397 Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong,
398 Z., Que, H., Xie, Y., Holland, P.W., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., Peng, C., Meng,

399 J., Yang, L., Liu, J., Wen, B., Zhang, N., Huang, Z., Zhu, Q., Feng, Y., Mount, A., Hedgecock,
400 D., Xu, Z., Liu, Y., Domazet-Lošo, T., Du, Y., Sun, X., Zhang, S., Liu, B., Cheng, P., Jiang, X.,
401 Li, J., Fan, D., Wang, W., Fu, W., Wang, T., Wang, B., Zhang, J., Peng, Z., Li, Y., Li, N., Chen,
402 M., He, Y., Tan, F., Song, X., Zheng, Q., Huang, R., Yang, H., Du, X., Chen, L., Yang, M.,
403 Gaffney, P.M., Wang, S., Luo, L., She, Z., Ming, Y., Huang, W., Huang, B., Zhang, Y., Qu, T., Ni,
404 P., Miao, G., Wang, Q., Steinberg, C.E., Wang, H., Qian, L., Liu, X., Yin, Y., 2012. The oyster
405 genome reveals stress adaptation and complexity of shell formation. Nature 490, 49-54.
406
407

408 **Figure Legends**

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

413

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

429

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

433 in SUL-I are enclosed in boxes.

434

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

439

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

443

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

448

449 Fig. 7. Carbohydrate-binding specificity of rSUL-I examined by competitive inhibition of complex
450 formation between rSUL-I and lactose-PD by various carbohydrates. A, rSUL-I/lactose-PD complex
451 was pre-formed by mixing rSUL-I (12 μg/mL) and lactose-PD (2.1 μg/mL), and the indicated
452 carbohydrate solutions in the same buffer were serially added. Rhamnose (●), lactose (■),
453 methyl-α-D-galactoside (▲), methyl β-D-galactoside (♦), melibiose (○), mannose (□), galactose
454 (△), and glucose (◇). The initial light scattering intensity of the pre-formed complex was taken as
455 100%. The curves for rhamnose, lactose, methyl β-galactoside, methyl α-galactoside, and melibiose
456 were drawn by fitting the data to the logistic function using ImageJ (Schneider et al. 2012). B,

457 Comparison of the carbohydrate concentrations required for 50% inhibition of the complex

458 formation (IC_{50}) between rSUL-I and sugar-PDs. The IC_{50} values were calculated using ImageJ.

459

460 Fig. 8. Comparison between the homology model of domain 3 of SUL-I and the C-terminal domain

461 of CSL3. The homology model of SUL-I domain 3 (*red*) was constructed by the Swiss-Model server

462 (http://swissmodel.expasy.org/) and superposed with the C-terminal domain of CSL3 (PDB code

463 2ZX2) (*blue*) using the program PyMOL. A, Overall structure. B, Closeup view of the

464 carbohydrate-binding sites.

465

```
                             CCACATTTTCTGCTTTTGACTTCATCGATCAT   32

TTGTCTGTCGTCTGCTTATTGTTGTCACTACACTACCATTGGAAAGGATTCTTCTTGAAA   92

ATGGCTATGATAACAGGAAAATTGGTCCTATGTTGCTTTCTCATGGCTTCATCGATTGGA  152
 M  A  M  I  T  G  K  L  V  L  C  C  F  L  M  A  S  S  I  G   -5
-24         DF1                          IF1
ATGTCTAGTGCTGCTGTGGGAAGAACTTGTGAAGGAAAAAGTCTTGATCTCGAATGTCCT  212
 M  S  S  A  A  V  G  R  T  C  E  G  K  S  L  D  L  E  C  P   16
             +1                                      DR1
GAAGGATACATTATTAGCGTCAATTATGCCAATTATGGTCGTAATAGCCCGGGGATTTGC  272
 E  G  Y  I  I  S  V  N  Y  A  N  Y  G  R  N  S  P  G  I  C   36

CCACATAAGAGTTCCAACGCGCCACCGTGCTCTGCCTCCTCTTCCCTCCGTATCATCAAC  332
 P  H  K  S  S  N  A  P  P  C  S  A  S  S  S  L  R  I  I  N   56
                                               R2
GAGCACTGTGATGGAAGATCATCATGCAGTGTCCATGCAACCAATGATGTATTCGGCGAC  392
 E  H  C  D  G  R  S  S  C  S  V  H  A  T  N  D  V  F  G  D   76
          R1
CCTTGTCGTGGTGTTTACAAGTATCTCGAGGTAGACTACTCCTGTCGCCGTGATCCCGAC  452
 P  C  R  G  V  Y  K  Y  L  E  V  D  Y  S  C  R  R  D  P  D   96

TGTCAGAGAGAACTTGACTGCGAAGGAAATTCGATCAATATGCTTTGCCCTTATGCTGAG  512
 C  Q  R  E  L  D  C  E  G  N  S  I  N  M  L  C  P  Y  A  E  116

ACTCCGGCTATTCACATCTGTTATGCCATGTATGGACGGCAGACGTCCGAACCAGTTTGT  572
 T  P  A  I  H  I  C  Y  A  M  Y  G  R  Q  T  S  E  P  V  C  136

CCCTCAAAAAGTATTTCAACCACCAACTGCGCCGCTTCCAGCTCTTTATCCACAGCTCGA  632
 P  S  K  S  I  S  T  T  N  C  A  A  S  S  S  L  S  T  A  R  156

TCAGTCTGTGAAGGGCGATCCGAATGTTCCATTGCTGCTTCTAATGATGTATTTGGTGAC  692
 S  V  C  E  G  R  S  E  C  S  I  A  A  S  N  D  V  F  G  D  176

CCTTGCATTGGCACTTACAAGTACCTGGAGATTGACTACATATGTGCCAGACGTGGACGA  752
 P  C  I  G  T  Y  K  Y  L  E  I  D  Y  I  C  A  R  R  G  R  196

TCATGTGAAGGGAGTAGCCTGACCCTTAGCTGTTCATCTGGGCAGACCATCTCGGTCTTG  812
 S  C  E  G  S  S  L  T  L  S  C  S  S  G  Q  T  I  S  V  L  216
                                                F1
GATGCATTCTATGGTCGCACAGCAGGACCAGAGATCTGTAAAGGAAACGCGCAGGATCAG  872
 D  A  F  Y  G  R  T  A  G  P  E  I  C  K  G  N  A  Q  D  Q  236

AACTGTCGTGCCGAGAGCAGTTTGAACATTGTTCAATCTGCATGCAATGGTCGATCATCA  932
 N  C  R  A  E  S  S  L  N  I  V  Q  S  A  C  N  G  R  S  S  256

TGTTCTGTGAACGCCAACAACAATGTCTTTGGAGATCCATGCGTGGGGACTTACAAGTAT  992
 C  S  V  N  A  N  N  N  V  F  G  D  P  C  V  G  T  Y  K  Y  176

CTCGAAGTTCTCTACAAATGTGCCTGAATGGCTGGGAATCAGCTGATCAGAGACAATGAC 1052
 L  E  V  L  Y  K  C  A  *                                   284

AGAACTCACCACACCAACCATGCCAACCTTTTGGAGCAAGAAATTCTGAAGTTCTCCCCC 1112
                                                F2
TTCTCGTCAAAATGTTTCTGATGTTTGGATTAATTTCATTATGGTTTTAACTGGTTTTAT 1172

ATCGTAGTTCTATTTCCATTGAAAACATTATTATTTCCATTGATGATCATACTTAGTAAG 1232

AATATTTA                                                     1240
```

Fig. 1

```
SUL-I        ------MAMITGKLVLCCFLMASSIGMSSAAVGRTCEGKSLDLECPEGYIISVNYANYGR 54
S.purpuratus -------MAVIGKLVLCCFLVAASISSSSAIVRRSCEGGALSLSCPSGTAIRINHANYGR 53
CSL3         ------------------------------------------------------------
SAL          -------MMLILKLSLLSLLIATPGLLVSGANMITCYGDVQKLHCETGLII-VKSSLYGR 52
C.gigas      ---------MLPKLLG----FVLLFGSTYAITERACEGSTLYLTCPQGQSINVTYANYGR 47
B.floridae   EDVSQEVQDSITDDVK----EVMSALEGTSGPRRVCEHQRLSINCPAGQQINIVSALYGR 176
N.vectensis  ---------------VCYLCIAANTCLGRPSVFRLCENRQGTLRCPKGKVIVVAYANYGR 45
SUEL         ------------------------------------------------------------

SUL-I        N-SPGICPHKS----SNAPPCSASSSLRIINEHCDGRSSCSVHATNDVFGDPCRGVYKYL 109
S.purpuratus TAGHGICPHRS----IRTTSCFASSSFSIVNNNCDGRTSCSVSATNGVFGDPCPGTYKYL 109
CSL3         ------------------------------------------------------------
SAL          T-DSTTCSTNRPPAQVAVTTC--SLPITTIGDRCNGLPDCELKTDLLGNTDPCQGTYKYY 109
C.gigas      S-NLFVCPAGG----QQNTNCYSGSSIQTVRNTCQGQNQCSISASDALFGDPCPRTYKYL 102
B.floridae   T-TRTVCPSGP----IRTTNCRSPDSLARVRTSCHGKSSCSVAASNSVFGDPCYGTFKYL 231
N.vectensis  T-AKGVCRHNS----IKITRCYSRKSKILIRKACHGENKCALNARNSVYGDPCYGTYKYI 100
SUEL         -----------------------------------------------------------E 1

SUL-I        EVDYSCRRD-PDCQRELDCEGNSINMLCPYAETPAIHICYAMYGRQTSE-PVCPSK---- 163
S.purpuratus QVDYSCEPK-SSCQVERTCEGGFIELHCPE-ETPAIHICEALYGRQLPGSVLCSHP---- 163
CSL3         -------------AISITCEGSDALLQCDGA---KIHIKRANYGRRQH--DVCSIGRPDN 42
SAL          NTSFDCING----NYAVICEHGYSTLDCGND---AILIVNANYGRASS--QICSNGLPNG 160
C.gigas      EVDYECEFQSPPGNRFHVCEGGSLYLYCPRGT--YLVIFSANFGRLSS--AICPGP---- 154
B.floridae   EVSSTCI---RPSGPRRVCEHQRLSINCPAGQ--QINIVSALYGRTTR--TVCPSG---- 280
N.vectensis  EVLYHCSYL-SSALVFRLCENRQGTLRCPKGK--VIVVAYANYGRTAK--GVCRHN---- 151
SUEL         LVSEFCL------KKERVCEDSSLTISCPEGE--GIVIYDAIYGRKRG--EVCPGLF-GA 50
                               **     : *       : :  * :**      :*

SUL-I        SISTTNCAASSSLSTARSVCEGRSECSIAASNDVFGDPCIGTYKYLEIDYICAR-----R 218
S.purpuratus KIGTTNCAAFSSMHVVQSACQGRATCSVAASNNVFGDPCVGTYKYLEIEYTCAR-----R 218
CSL3         QLTDTNCLSQSSTSKMAERCGGKSECIVPASNFVFGDPCVGTYKYLDTKYSCVQQQETIS 102
SAL          LTQNTNCYAANTLTTVAGLCNGKKSCTVEALNTIFSDPCSGTVKYLTVTYICTKE----- 215
C.gigas      GSNNVNCVSSNALSVVRNSCEGYPSCQLEAINNVFGDPCPGTYKYLEVNVGCSYF----- 209
B.floridae   PIRTTNCRSPDSLARVRTSCHGKSSCSVAASNSVFGDPCYGTFKYLEVSSTCIRTPVR-S 339
N.vectensis  SIKTTRCYSRKSKILIRKACHGENKCALNARNSVYGDPCYGTYKYIEVLYHCIRRRNSSV 211
SUEL         FTKNRKCRSSNSQQVVENSCEGKSSCTVLASNSVFGDPCPGTAKYLAVTYICSFL----- 105
                  .* : .:       * *   * : * * ::.*** ** **:     *

SUL-I        GRSCEGSSLTLSCSSGQTISVLDAFYGRTAGPEIC----KGN-AQDQNCRAESSLNIVQS 273
S.purpuratus GRSCEGGTLALSCSSGQTILVLDAFYGRMAGPEIC----PHPQVSNQHCRASSSLPIVKG 274
CSL3         SIICEGSDSQLLCDRG-EIRIQRANYGRRQH-DVCSIGRPHQQLKNTNCLSQSTTSKMAE 160
SAL          MVVCEGGSASINCGAQ-TIKTIWANYGRTDS-TVCSTGRPGSQLLNTNCYTSDTLNKVAA 273
C.gigas      ------------------------------------------------------------
B.floridae   ASACEHQTVTLRCSTGQRLNIVSALYGRTTR-AFC----PSGPIRTTNCRSANSLARVRT 394
N.vectensis  FHLCENRQGTLRCPKGKVIVVAYANYGRTAK-GVC----RHNSMKTTRCYSRKSKILIRK 266
SUEL         ------------------------------------------------------------

SUL-I        ACNGRSSCSVNANNNVFGDPCVGTYKYLEVLYKCA---- 308
S.purpuratus ICNGQTSCSVSATNNVFGDPCVHTYKYLEVLYECA---- 309
CSL3         RCDGKRQCIVSVSNSVFGDPCVGTYKYLDVAYTCD---- 195
SAL          GCDHLSTCTIPANNNFFGDPCPNTYKYLRIVYACV---- 308
C.gigas      ---------------------------------------
B.floridae   SCQGKSSCSVAASNSVFGDPCYGTFKYLDVKYTCICKYQ 433
N.vectensis  ACHGENKCALNARNSVYGDPCYGTYKYIEVLYHCV---- 301
SUEL         ---------------------------------------
```

Fig. 2

```
SUL-I domain1  ----AVGRTCEGKSLDLECPEGY--IISVNYANYGRN-SPGIC----PHKSSNAPPCSAS  49
SUL-I domain2  ----QRELDCEGNSINMLCPYAETPAIHICYAMYGRQTSEPVC----PSKSISTTNCAAS 149
SUL-I domain3  ---ARRGRSCEGSSLTLSCSSGQ--TISVLDAFYGRTAGPEIC-----KGNAQDQNCRAE 241
CSL3 N-domain  ----AISITCEGSDALLQCDGAK---IHIKRANYGRRQ-HDVCSIGRPDNQLTDTNCLSQ  52
CSL3 C-domain  QQETISSIICEGSDSQLLCDRGE---IRIQRANYGRRQ-HDVCSIGRPHQQLKNTNCLSQ 152
                        ***..  : *  .    * :  * ***              .     * :.

SUL-I domain1  SSLRIINEHCDGRSSCSVHATNDVFGDPCRGVYKYLEVDYSCRRDPDC  97
SUL-I domain2  SSLSTARSVCEGRSECSIAASNDVFGDPCIGTYKYLEIDYIC------ 191
SUL-I domain3  SSLNIVQSACNGRSSCSVNANNNVFGDPCVGTYKYLEVLYKCA----- 284
CSL3 N-domain  SSTSKMAERCGGKSECIVPASNFVFGDPCVGTYKYLDTKYSCVQ----  96
CSL3 C-domain  STTSKMAERCDGKRQCIVKVSNSVFGDPCVGTYKYLDVAYTCD----- 195
               *:     . * *: .* : ..* ****** *.****:  * *
```

Fig. 3

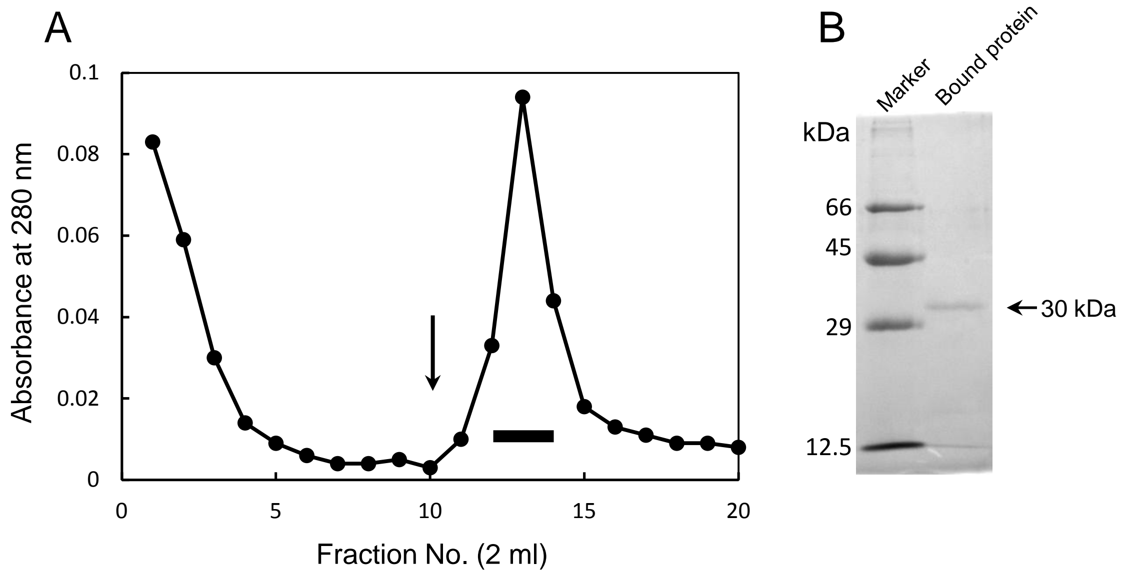


Fig. 4

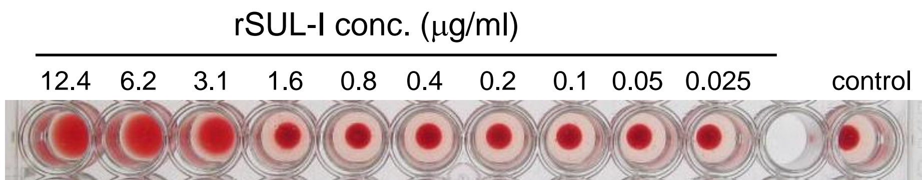


Fig. 5

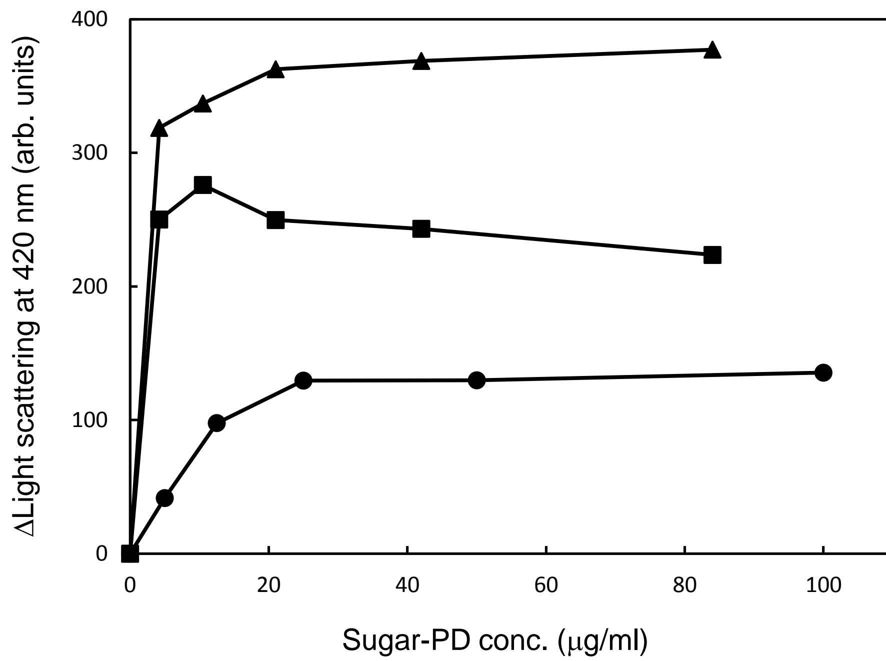
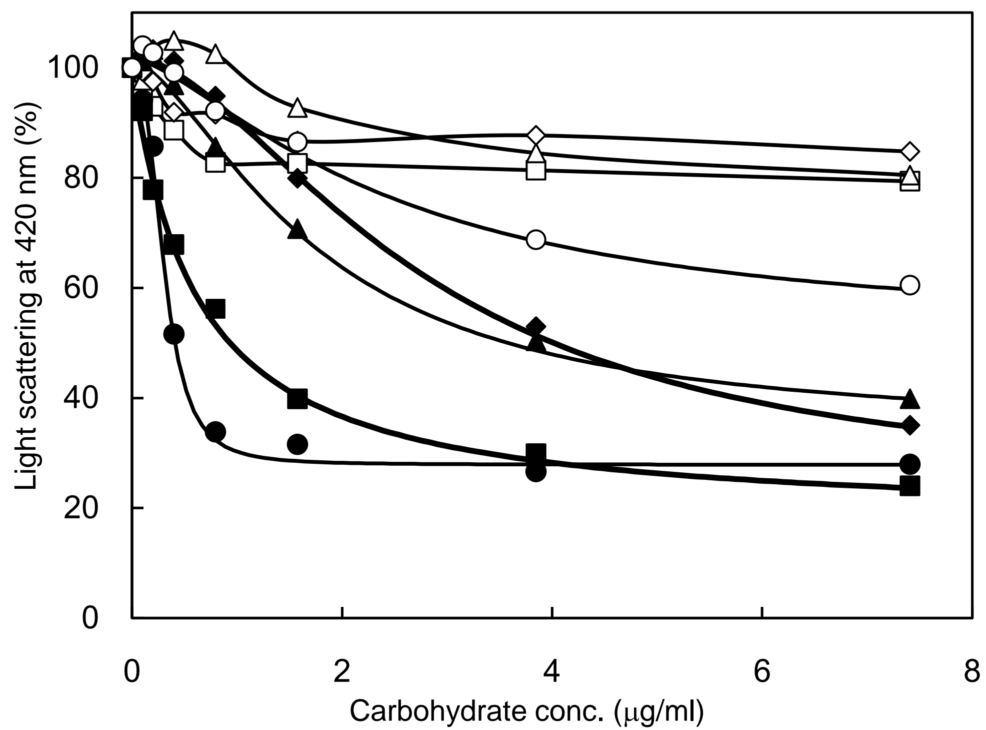


Fig. 6

A



B

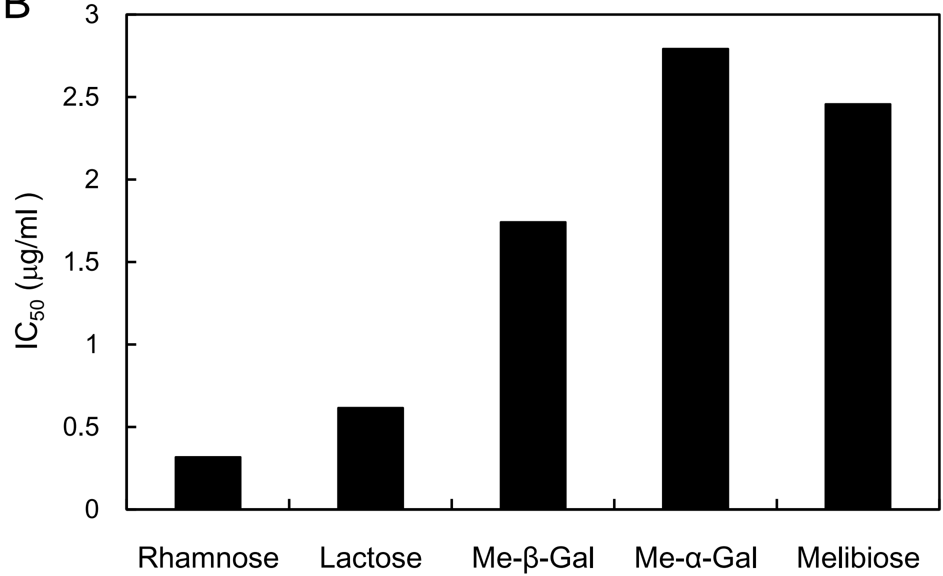


Fig. 7

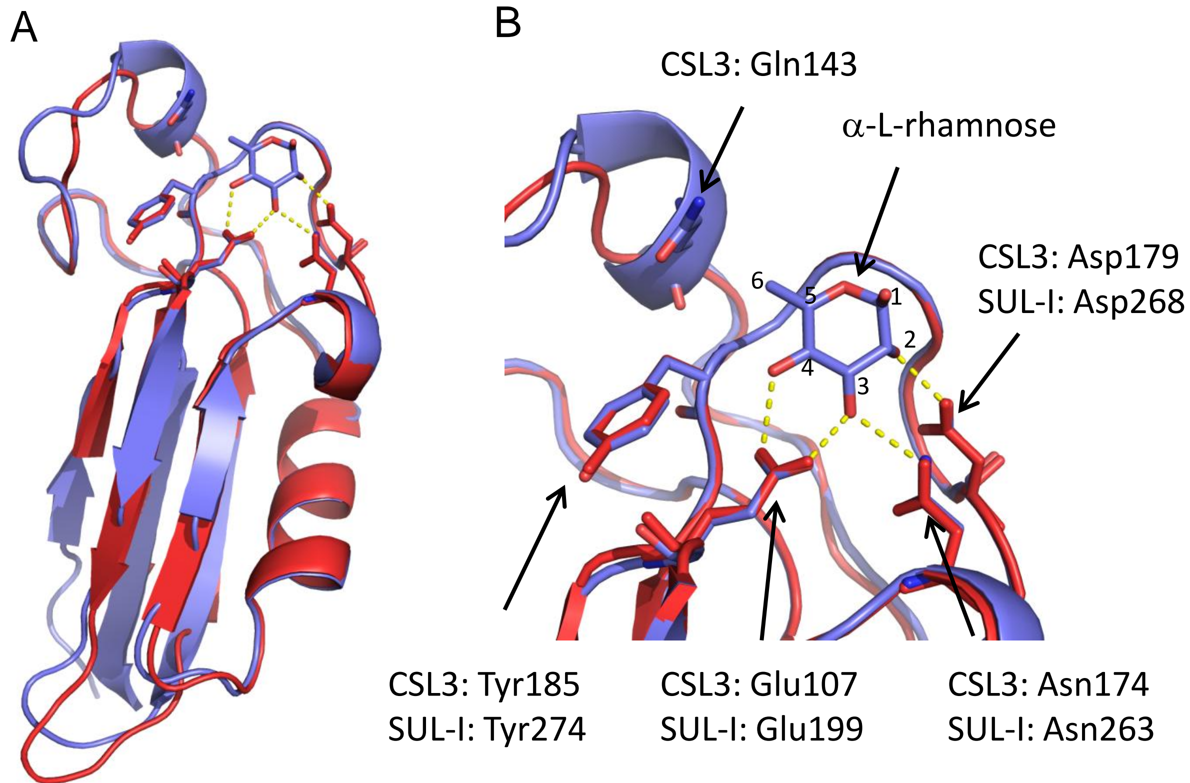


Fig. 8