Tandem repeat structure of rhamnose-binding lectin from catfish (Silurus asotus) eggs

Masahiro Hosono a, Kazunori Ishikawa a, Reiko Mineki b, Kimie Murayama b, Chifumi Numata a, Yukiko Ogawa a, Yoshio Takayanagi a, Kazuo Nitta a,*

a Cancer Research Institute, Tohoku Pharmaceutical University, 4-4-1 Komatsushima, Aoba-ku, Sendai 981-8558, Japan
b Division of Biochemical Analysis, Central Laboratory of Medical Sciences, Juntendo University School of Medicine, 2-2-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan

Received 10 June 1999; received in revised form 17 September 1999; accepted 20 September 1999

Abstract

The primary structure of catfish (Silurus asotus) egg lectin (SAL) was determined. SAL cDNA contained 1448-bp nucleotides and 308 amino acid residues, deduced from open reading frame. The SAL mature protein composed of 285-amino acid residues was followed by a predicted signal sequence having 23 residues. The mRNA of SAL was found to be expressed in eggs, but not in liver. SAL is composed of three tandem repeat domain structures divided into exactly 95 amino acid residues each, and all cysteine positions of each domain were completely conserved. Sequence homologies between the three domains, termed D1 (1–95), D2 (96–190) and D3 (191–285), were as follows; D1–D2, 28%; D2–D3, 33%; D1–D3, 43%. Two conserved peptide motifs, -(AN)YGR(TD)S(T)XCS(TGR)P- and -DPCX(G)T(Y)KY(L)-, appear to exist at the N- and C-terminal regions of each domain, respectively. The kinetic parameters of SAL obtained by measuring surface plasmon resonance were as follows: $K_a$ (M$^{-1}$) for neohesperidosyl-BSA, 7.1$\times$10$^6$; for melibiosyl-BSA, 4.9$\times$10$^6$; and for lactosyl-BSA, 5.2$\times$10$^5$. These results show that RBLs including SAL comprise a family of $\alpha$-galactosyl binding lectins having characteristic tandem repeat domain structures. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Lectin; Fish egg; cDNA sequence; Primary structure

1. Introduction

Rhamnose-binding lectins (RBLs) are widely found in fish eggs [1,2]. Other sources from which rhamnose-inhibitable lectins have been found so far include bacteria (Streptomyces sp.) [3], cockroach [4] and acorn barnacle [5]. In common, RBLs preferentially agglutinate rabbit and human type B erythrocytes and the activity is strongly inhibited by l-rhamnose or $\alpha$-galactoside rather than $\beta$-galactoside. Additionally, they do not require divalent cations such as Ca$^{2+}$ and Mg$^{2+}$ for their hemagglutinating activity [1,2]. Owing to a lack of structural informa-

Abbreviations: RBL, rhamnose-binding lectin; SAL, Silurus asotus lectin; OLL, Osmerus lanceolatus lectin; STL, steelhead trout lectin; BSA, bovine serum albumin; Lac-BSA, lactosyl BSA; Mel-BSA, melibiosyl BSA; Nhe-BSA, neohesperidosyl (Rha$_1$$\alpha$2Glc) BSA; RACE, rapid amplification of cDNA ends; AUAP, abridged universal amplification primer; PBS, phosphate-buffered saline; OKL, Oncorhynchus kisutch lectin; SSA, Sambucus sieboldiana agglutinin; CRD, carbohydrate recognition domain

* Corresponding author. Fax: +81-22-275-2013; E-mail: knitta@tohoku-pharm.ac.jp

1 The nucleotide sequence reported in this paper has been submitted to the DDBJ/EMBL/GenBank Data Bank with accession number AB020571.

© 1999 Published by Elsevier Science B.V. All rights reserved.
tion, classification of RBLs is still undefined. However, it seems that RBLs from fish eggs do not belong under any lectin categories typically known as C-type lectins or galectins because of their biochemical properties [2]. In comparisons of N-terminal amino acid sequences, the conserved tetrapeptide sequence, -YGRT-, was found in some RBLs, such as Silurus asotus (SAL) [6], Osmerus lanceolatus (OLL) [7], and Osmerus mordax dentex [8], so that RBLs may constitute a novel lectin family. Recently, two RBLs from steelhead trout (Oncorhynchus mykiss) eggs, STL1 and STL2, were isolated and their amino acid sequences were reported [9]. Among them, STL2 consists of 195 amino acids as well as two tandem repeat domains which have about 50% similarity with each other. In addition, STL2 shows 50% homology with SUEL from Anthocidaris crassispina eggs [10] and contains -YGRT- sequence as mentioned above.

RBL isolated from catfish (S. asotus) eggs, SAL, shows a molecular mass of 31750 by electrospray ionization mass spectrometry and exists mainly as a non-covalent trimer structure in solution [11]. In this paper, we describe the cDNA sequence of SAL and its deduced amino acid sequence, and its remarkable structural features.

2. Materials and methods

2.1. Materials

SAL was prepared according to the method described previously [6]. Antiserum against SAL was prepared in rabbit (Japanese white) with Freund’s incomplete adjuvant. SUEL and its antiserum were kindly provided by Dr. Y. Ozeki (Yokohama City University, Yokohama, Japan). Lac-BSA (13.4 mol lactose/mol albumin) and Mel-BSA (9.0 mol melibiose/mol albumin) were purchased from Sigma (St. Louis, MO). Nhe-BSA (14.0 mol neohesperidose/mol albumin) was synthesized by the method of Schwartz and Gray [12].

2.2. cDNA sequencing

In the previous report, we reported the N-terminal 30 amino acid sequence of SAL [6]. We have now obtained the next 18 residues as shown in Fig. 1. Two mixed primers, SALF1 and SALR1, for degenerated PCR were designed based on the corresponding N-terminal peptide sequence (see Fig. 1). Total RNA was extracted using RNeasy System (QIAGEN, Hilden, Germany) according to the manufacturer’s instructions. Nucleotide concentration was determined using GeneQuant apparatus (Pharmacia, Uppsala, Sweden). Poly(A)+RNA was directly purified with QuickPrep Micro mRNA Purification kit (Amersham Pharmacia), and reverse-transcribed with an oligo(dT)/T7 anchored primer by M-MLV reverse transcriptase (Toyobo, Tokyo, Japan). PCR was performed with a combination of SALF1 and SALR1, and first strand cDNA as a template using Ready-To-Go PCR beads (Amersham Pharmacia). Amplification conditions were as follows: 94°C for 1 min, 45°C for 1 min, 72°C for 2 min, 35 cycles. A DNA fragment approximately 150 bp long was amplified specifically, isolated from 1.5% agarose gel using QIAquick Gel Extraction kit (QIAGEN), and subsequently cloned into pGEM-T vector (Promega, Madison, WI). Inserted DNA was sequenced with universal vector primer (T7 and SP6) by use of dye terminator cycle sequencing method on an ABI373A automated DNA sequencer (PE Applied Biosystems, Foster City, CA). The flanking regions of the cloned cDNA fragment were obtained by RACE methods toward both 5’- and 3’-directions using poly(A)+ RNA. The gene-specific primers were designed based on the sequence of cloned cDNA fragment sequence.
as mentioned above. The other primer sets were occasionally prepared according to information from primer-walk sequencing. 3'-RACE was performed with adaptor primer and SuperScript II reverse transcriptase (Gibco-BRL, Rockville, MD) for the first strand cDNA synthesis, and with AUAP and GSPF1 for the following PCR (94°C for 1 min, 45°C for 1 min, 72°C for 2 min, 35 cycles). A PCR product approximately 1070 bp long was obtained by nested PCR primed with GSPF2 and AUAP. 5'-RACE was performed with terminal deoxynucleotidyl transferase, 5'-RACE abridged anchor primer (Gibco-BRL) and GSPR1. The amplified upstream region was approximately 450 bp long primed between GSPR2 and AUAP. Both RACE products were sequenced as described above.

2.3. Northern blot analysis

Northern blot analysis was performed with Alkphos Direct system (Amersham Pharmacia) according to the manufacturer's instructions. The PCR product primed with GSPF2 and AUAP was labeled with thermostable alkaline phosphatase, and used as a probe. The blotted filter (Hybond N, Amersham) was fixed by ultraviolet, and hybridized for 16 h at 55°C followed by two-step washing procedure, first at 55°C and second at room temperature. The resulting filter was treated with CDP-Star detection reagent (Amersham Pharmacia) for 5 min and exposed to Hyperfilm-ECL (Amersham) for 1 h.

2.4. Hemagglutination assay and immunodiffusion test

Hemagglutination and inhibition assay were performed according to the method described previously [13] using 2% rabbit erythrocyte suspension. Immunodiffusion test was performed according to Ouchterlony's method using 1% agarose gel plate.

2.5. Surface plasmon resonance analysis

To determine the interactions between SAL and its ligand carbohydrates, we used SPR670 Biosensing System (Nippon Laser and Electronics, Nagoya, Japan) applied to a solid phase binding assay system based on surface plasmon resonance. Lac-, Mel- and Nhe-BSA (5 μg/ml each) were immobilized on a dithiobutyric acid-coated sensor chip as ligands, respectively, after activation with water-soluble carbodiimide (16 mM) and N-hydroxysuccinimide (13 mM). PBS (10 mM, pH 7.4) was used as a mobile phase at a flow rate of 15 μl/min at 25°C. Ligand-immobilized sensor chip was then blocked with 1 M glycine and washed with 10 mM HCl.

3. Results

3.1. cDNA and deduced amino acid sequence of SAL

Fig. 2 shows the complete cDNA sequence (1448 bp) and the deduced amino acid sequence (308 residues) of SAL. SAL cDNA contained a 217-nucleotide of 5'-untranslated region, a 855-nucleotide of open reading frame (285-amino acid residues), and a 376-nucleotide of 3'-untranslated region including typical polyadenylation sequence. There were a couple of predicted translation initiation codons indicated by double circles.

The preceding region composed of 23 amino acid residues showed characteristic high hydrophobicity, so that it is considered to be a signal sequence. From the result of the predicted peptide sequence, mature protein contained N-terminal sequence as shown in Fig. 1, and had 285-amino acid residues with calculated molecular mass of 30 390 (Fig. 2). The consensus sequence for N-glycosylation site was found at position 88, and it is likely glycosylated due to the difference of molecular mass analyzed by SDS-PAGE or mass spectrometry as described above. The hydropathy profile of SAL is illustrated in Fig. 3. SAL showed a tendency of hydrophobic protein, although it was highly soluble in buffer solution. According to the secondary structure prediction [15], SAL, as a whole, is extremely rich in β-structure (data not shown).

3.2. Expression and structural characteristics of SAL

In Northern blot analysis, mRNA of SAL was expressed in eggs with RNA size of 1.6 kb, but not in liver (Fig. 4). Interestingly, SAL is composed of three tandem repeat domain structures divided exactly into 95 amino acid residues each. Furthermore, all cysteine positions of each domain were completely
conserved (Fig. 5). Sequence homologies between three domains named D1 (1-95), D2 (96-190) and D3 (191-285) were as follows; D1-D2, 28%; D2-D3, 33%; D1-D3, 43%. Two very characteristic peptide motifs, -(AN)YGR(TD)S(T)XCS(TGR)P-(YGR-motif) and -DPCX(G)T(Y)KY(L)-(DPC-motif), appear to exist at the N- and C-terminal regions of each domain, respectively. Fig. 5 also shows sequence similarity of SAL between STL2 [9], OKL from coho salmon eggs [16] and SUEL. Sequence of SAL was highly homologous to that of STL2 (40%, 78/195 residues), OKL (36%, 16/45 residues)
and SUEL (40%, 42/105 residues). OKL was purified by α-galactose-Sepharose affinity chromatography along with SAL, and its carbohydrate-binding properties were evaluated by adsorption test against some bacterial cells, such as Aeromonas salmonicida [16]. According to these results, l-rhamnose was effective to inhibit lectin adsorption to the bacterial cells, so that OKL is considered to be a member of the RBLs. Proposed tandem repeat structure models of RBLs are illustrated in Fig. 6.

3.3. Comparison of carbohydrate-binding specificity with SUEL

Unexpectedly, no precipitation line was observed between SAL and anti-SUEL antiserum by immunodiffusion test, whereas there was a strong line between SAL and anti-SAL antiserum (data not shown). In contrast, obvious precipitation lines were observed in anti-chinook salmon lectin serum against coho salmon and steelhead trout egg extract [17]. To determine whether SUEL recognizes l-rhamnose or galactosyl anomic conformation, we performed comparative inhibition assay of SUEL- versus SAL-induced hemagglutination of rabbit erythrocytes. As shown in Table 1, the hemagglutinating activity of SUEL was inhibited by l-rhamnose and by melibiose rather than lactose, so that SUEL is α-galactoside-binding lectin as well as SAL. These results indicate that the SUEL molecular structure is to display both hemagglutinating activity and l-rhamnose-binding potency.

3.4. Interaction of SAL with immobilized neoglycoproteins

According to the results of surface plasmon resonance analysis between SAL and three neoglycoproteins, the $K_a$ value (M$^{-1}$) of Nhe-BSA ($7.1 \times 10^6$) was higher than those of Mel- (4.9 $\times 10^6$) and Lac-BSA (5.2 $\times 10^5$) (Table 2). The order of these constant values was the same as the order of minimum inhibition concentrations for hemagglutination. Comparison of the kinetic parameters obtained for three neoglycoproteins revealed that $k_{ass}$ values (M$^{-1}$ s$^{-1}$) were higher for Nhe- (1.1 $\times 10^4$) and Mel-BSA (1.2 $\times 10^4$) than for Lac-BSA (3.6 $\times 10^3$), and $k_{diss}$ values (s$^{-1}$) were lower for Nhe- (1.5 $\times 10^{-3}$) and...
Mel-BSA ($2.4 \times 10^{-3}$) than for Lac-BSA ($6.9 \times 10^{-3}$). This finding suggests that SAL tends to associate with Lac-BSA more slowly and to dissociate more quickly compared with Nhe- and Mel-BSA. Although $k_{ass}$ value of Nhe-BSA was very similar to that of Mel-BSA, the difference of their $K_a$ values resulted from their dissociation rates. Recently, Hassegawa et al. reported kinetic measurement of SSA–fetuin interaction by a similar method [18]. As shown in Table 2, the $K_a$ value of SAL for Nhe-BSA was slightly lower than that of SSA for fetuin.

4. Discussion

In the past three decades, a number of fish egg lectins have been isolated and well characterized in terms of biological properties. In the case of RBLs, however, the biological functions and structures remain unclear.

Tateno et al. revealed the primary structure of STL2, an RBL from steelhead trout eggs, by use of amino acid sequencing technique [9]. STL2 is composed of two homologous domains in which all positions of cysteine residue are conserved. They also pointed out the sequence homology of several RBLs, including SAL, especially the conserved segment -YGR-. In the present study, we confirmed the amino acid sequence of SAL by use of cDNA sequencing strategy. The deduced sequence includes the N-terminal 48 amino acid sequence shown in Fig. 1 except twenty-fifth residue, S instead of C, and C-terminal four amino acid residues, -YACV, which were analyzed on HPG1009A C-terminal protein sequencer.

| SAL-D1 | (1-36) A N M I T C - Y G D - V Q K L - H C E T G L - I V K S S L Y G R T D S T T C S |
| SUEL | (10-44) K E R V - C E - - - D S S L T I S C E G E G - I V D A T Y G R K R G E Y C P |

Fig. 5. Alignment of amino acid sequence of SAL and other RBLs. Numerals in parentheses are the position number of amino acid. SAL-D1-D3, domain-1 (N-terminal unit of 95 residues), domain-2 (middle unit of 95 residues), and domain-3 (C-terminal unit of 95 residues); STL2-N and STL2-C, N-terminal 99 and C-terminal 96 residues of STL2 [9]; OKL, N-terminal 45 residues of OKL [16]; SUEL, SUEL sequence excepted N-terminal 9 residues [10]. Identical amino acid residues with SAL-D2 are in the boxes. Shaded boxes indicate the conserved cysteine residues. Hyphens represent gaps in the alignment.

| SAL-D1 | (73-95) L G N T D P C O G T Y K Y V N T S F D C I N G |
| SAL-D2 | (170-190) T I F S D P C S G T Y K Y L T V Y I C T |
| STL2-N | (77-99) F R T S D P C V G L Y E T Y T C I P A |
| STL2-C | (175-195) S L Y G D P C V G T Y K Y L E V A Y T C G |
| SUEL | (83-105) S V F G D P C P G T A K Y L A V Y T C I S F L |

Fig. 6. Domain structure model of RBLs. Schematic models were designed based on the SAL sequence data from Fig. 2 and reported sequence data of other RBLs. Abbreviations and references are described in the text. Open bar, determined sequences; lightly dotted bar, not determined area; diagonally hatched bar, YGR-motif; heavily dotted bar, DPC-motif; closed bar, N-terminal nine residues of SUEL.
(Hewlett Packard, Palo Alto, CA) with the Routine 2.0 program (data not shown). A proposed model of primary structures demonstrates clearly that RBLs are constructed of tandemly repeated domain units possessing sequence homology each other (Fig. 6). We call the three domains SAL D1 (N-terminal), D2 (middle) and D3 (C-terminal). Accordingly, STL2 can be represented as D1–D2 form from its sequence similarity. Although the complete amino acid sequences of OKL and OLL have not been clarified, they are thought to be dimeric form because of their reported molecular mass and of being ANYGR-tag which is found in D2 and D3 domain of SAL (Fig. 5). SUEL possesses lectin activity as well as SAL (Table 1), so that SUEL itself has the minimum active structure needed to act as a ‘hemagglutinin’. In addition, SAL, STL2 and SUEL have two conserved peptide motifs, termed YGR- and DPC-motif as mentioned above. Thus, it is possible that these two motifs construct a novel CRD for rhamnose.

Galectins, one of the major lectin families, are known to recognize β-galactoside moiety of glycoconjugates but not α-galactoside moiety. In contrast, RBLs show preferential affinity for α-galactoside [1,2]. We also report here the kinetic constant values of SAL for neoglycoproteins (Table 2). The $K_a$ value of Mel-BSA is one order higher than that of Lac-BSA. The results indicate the possibility that RBLs are present as α-galactoside-binding lectin family in contrast to galectin family [19]. However, there is an enigma about the function of the ‘rhamnose’-binding property. If the ligand of RBL is restricted to endogenous glycoprotein, there is not much meaning to estimate affinity to rhamnosyl-moiety, because rhamnose is a very rare sugar in animals. In this case, rhamnose may act as a ‘mimic’ sugar of a real endogenous substance. On the other hand, assuming that the ligand is an exogenous substance, it is possible that RBLs serve as part of a defense system because rhamnose is widely distributed in cell wall of bacteria. Only a few lines of evidence have been reported so far regarding physiological functions of RBLs, for example, contribution to the defense system with opsonin-like effect [20], and to fertilization [21,22] or oocyte maturation [1]. Tateno et al. also reported that STL2 shows sequence similarity to the ligand binding domains of LDL receptor superfamily including vitellogenin receptor, and that STL2 is associated with egg yolk proteins Lv and β', components of vitellogenin, by surface plasmon resonance analysis [9]. Vitellogenin is the serum phospho-lipoglycoprotein precursor to egg yolk, and has been noted as a potential biomarker for environmental estrogens [23]. Structural features of RBLs suggest that SUEL has a core structure of this family, and the others might be its oligomeric variant. Recently, a 32 kDa α-galactoside-binding and Ca$^{2+}$-independent lectin called SUL-1 was isolated from the large globiferous pedicellariae of sea urchin, Toxopneustes pileolus, which is different from the SUEL origin. The N-terminal 35 amino acid sequence of SUL-1 was analyzed as follows: AVGRTXEGKSLDEXPGENYISVNYANYGRNSPGI (H. Nakagawa, unpublished data). Interestingly, SUL-1 also has the ANYGR-tag and distinctive biological activity of chemotactic properties for guinea-pig neutrophils. These findings provide clues to elucidate the struc-

Table 1

<table>
<thead>
<tr>
<th>Concentration (mM)</th>
<th>SAL</th>
<th>SUEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>l-Rhamnose</td>
<td>0.012</td>
<td>0.38</td>
</tr>
<tr>
<td>d-Galactose</td>
<td>1.5</td>
<td>15</td>
</tr>
<tr>
<td>Lactose</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Melibiose</td>
<td>0.38</td>
<td>1.5</td>
</tr>
<tr>
<td>Neohesperidose</td>
<td>0.38</td>
<td>NT$^a$</td>
</tr>
</tbody>
</table>

$^a$Two percent of rabbit erythrocyte suspension and 8 hemagglutination units of each lectin were used.
$^b$Minimum concentration of saccharides required for complete inhibition.

$^c$NT, not tested.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>SAL</th>
<th>SSA$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lac-BSA</td>
<td>$K_a$ (M$^{-1}$)</td>
<td>5.2×10$^5$</td>
</tr>
<tr>
<td>Mel-BSA</td>
<td>$K_a$ (M$^{-1}$)</td>
<td>4.9×10$^6$</td>
</tr>
<tr>
<td>Nhe-BSA</td>
<td>$K_a$ (M$^{-1}$)</td>
<td>7.1×10$^6$</td>
</tr>
<tr>
<td>Fetuin</td>
<td>$K_a$ (M$^{-1}$)</td>
<td>1.4×10$^7$</td>
</tr>
<tr>
<td>$k_{diss}$ (s$^{-1}$)</td>
<td>9.4×10$^{-4}$</td>
<td></td>
</tr>
<tr>
<td>$k_{diss}$ (s$^{-1}$)</td>
<td>1.5×10$^{-7}$</td>
<td></td>
</tr>
<tr>
<td>$k_{diss}$ (s$^{-1}$)</td>
<td>6.0×10$^{-8}$</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Cited from [18].
$^b$Minimum concentration of neoglycoproteins for complete inhibition of SAL-induced hemagglutination (cited from [6]).
References


