CLONING cDNA SEQUENCE CODING THE KSA-1 LECTIN FROM RED ALGA KAPPAPHYCUS STRIATUM CULTIVATED IN VIETNAM

Le Dinh Hung, Dinh Thanh Trung

Nha Trang Institute of Technology Research and Application, Vietnam Academy of Science and Technology

To whom correspondence should be addressed. E-mail: ledinhungims@yahoo.co.uk

Received: 24.8.2016
Accepted: 30.11.2016

SUMMARY

The red alga Kappaphycus striatum is economically important food species and extensively cultivated in Vietnam. This cultivated alga may thus now contribute as a source of not only carrageenan, but also other bioactive compounds for biochemical and medicinal uses. KSA-1 lectin from this alga has been evaluated for biochemical properties, including high-mannose binding specificity, N-terminal amino acid sequence and molecular mass. In this study, we have cloned the cDNA clone encoding the lectin KSA-1. The full-length sequence of cDNA clone of KSA-1 lectin encoded a polypeptide of 268 amino acids including initiating methionine, with four tandemly repeated domains of about 67 amino acids, and sharing 43% sequence identity. The primary structure and the amino acid residues interact with mannopentaose core structure per a repeat domain of the KSA-1 lectin highly resemble and match those of anti-virus the lectin family in lower organisms, including bacteria [BOA from Burkholderia oklahomensis EO147, MBHA from Myxococcus xanthus and PFA from Pseudomonas fluorescens P10-1], cyanobacteria [OAA from Oscillatoria agardhii] and red algae [ESA-2 from Eucheuma serra, EDA-2 from Eucheuma denticulatum, KAA-1 and KAA-2 from Kappaphycus alvarezi and KSA-2 from K. striatum]. Analysis of predicted secondary structure of KSA-1 showed twenty β-strands. Each repeated domain comprises the five β-strands and matches with number of β-strands of BOA lectin. Therefore, the red alga K. striatum could be a good source for application in biomedicine and biochemistry as materials for production of the functional lectin(s).

Keywords: Amino acid sequence; cDNA of KSA-1; Kappaphycus striatum; Lectin; Lower organisms; mRNA; Red alga

INTRODUCTION

Lectins, or carbohydrate-binding proteins, are present in various organisms from virus to mammal, and serve as recognition molecules between cells, cell and matrix, and organisms. Owing to the capability of discriminating carbohydrate structures, lectins are used, not only, as valuable biochemical reagents in many research fields, including glycomics, but also as promising candidates for medicinal and clinical application (Sharon, Lis, 2003). Recently, several high-mannose N-glycan binding lectins are newly found and marked as potential microbicides having anti-HIV activities, such as OAA lectin from cyanobacterium Oscillatoria agardhii (Sato et al., 2007; Koharudin et al., 2011), actinohivin from actinomycete Longispora albida (Chiba et al., 2001), PLA from Pseudomonas fluorescens (Sato et al., 2012), BOA from Burkholderia oklahomensis (Whitley et al., 2013), MBHA from Myxococcus Xanthus (Koharudin et al., 2012), ESA-2 from Eucheuma serra (Hori et al., 2007), KAA-1 and KAA-2 from Kappaphycus alvarezi (Sato et al., 2011; Hirayama et al., 2016). These lectins show strong anti-HIV activities by blocking the entry of viruses into host cells through binding to the mannoside structures in the viral envelope glycoproteins, which are critical for the primary infection of viruses (Balzarini, 2006). Interestingly, these lectins commonly have tandem-repeated structures of a similar domain composed of about 67 amino acids and grouped into two types of four and two repeats (Hori et al., 2007; Sato et al., 2007; Whitley et al., 2013; Koharudin et al., 2012), and as exemplified with EDA-2 from Eucheuma denticulatum (Hung et al., 2015a) and KSA-2 from Kappaphycus striatum (Hung et al., 2015b). High-mannose N-glycan structures are also present on the
surfaces of the other enveloped viruses such as influenza, hepatitis C, human herpes, West Nile, severe acute respiratory syndrome-related coronavirus (SARS-CoV) and Ebola (Okuno et al., 1992; Vigerust, Shepherd, 2007), suggesting that high-mannose binding lectins could also inactivate the entries of those viruses into host cells (Barrientos et al., 2003; Helle et al., 2006; O’Keefe et al., 2010; Koharudin, Gronenborn, 2014).

Algae belonging to genus Eucheuma and Kappaphycus are widely cultivated as food and carrageenanophyte. Kappaphycus cultivation was originally started in the Philippines in latter half of the 1960s using the local varieties selected from the wild (Doty, 1973). Presently, K. striatum became the most widely cultivated commercial eucheumoid and has globally been introduced for cultivation and some experimental purposes in tropical and subtropical area, including Asia, Africa, the Pacific Island, and American countries (Ask, Azanza, 2002). We previously reported the KSA-1 and KSA-2 lectins isolated from the cultivated sample of K. striatum that preferentially recognized as high-mannose N-glycans (Hung et al., 2011). However, little information is known about genes encoding lectins from this alga, and recently the gene encoding KSA-2 lectin from K. striatum had been reported (Hung et al., 2015b). In this study, we cloned the cDNA encoding KSA-1 and elucidated the primary structure of deduced KSA-1 lectin. The results of this research provide valuable information regarding the relationship of the lectin group from these cultivated algae for future applications.

MATERIALS AND METHODS

Materials

The red alga K. striatum was collected at Ninh Thuan province, Vietnam, in March, 2013. A small portion of the alga was stored at -20°C in RNAlater solution (Invitrogen, USA) until used for the RNA extraction. GeneRacer kit was obtained from Life Technologies (Invitrogen, USA).

Rapid amplification of the 3’ cDNA ends (3’RACE) of KSA

Total RNA of K. striatum was extracted from 2 gram of the RNAlater-treated fresh algal tissues by using the plant RNA isolation reagents (Invitrogen, USA). Messenger RNA (mRNA) purification from the total RNA was performed using Oligotex™-dT30 mRNA purification Kit (TaKaRa, Japan). Complementary DNAs (cDNAs) were synthesized from 120 ng of mRNA using a GeneRacer kit (Invitrogen) according to the manufacturer’s instruction.

The first polymerase chain reaction (PCR) for rapid amplification of the cDNA 3’ end (3’RACE) was performed with eight aliquots of a 10 µL reaction mixture containing 1 µL of a 10 Blend Taq buffer (Toyobo, Osaka, Japan), 2 pmol of each deoxynucleotide triphosphate (dNTP), 6 pmol of the GeneRacer_3’_Primer, which was obtained from the GenRacer kit (Invitrogen), 2 pmol of KSA_3’ common-F1 primer, which was designed from the N-terminal amino acid sequence among lectins including ESA-2 (Hori et al., 2007) and KSA-2 (Hung et al., 2011), 0.2 µL of a 10-fold diluted synthesized cDNA, and 0.25 units of Blend Taq DNA polymerase (Toyobo). The reactions for eight aliquots were performed with a T Gradient Thermocycler (Biomat, Göttingen, Germany) under the following conditions: denaturation at 94 °C for 5 min, followed by 30 cycles consisting of denaturation at 94 °C for 30 s, annealing at 8 different temperatures starting from 50 to 64 °C (2 °C increments) for 30 s, and extension at 72 °C for 1 min, and the final extension step at 72 °C for 5 min. PCR products were subcloned into pGEM-T Easy vector (Promega, USA) and transformed into E. coli DH5a competent cells. Plasmids from the transformants were purified with a HiYield Plasmid Mini kit (RBC Biosciences, Taiwan) according to the manufacturer’s instructions. DNA sequencing was performed by using BigDye Terminator Cycle Sequencing kit ver. 3.1 with an ABI 3130 L genetic analyzer (Applied Biosystems).

Rapid amplification of the 5’ cDNA ends (5’RACE) of KSA

The first PCR of 5’RACE was performed in the same way as 3’RACE as described above, except that GeneRacer_5’_Primer (GenRacer kit, Invitrogen) and a degenerated primer KSA_5’ RACE_R1 as a primer pair. The nested PCR was performed with a 50 µL reaction mixture containing 5 µL of a 10 x Blend Taq buffer by the same method, except that 1 µL of a 100-fold diluted solution containing the first PCR product was used as a template. The nested PCR was performed with a primer pair of GeneRacer_5’_Nested Primer (GenRacer kit, Invitrogen) and the degenerated primer KSA_5’RACE_R2, which was designed from the conserved sequence among the lectin family of
red algae, carrageenophyte, including *E. serra* (ESA-2) (Hori et al., 2007) and *K. striatum* (KSA-2) (Hung et al., 2011). Subcloning and DNA sequencing were performed as described above.

To verify the sequence accuracy, full-length cDNA of KSA was further amplified by using a primer pair of KSA_5'End_F and KSA_3'End_R, which designated from the 5' and 3' terminal sequence of KSA cDNA obtained by 5'RACE and 3'RACE, respectively, (Table 1). Subcloning and DNA sequencing were then performed as described above.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (from 5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>KSA 3' common F1</td>
<td>AGAACCGTGGGGAGGATCT</td>
</tr>
<tr>
<td>KSA 5' RACE R1</td>
<td>AYTGRTTYCTACRTRTATIGRTC^</td>
</tr>
<tr>
<td>KSA 5' RACE R2</td>
<td>ATIGGICCYTCICCYTTARTAYTCG</td>
</tr>
<tr>
<td>KSA 5' End F</td>
<td>ATAGCTGAGTCAGTTACACAACC</td>
</tr>
<tr>
<td>KSA 3' End R</td>
<td>ACGAATTGTCAAAGCCCTCC</td>
</tr>
<tr>
<td>GeneRacer 3' Primer ^</td>
<td>GCTGTCACGATACGCTACGTAACG</td>
</tr>
<tr>
<td>GeneRacer 5' Primer ^</td>
<td>CGACTGACAGGACGGAGGAACCTGA</td>
</tr>
<tr>
<td>GeneRacer 5' Nested Primer ^</td>
<td>GGACACTGACATGGACTGAAGGAGTA</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

**Cloning cDNA encoding KSA-1 lectin**

cDNA cloning KSA-1 sequence was performed by 5' and 3'RACEs as described in the Materials and Methods section. The isolated full-length cDNA sequence of KSA-1 lectin consisted of 1066 bp containing 76 bp of a 5'untranslated region (5'UTR), 804 bp of an open reading frame (ORF), and 186 bp of 3'UTR (Figure 1). ORF coded a polypeptide of 268 amino acids including an initiating methionine (Figure 1). The calculated molecular mass of the deduce amino acid sequence in KSA-1 lectin cDNA was 27,825.1 Da. The 20 N-terminal amino acid sequence of KSA-1 (GRYTVQNWGGSSAPWNDAG), which had been determined by Edman degradation (Hung et al., 2011), was found in the deduced amino acid sequence of KSA-1 cDNA. The primary structure of KSA-1 has four tandemly repeated domains, each consisting of 67 amino acids and sharing 43 % sequence identity (Figure 2). cDNA clone has also elucidated that the transcripts of KSA-1 do not contain the region encoding a signal peptide sequence (Figure 1), which targets the proteins from the cytosol to the cytoplasmic membrane (prokaryotes) or to the endoplasmic reticulum membrane (eukaryotes). In addition, there are not any known cellular localization signals in the primary structure of KSA-1. It suggests that KSA-1 is also synthesized on free ribosomes and located in nucleocytoplasm.

Homologous sequences were identified with the basic local alignment search tool program (BLAST). The amino acid sequence comparison was performed using the CLUSTALW 2.0 program (Larkin et al., 2007). Secondary structure was predicted using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred). A phylogenetic tree was constructed based on the amino acid sequences by Mega v. 6.0 software using Maximum likelihood and bootstrap methods with 1000 bootstrap replications (Tamura et al., 2013).

Analysis of predicted secondary structure of KSA-1 using PSIPRED prediction program (http://bioinf.cs.ucl.ac.uk/psipred) showed twenty β-strands. Each repeated domain comprises the five β-strands, denoted as β1 to β5, β6 to β10, β11 to β15 and β16 to β20 for the first, second, third and fourth repeats, respectively. The β5-strand of the first repeated domain was connected with β6-strand of the second through a four-residue linker from Glycine 66 to Asparagine 69. The β10-strand of the second repeated domain was connected with β11-strand of the third repeated domain through a three-residue linker from Glycine134 to Aspartic acid 136. The β15-strand of the third repeated domain was connected with β16-strand of the fourth repeated domain through a four-residue linker from Glycine...
201 to Asparagine 204 (Figure 2). The predicted secondary structure of KSA-1 conformed well to the tertiary structure of the lectin BOA from *B. oklahomensis* EO147 (276 aa) (ZP_02360833). The number of β-strands of KSA-1 matches with that of BOA lectin (Whitley *et al.*, 2013).

**Figure 1.** Nucleotide and deduced amino acid sequences of KSA-1, a lectin from *K. striatum*. The stop codon TAA is shown as an asterisk. Underline indicates 20 N-terminal amino acid sequence determined by Edman degradation method (Hung *et al.*, 2011). The italicized and nonitalicized numbers represent the positions of nucleotides and amino acids, respectively.

**Figure 2.** The amino acid sequence alignment of the repeated domains and predicted secondary structure of KSA-1. Identical amino acids among the four repeated domains in the lectin molecule are indicated in red color. Underlines indicate the connected residue linkers between strands β5 and β6; β10 and β11; β15 and β16.
The nine other homologous sequences were found in database search, including five lectins from marine red algae, *E*. *serra* (ESA-2) (268 aa) (P84331), *K*. *alvarezii* (KAA-1) (LC007080) and KAA-2 (LC007081), *E*. *denticulatum* (EDA-2) (LC057379), *K*. *striatum* (KSA-2) (LC057282). All of them contain four tandemly repeated homologous domains of about 67 amino acids, except *O. agardhii* and *P. fluorescens*, which are composed of only two tandemly repeated homologous domains (Figure 3).

![Figure 3](image)

**Figure 3.** Multiple alignments of KSA-1 with related proteins. Multiple sequence alignments were carried out using the CLUSTALW 2.0 program (Larkin et al., 2007). The identical amino acids were indicated in red color. The following sequences were obtained from GenBank: *O. agardhii* (OAA) (P84330); *B. oklahomensis* EO147 (BOA) (ZP_02360833); *P. fluorescens* Pf0-1 (PFA) (YP_346241); *M. xanthus* (MBHA) (YP_635174); *E. serra* (ESA-2) (P84331); *K. alvarezii* (KAA-1) (LC007080) and KAA-2 (LC007081); *E. denticulatum* (EDA-2) (LC057379); *K. striatum* (KSA-2) (LC057282). Amino acids that interact with mannopentaose (Manα1-6(Manα1-3)Manα1-6(Manα1-3)Man) of M8/9 core unit in *B. oklahomensis* agglutinin (BOA) (Whitley et al., 2013) are indicated with asterisks.

The nine other homologous sequences were found in database search, including five lectins from marine red algae, *E. serra* (ESA-2) (268 aa) (P84331), *K. alvarezii* (KAA-1) (267 aa) (LC007080) and KAA-2 (268 aa) (LC007081), *E. denticulatum* (EDA-2) (268 aa) (LC057379), *K. striatum* (KSA-2) (268 aa) (LC057282), one protein from the cyanobacterium, *Oscillatoria agardhii* (OAA) (132 aa) (P84330), three proteins from bacteria, *B. oklahomensis* EO147 (BOA) (276 aa) (ZP_02360833), *P. fluorescens* Pf0-1 (PFA) (133 aa) (YP_346241) and *M. xanthus* (MBHA) (267 aa) (M13831). All of them contain four tandemly repeated homologous domains of about 67 amino acids, except *O. agardhii* and *P. fluorescens*, which are composed of only two tandemly repeated homologous domains (Figure 3). The degrees of similarity for identical amino acids of EDA-2, KSA-2, ESA-2, KAA-1, KAA-2, MBHA and BOA with KSA-1 amino acid sequences were 94.0, 98.9, 98.1, 97.5 and 93.8, respectively.

---

***

**KSA-1** 68

**KSA-2** 68

**EDA-2** 68

**KSA-2** 68

**EDA-2** 68

**BOA** 76

**MBHA** 67

**OAA** 67

**PFA** 67

---

***

**KSA-1** 135

**KSA-2** 135

**EDA-2** 135

**KSA-2** 135

**EDA-2** 135

**BOA** 135

**MBHA** 134

---

***

**KSA-1** 202

**KSA-2** 202

**EDA-2** 202

**KSA-2** 202

**EDA-2** 202

**BOA** 209

**MBHA** 201
99.3, 98.5, 59.8 and 57.0 %, respectively, whereas OAA from *O. agardhii* and PFA from *P. fluorescens* (each 132 residues in the N-terminal portions) showed lower sequence similarity to KSA-1 with values of 63.6 and 62.9 %, respectively.

Interestingly, several amino acid residues (23 amino acids per a single repeat domain) are absolutely conserved in all the domains among 10 lectins shown in Figure 3. The amino acid residues (7 amino acids per a single repeat domain) in BOA (Whitley *et al.*, 2013) which interact with mannopentaose (Figure 4a) are absolutely conserved in all repeated domains of the lectins (indicated with asterisks in Figure 3), except only a replacement at position 259 of proline by alanine in the fourth repeat domain of KSA-1 lectin. This structural kinship suggested that the lectins belonging to this lectin family might have similar carbohydrate-binding specificities. Three-dimensional structural analysis of BOA lectin from *B. oklahomensis* bound mannopentaose (Figure 4b) supported their strict binding specificities (Whitley *et al.*, 2013), the amino acid residues which were revealed to interact with branched mannoside in the high-mannose glycans were absolutely conserved in all the lectins of this family (Figure 3).

From a phylogenetic tree constructed based on comparison of the amino acid sequences, the lectins from the red algae, Carrageenophytes, such as *K. striatum* (KSA-1, KSA-2), *K. alvarezii* (KAA-1, KAA-2), *E. denticulatum* (EDA-2) and *E. serra* (ESA-2) showed an evolutionarily direct relationship, as expected from their taxonomic classifications, while lectin from cyanobacterium *O. agardhii* exhibited an evolutionarily closer to that of BOA lectin from *B. oklahomensis*. The results suggested that red algae and proteobacteria are closely related to each other, because the marine algal lectins from *K. striatum*, *K. alvarezii*, *E. denticulatum* and *E. serra* showed a nearer evolutionary distance to the bacterial lectins *P. fluorescens* (PFA) and *M. Xanthus* (MBHA) than the cyanobacterial OAA lectin (Figure 5), and these lectins seem to be evolved from bacterial origin. According to the endosymbiont hypothesis, the origin of chloroplasts of marine red algae is a cyanobacterial symbiont (Moreira *et al.*, 2000; Cavalier-Smith, 2000). In fact, the intense similarity of protein structures within this same family from different biological sources suggests that they are evolutionarily related. However, the evolutionary lineage of algal lectins is still in question. Considering the evolutionary distance between the red algal, cyanobacterial and bacterial lectins, we could not rule out the possibility that the lectins from cyanobacteria and macroalgae are derived from the same bacterial symbionts, since some bacteria have been isolated from cytoplasmic fluids of macroalgae and such symbionts are frequently observed on cyanobacterial surfaces (Fourcans *et al.*, 2004). For instance, *P. fluorescens* has been found in the rhizoid of marine algae (Chisholm *et al.*, 1996). This same lectin family discovered in lower organisms may provide new insight into the relationship between the algal, bacterial and cyanobacterial lectins.
CONCLUSION

The primary structure and the amino acid residues per a single repeat domain which interact with mannopentaose of the KSA-1 lectin resembled and matched highly those of the anti-viral lectin family in lower organisms, suggested that lectins from the cultivated alga *K. striatum* have potentiality to represent broad antiviral activities against viruses possessing high-mannose glycans on their envelope. It is important that lectins are isolated from a globally cultivate edible algal species in contrast that any other homologous anti-HIV lectins are derived from bacteria, cyanobacterium and uncultivated algae. These functional lectins will become a strong tool by being supplied in bulk as native form, which are almost identical in their carbohydrate specificities and structures.

Acknowledgements: This research was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106-YS.06-2015.16.

REFERENCES


**NHÂN DỒNG cDNA MÀ HÒA LECTIN KSA-1 TỪ RONG DÔ KAPPAPHYCUS STRIATUM BUÔC NUÔI TRỌNG Ô VIỆT NAM**

Lê Đình Hưng, Đình Thành Trung

*Viện Nghiên cứu và Đnk dụng công nghệ Nha Trang, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*

**TÓM TÀT**

Rồng dô *Kappaphycus striatum* là loài rồng thực phẩm có giá trị kinh tế cao và được nuôi trồng rộng rãi ở...
Việt Nam. Rong nuôi tròng này không chỉ là nguồn carrageenan, mà còn có các hợp chất hoạt tính sinh học khác cho sử dụng trong hóa sinh và y sinh. Lectin KSA-1 từ loài rong này đã được tách chất và đánh giá tính chất hóa sinh, bao gồm đặc tính liên kết high-mannose, trình tự acid amin ở đầu cuoi N và khối lượng phân tử. Trong nghiên cứu này, chúng tôi nhận dòng cDNA mới mã hóa lectin KSA-1. cDNA được nhân dòng của lectin KSA-1 mã hóa một polypeptide gồm 268 acid amin được bao bọc bằng methionine. Chúi polypeptide suy diến chứa 4 vùng lặp lại nhỏ tiếp nhau, mỗi vùng bao gồm khoảng 67 acid amin và có giá trị tương đối lớn nhất là 43%. Cấu trúc bậc một và số các gốc acid amin tương tác với cấu trúc lồi pentamannose trên mỗi vùng lặp lại của lectin KSA-1 có sự tương đồng với cấu trúc bậc một và số các gốc acid amin tương tác với cấu trúc lồi pentamannose trên mỗi vùng lặp lại của họ lectin kháng virus trong vật bậc thấp, bao gồm lectin từ vi khuẩn, Burkholderia oklahomensis EO147, Myxococcus xanthus và Pseudomonas fluorescens P8-1, vi khuẩn lam, Oscillatoria agardhii và rong dò, Eucheuma serrae, Eucheuma denticulatum, Kappaphycus alvarezii và K. striatum. Phân tích cấu trúc bậc hai của KSA-1 cho thấy có 20 dải β. Mỗi khu vực lặp lại chứa 5 dải β và trùng khớp với số dải β của lectin BOA. Vì vậy, rong dò K. striatum hứa hẹn trở thành nguồn nguyên liệu tốt sử dụng trong hóa sinh và y sinh như là vật liệu để thuần lectin chức năng.

Từ khóa: Dòng cDNA của KSA-1; Kappaphycus striatum; Lectin; mRNA; Rong dò; Sinh vật bậc thấp; Trình tự acid amin