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: ledinhhungims@yahoo.co.uk

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 organinsms, including bacteria [BOA from *Burkholderia oklahomensis* EO147, MBHA from *Myxococcus xanthus* and PFA from *Pseudomonas fluorescens* Pf0-1], 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 alvarezii* 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 alvarezii* (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 carageenophyte. *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 (Biometra, Göttingen, Germany) under the following conditions: denaturation at 94 ^{o}C for 5 min, followed by 30 cycles consisting of denaturation at 94 ^{o}C for 30 s, annealing at 8 different temperatures starting from 50 to 64 ^{o}C (2 ^{o}C increments) for 30 s, and extension at 72 ^{o}C for 1 min, and the final extension step at 72 ^{o}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 Bioscience, 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.

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).

Table 1. Primer sequence used for the cDNA cloning of KSA-1. [a] These primers were inferred from the GenRacer kit (Invitrogen); [b] Key to symbols of the degenerated nucleotides: I represents inosine, Y represents C and T, R represents A and G

Primer	Sequence (from 5' to 3')
KSA 3' common F1	AGAACCAGTGGGGAGGATCT
KSA 5' RACE R1	AYTGRTTYTCIACRTTRTAITGRTC [b]
KSA 5' RACE R2	ATIGGICCYTCICCYTTRTAYTGC
KSA 5' End F	ATAGCTGAGTCAAGTTACACCAAC
KSA 3' End R	ACGAAATTGTCAAAGCCTCC
GeneRacer 3' Primer [a]	GCTGTCAACGATACGCTACGTAACG
GeneRacer 5' Primer [a]	CGACTGGAGCACGAGGACACTGA
GeneRacer 5' Nested Primer [a]	GGACACTGACATGGACTGAAGGAGTA

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 (GRYTVQNQWGGSSAPWNDAG), 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.

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).

```
ATAGCTGAGTCAAGTTACACCAACCTCTAAAACGCCACATCAGACTCTCTAATCTCTTCAT      61
TATCTTCAATTTGTGATGGGACGTTACACAGTTCAGAACCAATGGGGGGGATCCTCTGCC     121
                    M  G  R  Y  T  V  Q  N  Q  W  G  G  S  S  A      15
CCCTGGAATGATGCCGGCTTGTGGATCCTTGGCAGTCGTGGCAACCAGAATGTGATGGCT     181
 P  W  N  D  A  G  L  W  I  L  G  S  R  G  N  Q  N  V  M  A      35
ATTGATGTTAACTCCAGCGATGGAGGTGCCAACCTGAACGGTACGATGACCTACTCCGGT     241
 I  D  V  N  S  S  D  G  G  A  N  L  N  G  T  M  T  Y  S  G      55
GAAGGCCCGATTGGATTCAAGGGGGCACGCCGTGGTGAGTCAAATGTTTACGATGTAGAG     301
 E  G  P  I  G  F  K  G  A  R  R  G  E  S  N  V  Y  D  V  E      75
AACCAGTGGGGAGGATCTTCCGCCCCCTGGCACGCTGGAGGCCAATTCGTTATCGGTTCC     361
 N  Q  W  G  G  S  S  A  P  W  H  A  G  G  Q  F  V  I  G  S      95
AGGTCTGGACAGGGAGTGCTTGCCGTGAACATTACTTCTTCAGATGCGGGAAAGACATTG     421
 R  S  G  Q  G  V  L  A  V  N  I  T  S  S  D  G  G  K  T  L      115
ACTGGCACCATGACCTATGAGAGAGAAGGTCCAATTGGATTCAAGGGTACTCAGTCTGGC     481
 T  G  T  M  T  Y  E  R  E  G  P  I  G  F  K  G  T  Q  S  G      135
GGTGACACTTACAACGTAGAGAACCAGTGGGGTGGGTCCTCGGCGCCGTGGAACAAGGCA     541
 G  D  T  Y  N  V  E  N  Q  W  G  G  S  S  A  P  W  N  K  A      155
GGTATTTGGGCGCTCGGGGATCGCAGTGGGCAAGCAATGATTGCTATGGATGTTTCGTCC     601
 G  I  W  A  L  G  D  R  S  G  Q  A  M  I  A  M  D  V  S  S      175
TCGGATGGTGGAAAAACTCTAGAGGGGACGATGCAGTACAAAGGAGAGGGACCGATTGGA     661
 S  D  G  G  K  T  L  E  G  T  M  Q  Y  K  G  E  G  P  I  G      195
TTCCGTGGCAAGCTGAGCGGTGCCAACAACTACAGTGTCGAGAACCAGTGGGGCGGTTCC     721
 F  R  G  K  L  S  G  A  N  N  Y  S  V  E  N  Q  W  G  G  S      215
TCTGCTCCTTGGAACAAGGCTGGAGACTGGTTGATTGGAGACCGCCACAATCAAAACATA     781
 S  A  P  W  N  K  A  G  D  W  L  I  G  D  R  H  N  Q  N  I      235
ACCGCAGTCAAGGTGTCATCTGACAATGATGGAAAGAACCTCGACGGCACATGCACGTAC     841
 T  A  V  K  V  S  S  D  N  D  G  K  N  L  D  G  T  C  T  Y      255
GAGAGCGAGGGTGCCATTGGTTTCAAGGGAGTCGCTTCCTAATCTAAAGAGCGTTTGAGC     901
 E  S  E  G  A  I  G  F  K  G  V  A  S  *                        268
GCAAGGATCAGAACACTCCTTGGCGGATAGATCCGAACATTTAGTAGGTGTTTCTTGTTT     961
CCCTCATGATCAGTAGCCAATCATCTCTCAAAGTTCACAACAGAGTAAAAACGTGATCAC    1021
CCTCTAAGAGTTTGCAACTTAGGAAGGAGGCTTTGACAATTTCGT                   1066
```

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.
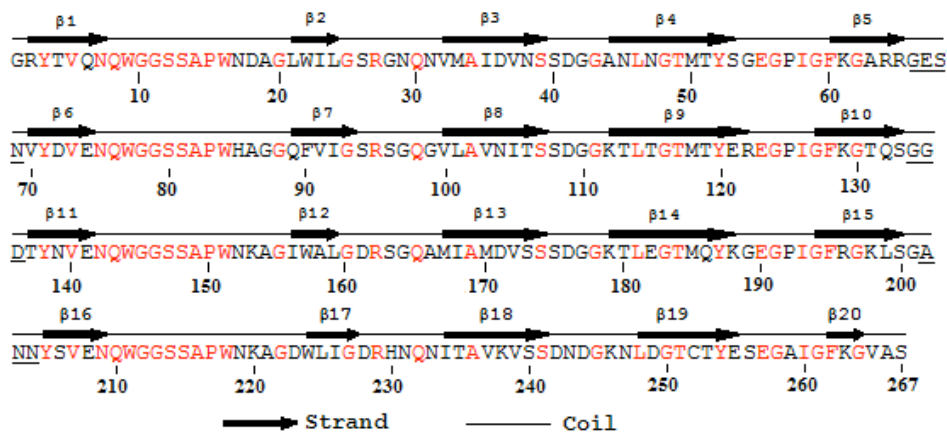

Figure 2. The amino acid sequence alignment of the repeated domains and predicted secondary structure of KSA-1. Identical amino acids among the four repeated domains in the lectin molecule are indicated in red color. *Underlines* indicate the connected residue linkers between strands β5 and β6; β10 and β11; β15 and β16.

```
               ***             *                      ***
KSA-1    1 GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAIDVNSSDGGANLNGTMTYSGEGPIGFKGARRGE  67
KSA-2    1 GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAIDVNSSDGGANLNGTMTYSGEGPIGFKGARRGE  67
EDA-2    1 GRYTVQNQWGGSSAPWNDAGLCILGSRGNQNVIAVDVTSSDGGANLGGTMTYSGEGPIGFKGARRGE  67
ESA-2    1 GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAVDVNSSDGGANLNGTMTYSGEGPIGFKGARRGE  67
KAA-1    1 GRYTVQNQWGGSSAPWNDAGLWILGSRSNQNVMAIDVNSSDGGANLNGTMTYSGEGPIGFKGARRGE  67
KAA-2    1 GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAIDVNSSDGGANLNGTMTYSGEGPIGFKGARRGD  67
BOA     10 NLQHVQNQWGGSSAPWHEGGMWVLGCRSGQNVVALNIKSDGGRTLTGTMTYVGEGPIGFRATLTQ-  75
MBHA     1 AAYLVQNQWGGSQATWNPGGLWLIGARDKQNVVALDIKSDDGKTLKGTMTYNGEGPIGFRGTLSS-  66
OAA      1 ALYNVENQWGGSSAPWNEGGQWEIGSRSDQNVVAINVESGDDGQTLNGTMTYAGEGPIGFRATLLG-  66
PFA      1 SKYAVANQWGGSSAPWHPGGTWVLGARDNQNVVAIEIKSGDGGKSFTGTMTYAGEGPIGFKAQRTG-  66

               ***             *                      ***
KSA-1   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
KSA-2   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
EDA-2   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVTALSVTSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
ESA-2   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
KAA-1   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
KAA-2   68 SNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSG  134
BOA     76 SNTYAVENQWGGSSAPWHPGGTWVIGCRVNQQVVALDIESGDQGATLAGTMTYAGEGPIGFKSQQAD  142
MBHA    67 ANNYTVENQWGGTSAPWQPGGVWWLGARDKQNIVAVSIKSNDGGKTLTGTTTYNGEGPIGFKSEVTD  133
OAA     67 NNSYEVENQWGGDSAPWHSGGNWILGSRENQNVVAINVESGDDGQTLNGTMTYAGEGPIGFKGTLT-  132
PFA     67 QNYYNVENQWGGNDAPWHPGGKWVIGGRDNQNVIALSVTSSDGGKNLSGTNTYANEGPIGFRGQIE-  132

               ***             *                      ***
KSA-1  135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  201
KSA-2  135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  201
EDA-2  135 GDTYNVENQWGGSSAPWNKAGIWALGDRNGQAMIAMDVSSSDGGQTLEGTMQYKGEGPIGFRGKLSG  201
ESA-2  135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  201
KAA-1  135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  201
KAA-2  135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  201
BOA    143 GGVYAVENQWGGSSAPWHNGGVWVIGARDQ-AVVAVSIGSTDSGKTLNGNMTYAGEGPIGFKGNSVA  208
MBHA   134 GDTYSVENQWGGSAAPWHSGGVWVLGTRGKQNVINVDAKSNDGGKTLSGTMTYNGEGPIGFRGTLTS  200

               ***             *                      ***
KSA-1  202 ANNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYESEGAIGFKGVAS--  267
KSA-2  202 ANNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  268
EDA-2  202 ANNYAVENQWGGSSAPWNKAGDWLIGDRYNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  268
ESA-2  202 ANNYSVENQWGGSSAPWNAAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  268
KAA-1  202 ANNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYESEGPIGFKGVAS--  267
KAA-2  202 ANNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  268
BOA    209 GNNYAVENQWGGTSAPWHPGGIWLLGCRSGQNVVELYITSGDNGNTFHGSMTYSGEGPIGFRAMALPQ  276
MBHA   201 PDTYTVENQWGGSTAPWNPGGFWMIGARNGQNVVALNVASSDGGKTLAGTMIYNGEGPIGFRARLG--  266
```

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,

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 (Manα1-6(Manα1-3)Manα1-6(Manα1-3)Man) (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 analyse 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).
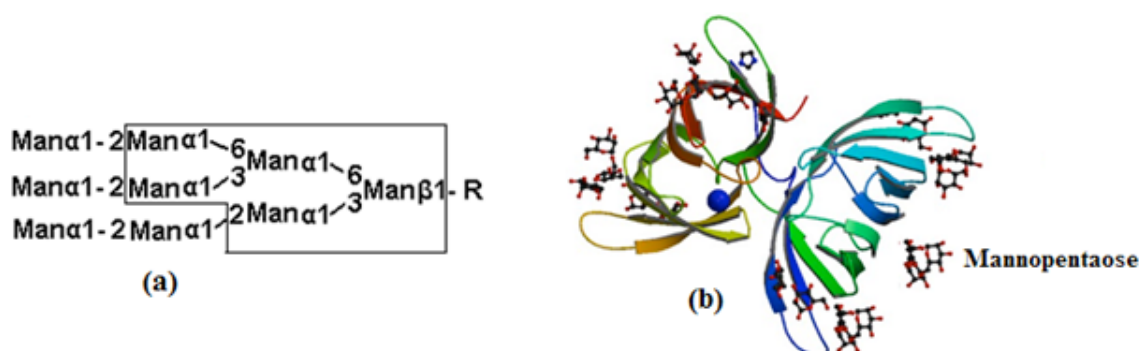

Figure 4. (a) Structure of Man9. Mannopentaose core structure (Manα1-6(Manα1-3)Manα1-6(Manα1-3)Man) was indicated in frame. Man: mannose. (b) Crystal structure of BOA lectin from *B. oklahomensis* bound to mannopentaose (Whitley *et al.,* 2013).

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.
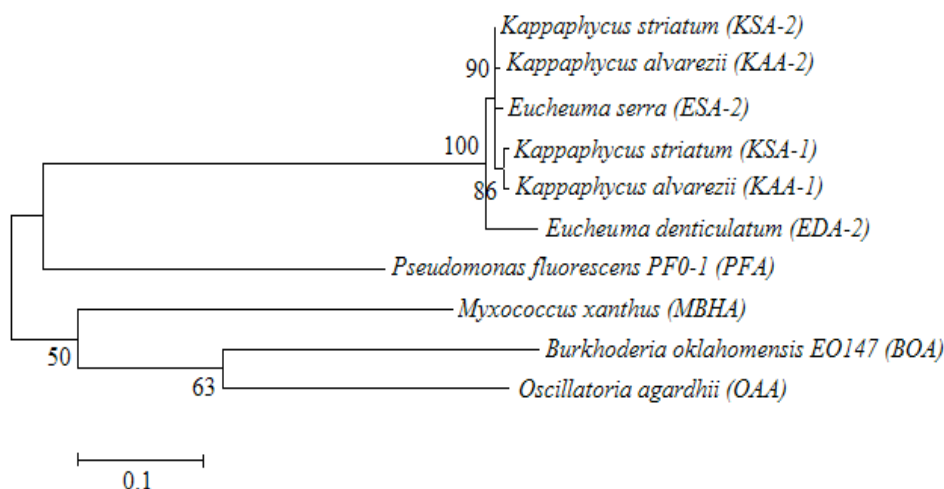
Figure 5. Phylogenetic analysis of the same lectin family in lower organisms. The phylogenetic tree was constructed from multiple alignments of amino acid sequences by Mega v. 6.0 software using Maximum likelihood and bootstrap methods with 1000 bootstrap replications (Tamura *et al.,* 2013). Scale bar represents 0.1 substitution per site. The following sequences were obtained from GenBank database: KSA-2 from *K. striatum* (LC057282); KAA-1 (LC007080) and KAA-2 (LC007081) from *K. alvarezii*; ESA-2 from *E. serra* (P84331); EDA-2 from *E. denticulatum* (LC057379); MBHA from *M. xanthus* (WP_011556980); BOA from *B. oklahomensis* EO147 (ZP_02360833); OAA from *O. agardhii* (P84330); PFA from *P. fluorescens* Pf0-1 (WP_011332171).

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 organinsms, 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 cultivated 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.

REFERENCES

Ask EI, Azanza RV (2002) Advances in cultivation technology of commercial eucheumatoid species: a review with suggestions for future research. *Aquaculture* 206: 257-277.

Balzarini J (2006) Inhibition of HIV entry by carbohydrate-binding proteins. *Antiviral Res* 71: 237-247.

Barrientos LG, O'Keefe BR, Bray M, Sanchez A, Gronenborn AM, Boyd MR (2003) Cyanovirin-N binds to the viral surface glycoprotein, GP1,2 and inhibits infectivity of Ebola virus. *Antiviral Res* 58: 47-56.

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci* 5: 174-182.

Chiba H, Inokoshi J, Okamoto M, Asanuma S, Matsuzaki K, Iwama M, Mizumoto K, Tanaka H, Oheda M, Fujita K, Nakashima H, Shinose M, Takahashi Y, Omura S (2001) Actinohivin, a novel anti-HIV protein from an actinomycete that inhibits syncytium formation: isolation, characterization, and biological activities. *Biochem Biophys Res Commun* 282: 595-601.

Chisholm JRM, Dauga C, Ageron E, Grimont PAD, Jaubert JM (1996) Roots in mixotrophic algae. *Nature* 381, 382.

Doty MS (1973) Farming the red seaweed, Eucheuma, for carrageenans. *Micronesia* 9: 59-73.

Fourçans A, De Oteyza TG, Wieland A, Solé A, Diestra E, Bleijswijk JV, Grimalt JO, Kühl M, Esteve I, Muyzer G, Caumette P, Duran R (2004) Characterization of functional bacterial groups in a hypersaline microbial mat community (Salins-de-Giraud, Camargue, France). *FEMS Microbiology Ecology* 51: 55-70.

Helle F, Wychowski C, Vu-Dac N, Gustafson KR, Voisset C, Dubuisson J (2006) Cyanovirin-N inhibits hepatitis C virus entry by binding to envelope protein glycans. *J Biol*

Chem 281: 25177-25183.

Hirayama M, Shibata H, Imamura K, Sakaguchi T, Hori K (2016) High-mannose specific lectin and its recombinants from a Carrageenophyta *Kappaphycus alvarezii* represent a potent anti-HIV activity through high-affinity binding to the viral envelope glycoprotein gp120. *Mar Biotechnol* 18: 144-160.

Hori K, Sato Y, Ito K, Fujiwara Y, Iwamoto Y, Makino H, Kawakubo A (2007) Strict specificity for high-mannose type *N*-glycans and primary structure of a red alga *Eucheuma serra* lectin. *Glycobiology* 17: 479-491.

Hung LD, Sato Y, Hori K (2011) High-mannose N-glycan-specific lectins from the red alga *Kappaphycus striatum* (Carrageenophyte). *Phytochemistry* 72: 855-861.

Hung LD, Hirayama M, Ly BM, Hori K (2015a) Purification, primary structure and biological activity of high-mannose *N*-glycan-specific lectin from the cultivated *Eucheuma denticulatum*. *J Appl Phycol* 27: 1657-1669.

Hung LD, Hirayama M, Ly BM, Hori K (2015b) Biological activity, cDNA cloning and primary structure of lectin KSA-2 from the cultivated red alga *Kappaphycus striatum* (Schmitz) Doty ex Silva. *Phytochemistry Letters* 14: 99-105.

Koharudin LMI, Furey W, Gronenborn AM (2011) Novel fold and carbohydrate specificity of from *Oscillatoria agardhii* the potent Anti-HIV cyanobacterial lectin. *J Biol Chem* 286: 1588-1597.

Koharudin LMI, Kollipara S, Aiken C, Gronenborn AM

(2012) Structural insights into the anti-HIV activity of the *Oscillatoria agardhii* agglutinin homolog lectin family. *J Biol Chem* 287: 33796-33811.

Koharudin LMI, Gronenborn AM (2014) Antiviral lectins as potential HIV microbicides. *Curr Opin Virol* 7: 95-100.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0". *Bioinformatics* 23: 2947-2948.

Moreira D, Le Guyader H, Philippe H (2000) The origin of

red algae and the evolution of chloroplasts. *Nature* 405: 69-72.

O'Keefe BR, Giomarelli B, Barnard DL, Shenoy SR, Chan PK, McMahon JB, Palmer KE, Barnett BW, Meyerholz DK, Wohlford-Lenane CL, McCray PB Jr (2010) Broad-spectrum in vitro activity and in vivo efficacy of the antiviral protein griffithsin against emerging viruses of the family Coronaviridae. *J Virol* 84: 2511-2521.

Okuno T, Shao H, Asada H, Shiraki K, Takahashi M, Yamanishi K (1992) Analysis of human herpesvirus 6 glycoproteins recognized by monoclonal antibody OHV1. *J Gen Virol* 73: 443-447.

Sato Y, Okuyama S, Hori K (2007) Primary structure and carbohydrate-binding specificity of a potent anti-HIV lectin isolated from the filamentous cyanobacterium, *Oscillatoria agardhii*. *J Biol Chem* 282: 11021-11029.

Sato Y, Morimoto K, Hirayama M, Hori K (2011) High mannose-specific lectin (KAA-2) from the red alga *Kappaphycus alvarezii* potently inhibits influenza virus infection in a strain-independent manner. *Biochem Biophys Res Commun* 405: 291-296.

Sato Y, Morimoto K, Kubo T, Yanagihara K, Seyama T (2012) High mannose-binding antiviral lectin PFL from *Pseudomonas fluorescens* Pf0-1 promotes cell death of gastric cancer cell MKN28 via interaction with α2-integrin. *PLos One* **7,** e45922.

Sharon N, Lis H (2003) Lectins, second ed. Kluwer Academic Publishers, The Netherlands.

Vigerust DJ, Shepherd VL (2007) Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* 15: 211-218.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30: 2725-2729.

Whitley MJ, Furey W, Kollipara S, Gronenborn AM (2013) *Burkholderia oklahomensis* agglutinin is a canonical twodomain OAA-family lectin: structures, carbohydrate binding, and anti-HIV activity. *FEBS J* 280: 2056-2067.

NHÂN DÒNG cDNA MÃ HÒA LECTIN KSA-1 TỪ RONG ĐỎ *KAPPAPHYCUS STRIATUM* ĐƯỢC NUÔI TRỒNG Ở VIỆT NAM

Lê Đình Hùng, Đinh Thành Trung

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

TÓM TẮT

Rong đỏ *Kappaphycus striatum* là loài rong thực phẩm có giá trị kinh tế cao và được nuôi trồng rộng rãi ở

Việt Nam. Rong nuôi trồng 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 chiết và đánh giá tính chất hóa sinh, bao gồm đặc tính liên kết high-mannose, trình tự acid amin ở đầu cuối N và khối lượng phân tử. Trong nghiên cứu này, chúng tôi nhân 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 bắt đầu bằng methionine. Chuỗi polypeptide suy diễn chứa 4 vùng lặp lại nối tiếp nhau, mỗi vùng bao gồm khoảng 67 acid amin và có giá trị tương đồng lẫn nhau là 43 %. Cấu trúc bậc một và số các gốc acid amin tương tác với cấu trúc lõi pentamannose trên mỗi vùng lập 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 lõi pentamannose trên mỗi vùng lập lại của họ lectin kháng virus trong sinh vật bậc thấp, bao gồm lectin từ vi khuẩn, *Burkholderia oklahomensis* EO147, *Myxococcus xanthus* và *Pseudomonas fluorescens* Pf0-1, vi khuẩn lam, *Oscillatoria agardhii* và rong đỏ, *Eucheuma serra, Eucheuma denticulatum, Kappaphycus alvarezii* và *K. striatum*. Phân tích cấu trúc bậc hai của KSA-1 cho thấy có 20 dãi β. Mỗi khu vực lập lại chứa 5 dãi β và trùng khớp với số dãi β của lectin BOA. Vì vậy, rong đỏ *K. striatum* hứa hẹn trở thành nguồn nguyên liệu tốt sử dụng trong hóa sinh và y sinh như là vật liệu để thu nhận lectin chức năng.

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