Tạp chí Công nghệ Sinh học **14**(2): 327-336, 2016

CLONING AND CHARACTERIZING cDNA SEQUENCES CODING HIGH-MANNOSE N-GLYCAN BINDING LECTINS FROM CULTIVATED RED ALGAE *EUCHEUMA DENTICULATUM* AND *KAPPAPHYCUS STRIATUM*

Le Dinh Hung[1], Makoto Hirayama[2], Kanji Hori[2]

[1]Nha Trang Institute of Technology Research and Application, Vietnam Academy of Science and Technology
[2]Graduate School of Biosphere Science, Hiroshima University. Kagamiyama 1-4-4, Higashi-Hiroshima 739 - 8528, Japan

Received: 14.11.2014
Accepted: 29.12.2015

SUMMARY

The red algae, *Eucheuma denticulatum* and *Kappaphycus striatum* have been widely cultivated in Vietnam as a source of carrageenophytes for industry. In the past, biochemical properties of lectins isolated from these algae has been characterized and evaluated extensively. However, gene coding for such lectins isn't studied yet. In this study, their full length cDNA is amfplified using cDNA ends (RACE) methods. Sequence analysis revealed that cDNA of EDA-2 from *E. denticulatum* consisted of 1,158 bp containing 103 bp of a 5'untranslated region, 248 bp of 3'untranslated region, and 807 bp of an open reading frame; and cDNA of KSA-2 from *K. striatum* consisted of 1174 bp containing 94 bp of the 5'-untranslated region, 273 bp of 3'untranslated region and 807 bp of the open reading frame. The cDNA of both EDA-2 and KSA-2 encoded for a polypeptide of 269 amino acids including an initiating methionine, but differed in sequences and molecular masses. The deduced amino acid sequences of EDA-2 and KSA-2 composed of four tandem repeated domains with about 67 amino acids each. The primary structure of EDA-2 and KSA-2 is highly similar to those of the high mannose N-glycan specific lectins including OAA from cyanobacterium, BOA, MBHA and PFA from bacteria, and ESA-2, KAA-1, KAA-2 from macro red algae, which showed strong anti-HIV and anti-influenza virus activities. These results indicate that these cultivated algae are becoming promising materials for production of anti-virus reagent or functional food that can prevent virus infection in future.

Keywords: *Red algae; Eucheuma denticulatum; Kappaphucus striatum; Lectin genes; EDA-2; KSA-2*

INTRODUCTION

Lectins are proteins of non-immunoglobulin nature, capable of recognition and reversible binding to carbohydrate moieties of complex glycocongugates. Lectins found across a wide variety of species in nature including prokaryotes, sea corals, algae, fungi, higher plants, invertebrates and vertebrates are involved in many biological processes such as host-pathogen interactions, cell-cell communication, induction of apoptosis, cancer metastasis and differentiation, targeting of cells as well as recognizing and binding carbohydrates (Sharon, Lis, 2003).

Anti-human immunodeficient virus (HIV) and/or anti-influenza virus lectins from bacteria, algae, fungi, and land plants share the common property in binding to high-mannose *N*-glycans, thereby 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). The anti-HIV lectins from cyanobacteria (blue-green algae) and eukaryotic macroalgae are presently marked as a new antiviral agent because their HIV-inhibiting activities are extremely strong compared to those from other biological groups (Ziolkowska, Wlodawer, 2006; Balzarini, 2007). One of lectins selected as a candidate for lectin-based microbicides is cyanovirin-N (CV-N) (O'Keefe *et al*., 2000; Boyd *et al*., 1997; Barrientos *et al*., 2006), that exhibits potent antiviral activity by blocking virus entry into the host target cells through specific and tight binding to Manα(1–2)Man-linked mannose substructures (Botos *et al*., 2002; Barrientos, Gronenborn, 2005; Barrientos *et al*., 2006; Shenoy *et al*., 2002). Notably, the target epitopes on Man-8/9 and binding modes for the different lectins are quite distinct (Williams *et al*., 2005; Ziolkowska *et al*.,

2006). Possible mannose-targeting lectins structurally, as well as their recognition epitopes, a novel antiviral lectin from the cyanobacterium *Oscillatoria agardhii* (named OAA standing for *O. agardhii* agglutinin) was investigated extensively (Koharudin *et al*., 2011; Koharudin, Gronenborn, 2011). Gene encoding for OAA homologous proteins (OAAHs) have recently been discovered in a number of other prokaryotic microorganisms, including cyanobacteria, proteobacteria, and chlorobacteria, as well as in eukaryotic marine red algae (Sato, Hori, 2009). Similar to OAA, OAAHs contain a sequence repeat of about 66 amino acids, with the number of repeats varying for different family members. For example, OAAH of *Pseudomonas fluorescens* and *Herpetosiphon aurantiacus* contain 133 residues and two sequence repeats, like OAA, whereas *Lyngbya* sp., *Burkholderia oklahomensis* EO147, *Stigmatella aurantiaca* DW4/3-1, *Myxococcus xanthus*, and *Eucheuma serra* homologs contain four sequence repeats over the length of 246–268 residues (Hori *et al*., 2007). In addition, several potent anti-HIV and anti-influenza virus lectins from marine macroalgae, such as griffithsin from the red alga *Griffithia* sp. (Mori *et al*., 2005), KAAs from the red alga *Kappaphycus alvarezii* (Sato *et al*., 2011a; Hirayama *et al.,* 2016) and BCA from the green alga *Boodlea coacta* (Sato *et al*., 2011b) were investigated in more details. Although these anti-HIV lectins share both binding specificity for high-mannose *N*-glycans and repeated domain structures, they differ to each another in amino acid sequences and recognizing branched mannoside structures, which may lead to subtle difference in the degree of inhibiting activities.

The algae *Kappaphycus alvarezii*, *K. striatum* and *Eucheuma denticulatum* are economically important species, which were extensively cultivated either for edible purposes or as a source of carrageenophyte for industry and have been introduced in more than 20 countries for mariculture purposes (Ask, Azanza, 2002). In Vietnam, *E. denticulatim* and *K. striatum* cultivation began in 2005 using algal seeds imported from Bohol, Philippines. Apart from carrageenan source, these cultivated algae may provide bioactive compounds for biochemical and medicinal uses. Lectins from algae of *Eucheuma* and *Kappaphycus* genera have been isolated and well characterized and showed the novel properties in carbohydrate binding specificity with high-mannose type *N*-glycans (Hori *et al*., 2007; Hung *et al*., 2009, 2011, 2015; Sato *et al*., 2011a; Hirayama *et al.,* 2016), which were distinct from those of other organisms. Therefore, they are promising sources for the development of new reagents using in biochemistry and biomedicine. In this study, genes coding lectins from these cultivated algae were isolated. The nucleotide sequences and their protein primary structures were investigated for further application of cultivated red algae, and carrageenophytes for human health purposes.

MATERIALS AND METHODS

Materials

The red algae *Eucheuma denticulatum* and *Kappaphycus striatum* were 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 RNA extraction. GeneRacer kit was obtained from Life Technologies (Invitrogen, USA). All other chemicals used in this study were of the highest purity available.

Extract and purification RNA and mRNA from *E. denticulatum* and *K. striatum* tissues

Total RNA of each *E. denticulatum* and *K. striatum* was extracted from 2 gram of the RNAlater-treated fresh algal tissues using the plant RNA isolation reagents (Invitrogen, USA). mRNA purification from the total RNA was performed using Oligotex ™-dT30 mRNA purification Kit (TaKaRa, Japan). RNA and mRNA quantities were determined on the NanoDrop® ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific). cDNAs were synthesized from 150 ng of mRNA using a GeneRacer kit (Invitrogen) according to the manufacturer's instruction.

Amplification of the 3′ and 5' cDNA ends (3' and 5'RACE) of EDA and KSA

The first polymerase chain reaction (PCR) for amplification of the cDNA 3'end (3'RACE) was performed with 8 aliquots of a 10 µL reaction mixture containing 1 µL of a 10 x Blend Taq buffer (Toyobo, Osaka, Japan), 2 pmol of each dNTP, 6 pmol of the GeneRacer_3'_Primer, and 2 pmol of the common-F1 primer, which was designed from the N-terminal amino acid sequence of lectins (Hung *et al*., 2011) (Table 1), 0.2 µL of a 10-fold diluted synthesized cDNA, and 0.25 units of Blend Taq DNA polymerase (Toyobo). The reactions for 8 aliquots were performed with a T Gradient Thermocycler (Biometra, Göttingen, Germany) under the following conditions: denaturation at 94°C

for 5 min, follows by 30 cycles consisting of denaturation at 94°C for 30 s, annealing at 8 different temperatures of 50–64°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. The PCR products in 8 aliquots were pooled, diluted to 100-fold, and then used as a template for nested PCR. The nested PCR was performed with a 50 µL reaction mixture containing 5 µL of a 10 x Blend Taq buffer, 10 pmol of each dNTP, 2 pmol of GeneRacer_3'_Nested_Primer, and 50 pmol of the degenerated primer 3'RACE_d_F2 which was designed from the conserved sequence among the high-mannose specific lectin family including ESA-2 (Hori *et al*., 2007) and OAA (Sato *et al*., 2007), 1 µL of the dilution of the first PCR products, and 1.25 units of Blend Taq DNA polymerase. The nested PCR reaction was carried out by the same method of the first PCR, except for the annealing temperature of 60^{o}C used. Nested PCR products were subcloned into pGEM-T Easy vector (Promega, WI, USA) and transformed into *Escherichia coli* DH5α competent cells. Plasmids from *E. coli* clones formed in selection culture mediu,m were purified using HiYield Plasmid Mini kit (RBC Bioscience, New Taipei City, Taiwan) according to the manufacturer's instruction. DNA sequencing was performed using BigDye Terminator Cycle Sequencing kit ver. 3.1 and ABI 3130xl genetic analyser (Applied Biosystems).

The first PCR of 5'RACE was performed in the same way as 3'RACE as described above, except that GeneRacer_5'_Primer and the primer EDA_3'end_R or KSA_3'end_R, which designed from the 3' terminal sequence of each EDA and KSA cDNA obtained by 3'RACE (Table 1). The nested PCR was performed by the same method, except for using a 100-fold dilution of the first PCR products as a template, and the GeneRacer_5'_Nested_ Primer and the primer EDA_5'RACE_R1 or KSA5'RACE_R1 designated from the sequence of EDA and KSA cDNA obtained by 3'RACE, as the primer pair (Table 1). Subcloning and DNA sequencing were performed as described above.

Table 1. Primer sequence used for the cDNA cloning of EDA and KSA.

Primer	Sequence (from 5' to 3')
EDA_common F1	AGAACCAGTGGGGAGGATCT
EDA_3'RACE_d_F1	AYCAITAYAAYGTIGARAAYCARTGGGG [b]
EDA_5'RACE_R1	GCAATGTTCTTGGTAGCAGC
EDA_5'end_F	AGAAATTCAACACCACAACT
EDA_3'end_R	CTGCACAAAACGTAACAATATCTAT
KSA_ common F1	AGAACCAGTGGGGAGGATCT
KSA_ 3'RACE d F2	AYCAITAYAAYGTIGARAAYCARTGGGG [b]
KSA_ 5'RACE R1	AYTGRTTYTCIACRTTRTAIT [b]
KSA_ 5'RACE R2	ATIGGICCYTCICCYTTRTAYTGC [b]
KSA_ 5'End F	ACCACCAGCACCGTGCTACT
KSA_ 3'End R	GGGTATAGATAATAGGAATGTCC
GeneRacer 3′ Primer [a]	GCTGTCAACGATACGCTACGTAACG
GeneRacer 3′ Nested Primer [a]	CGCTACGTAACGGCATGACAGTG
GeneRacer 5′ Primer [a]	CGACTGGAGCACGAGGACACTGA
GeneRacer 5′ Nested Primer [a]	GGACACTGACATGGACTGAAGGAGTA

Note: [a] These primers were inferred from the GenRacer kit (Invitrogen). [b] Key to symbols of the degenerated nucleotides: represents inosine; Y represents C and T; R represents A and G.

To verify the sequence accuracy, full-length cDNA of EDA or KSA was further amplified using the high-fidelity DNA polymerase KOD Plus Neo (Toyobo), the primer pair of EDA_3'end_R and EDA_5'end_F, or KSA_3'end_R and KSA_5'end_F, which designated from the 5' terminal sequence of EDA or KSA cDNA (Table 1), and a 10-fold diluted synthesized cDNAs of *E. denticulatum* or *K. striatum* as a template. Before subcloning, the

PCR products were treated with a 10 x A-attachment mix (Toyobo) according to the manufacturer's instruction. 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).

RESULTS AND DISCUSSION

Extract and purification RNA and mRNA from *E. denticulatum* and *K. striatum* tissues

From 2 grams of the fresh tissues stored in RNAlater solution, total RNA and mRNA were extracted, efficiently purified and are summarized in table 2. The A260/280 ratio value of 1.8 - 2.0 indicates that the RNA or mRNA is pure.

Table 2. Total RNA and mRNA yields were purified from 2 gram of the RNAlater-treated fresh algal tissues.

Species	ng/µL RNA	Total RNA (ng)	A260/280 ratio	ng/µL mRNA	Total mRNA (ng)	A260/280 ratio
E. denticulatum	118.6	47.76	2.18	3.7	177.0	2.2
K. striatum	159.9	63.960	2.08	3.6	288.0	2.1

Primary structure of lectins EDA and KSA and comparison analysis to other proteins

cDNA cloning of lectin EDA and KSA were performed by 5' and 3'RACEs as described in the Materials and Methods (Figure 1). The obtained full-length cDNA of EDA lectin consisted of 1,158 bp containing 103 bp of a 5'untranslated region (5'UTR), 807 bp of an open reading frame (ORF), and 248 bp of 3'UTR (Figure 2). ORF coded a polypeptide of 269 amino acids including the initiating methionine amino acid. The calculated molecular mass of the deduce amino acid sequence from EDA cDNA was 27.834.19 Da, similar to that of EDA-2 (27,834.2 Da) determined by ESI-MS methods (Hung *et al*., 2015). Thus we can concluded that the isolated cDNA encoding EDA-2.

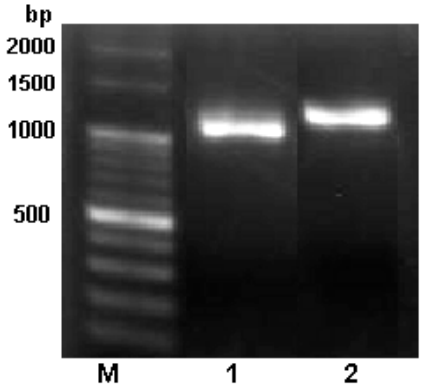

Figure 1. PCR products electrophoresis separated on 1% agarose. M: 100 bp DNA Ladder (Toyobo, Japan). 1: 1158 bp fragment of EDA. 2: 1174 bp fragment of of KSA.

```
AGAAATTCAACACCACAACTTTCAGCCTACATATTCGACCCACCCTCAACGACTCTACTCC  61
ATCAGACTTTTACCTTCTCATCTTCATCGCAAAAAGTAAGCTATGGGACGCTACACAGTT  121
                                              M  G  R  Y  T  V   6
CAGAATCAATGGGGTGGCTCTTCCGCGCCTTGGAACGATGCCGGATTGTGCATCCTTGGT  181
 Q  N  Q  W  G  G  S  S  A  P  W  N  D  A  G  L  C  I  L  G   26
AGCCGTGGTAACCAAAATGTGATCGCTGTTGATGTCACTTCTAGCGATGGAGGCGCTAAT  241
 S  R  G  N  Q  N  V  I  A  V  D  V  T  S  S  D  G  G  A  N   46
CTCGGCGGCACAATGACCTACTCGGGTGAGGGCCCCATCGGATTCAAGGGTGCCCGCCGC  301
 L  G  G  T  M  T  Y  S  G  E  G  P  I  G  F  K  G  A  R  R   66
GGTGAGTCAAATGTATATGACGTTGAGAATCAGTGGGGAGGATCTTCCGCTCCCTGGCAT  361
 G  E  S  N  V  Y  D  V  E  N  Q  W  G  G  S  S  A  P  W  H   86
GCTGGAGGTCAGTTCGTTATCGGCTCTCGCTCCGGGCAAGGGGTGACTGCTTTGAGCGTC  421
 A  G  G  Q  F  V  I  G  S  R  S  G  Q  G  V  T  A  L  S  V  106
ACATCTTCCGATGGCGGGAAGACATTGACAGGCACCATGACCTACGAAAGGGAGGGCCCT  481
 T  S  S  D  G  G  K  T  L  T  G  T  M  T  Y  E  R  E  G  P  126
ATTGGATTCAAGGGCACGCAGTCCGGAGGAGACACGTACAACGTCGAGAATCAGTGGGGT  541
 I  G  F  K  G  T  Q  S  G  G  D  T  Y  N  V  E  N  Q  W  G  146
GGATCATCTGCTCCCTGGAACAAAGCAGGCATCTGGGCGCTTGGGGATCGCAATGGTCAG  601
 G  S  S  A  P  W  N  K  A  G  I  W  A  L  G  D  R  N  G  Q  166
GCAATGATTGCGATGGACGTATCATCCTCAGATGGCGGGCAGACATTGGAAGGAACGATG  661
 A  M  I  A  M  D  V  S  S  S  D  G  G  Q  T  L  E  G  T  M  186
CAGTACAAGGGAGAGGGGCCGATTGGATTCCGGGGCAAGTTGAGCGGTGCTAACAACTAC  721
 Q  Y  K  G  E  G  P  I  G  F  R  G  K  L  S  G  A  N  N  Y  206
GCCGTGGAGAATCAGTGGGGCGGTTCCTCTGCTCCCTGGAATAAGGCAGGGGATTGGCTG  781
 A  V  E  N  Q  W  G  G  S  S  A  P  W  N  K  A  G  D  W  L  226
ATCGGAGACCGCTACAACCAAAATATCACTGCAGTCAAAGTGTCATCTGACAATGACGGA  841
 I  G  D  R  Y  N  Q  N  I  T  A  V  K  V  S  S  D  N  D  G  246
AAGAACCTTGACGGGACATGCACGTACGAGCGTGAGGGTCCGATTGGCTTCAAGGGGGTC  901
 K  N  L  D  G  T  C  T  Y  E  R  E  G  P  I  G  F  K  G  V  266
GCGACTTCTTAGGAAGGATCGGCTTATAGTTCTTAAGTGGTAGACACTTAATCATAGATT  961
 A  T  S  *                                                    269
TCGTTCTCCTTCTTATCCCGTTATGGAATGTGTGCTGCTACCAAGAACATTGCTCTGTAG  1021
GTTCCTGAATGAATGATGACTCGGACATGTTGGTGGATGTGTCCTTCTAGAATGCTTACG  1081
GATCTAGTCTGATCCAAAGATTTCCTCTAGAATGAAGTCATTAAAATCCCTTATAGATAT  1141
TGTTACGTTTTGTGCAG                                             1158
```

Figure 2. Nucleotide and deduced amino acid sequences of EDA-2, a lectin from *E. denticulatum* (registered as accession number LC057379 in GenBank). The stop codon TAG is shown as an asterisk. The *italicized* and *nonitalicized* numbers represent the positions of nucleotides and amino acids, respectively.

The full-length cDNA of KSA lectin consisted of 1174 bp containing 94 bp of 5'-untranslated region (UTR), 273 bp of 3' UTR, and 807 bp of the open reading frame (ORF) and encoded a polypeptide of 269 amino acids including an initiating methionine (Figure 3). The molecular mass of the deduce amino acid sequence from KSA lectin cDNA was calculated as 28,021.4 Da, which was similar to that of KSA-2 determined by ESI-MS method (28,021.3 Da) (Hung *et al*., 2011). The 20 N-terminal amino acid sequence of both EDA-2 and KSA-2, determined by Edman degradation (Hung *et al*., 2011) was also found in the deduced amino acid sequences from isolated cDNAs. The deduced amino acid sequences of both EDA-2 and KSA-2 have four tandem repeated domains and each domain consists of 67 amino acids. Four domains of EDA-2 shared 47 % sequence identity (Figure 4a), whereas those of KSA-2 shared 44.8 % sequence identity (Figure 4b). Clusters of identical amino acids among the four repeated domains of both lectins were located on both N- and C-terminal regions of each domain. The seven other homologous amino acid sequences were found in database, including one from the cyanobacterium, *Oscillatoria*

agardhii (OAA) (132 aa) (P84330), three from bacteria, *Burkholderia oklahomensis* EO147 (BOA) (276 aa) (ZP_02360833), *Pseudomonas fluorescens* Pf0-1 (PFA) (133 aa) (YP_346241) and *Myxococcus xanthus* (MBHA) (267 aa) (M13831), and three lectins from marine red algae, *Eucheuma serra* (ESA-2) (268 aa) (P84331), *Kappaphycus alvarezii* (KAA-1) (267 aa) (LC007080) and KAA-2 (268 aa) (LC007081). All of them contain four tandem repeated homologous domains of about 67 amino acids, except *O. agardhii* and *P. fluorescens*, which possess only two tandem repeated homologous domains (Figure 5). The similarity degrees between KSA-2, ESA-2, KAA-1, KAA-2, MBHA and BOA amino acid sequences with EDA-2 were 95.2, 95.2, 94.0, 94.8, 59.0 and 56.8 %, respectively, whereas sequence of OAA from *O. agardhii* and PFA from *P. fluorescens* (each 132 residues in the N-terminal portions) showed a considerable similarity to EDA-2 with 61.4 and 63.6 %, respectively.

```
ACCACCAGCACCGTGCTACTCGCTCGACATTAGATCCGCTACATCAACCTCATCATCTTTT  61
CTATCTCCTCACCTACATTCGCAACAGTAAGACATGGGACGTTACACAGTTCAGAACCAA  121
                                  M  G  R  Y  T  V  Q  N  Q    9
TGGGGAGGATCCTCTGCCCCCTGGAATGATGCCGGCTTGTGGATCCTTGGCAGTCGTGGC  181
 W  G  G  S  S  A  P  W  N  D  A  G  L  W  I  L  G  S  R  G   29
AACCAGAACGTGATGGCTATTGATGTCAACTCGAGCGATGGAGGTGCCAACTTGAACGGT  241
 N  Q  N  V  M  A  I  D  V  N  S  S  D  G  G  A  N  L  N  G   49
ACGATGACCTACTCCGGTGAAGGCCCGATTGGATTCAAGGGGGCACGCCGTGGTGAGTCA  301
 T  M  T  Y  S  G  E  G  P  I  G  F  K  G  A  R  R  G  E  S   69
AACGTCTACGATGTAGAGAACCAGTGGGGAGGATCTTCCGCCCCCTGGCACGCTGGAGGC  361
 N  V  Y  D  V  E  N  Q  W  G  G  S  S  A  P  W  H  A  G  G   89
CAATTCGTTATCGGTTCCAGGTCTGGACAGGGAGTGCTTGCCGTGAACATCACTTCTTCA  421
 Q  F  V  I  G  S  R  S  G  Q  G  V  L  A  V  N  I  T  S  S  109
GATGGCGGAAAGACATTGACTGGCACCATGACCTATGAGAGAGAAGGTCCAATTGGATTC  481
 D  G  G  K  T  L  T  G  T  M  T  Y  E  R  E  G  P  I  G  F  129
AAGGGTACTCAGTCTGGCGGTGACACTTACAACGTAGAGAACCAGTGGGGTGGGTCCTCG  541
 K  G  T  Q  S  G  G  D  T  Y  N  V  E  N  Q  W  G  G  S  S  149
GCACCGTGGAACAAGGCAGGCATTTGGGCGCTCGGGGATCGCAGTGGGCAAGCAATGATT  601
 A  P  W  N  K  A  G  I  W  A  L  G  D  R  S  G  Q  A  M  I  169
GCTATGGATGTTTCATCCTCGGATGGTGGAAAGACTCTAGAAGGGACGATGCAGTACAAA  661
 A  M  D  V  S  S  S  D  G  G  K  T  L  E  G  T  M  Q  Y  K  189
GGAGAGGGACCGATTGGATTCCGTGGCAAGCTGAGCGGTGCCAACAATTACAGCGTCGAG  721
 G  E  G  P  I  G  F  R  G  K  L  S  G  A  N  N  Y  S  V  E  209
AATCAGTGGGGCGGATCCTCTGCTCCTTGGAACAAGGCTGGAGACTGGTTGATTGGAGAC  781
 N  Q  W  G  G  S  S  A  P  W  N  K  A  G  D  W  L  I  G  D  229
CGCCACAACCAGAACATCACTGCTGTGAAGGTGTCATCTGACAATGACGGAAAGAACCTC  841
 R  H  N  Q  N  I  T  A  V  K  V  S  S  D  N  D  G  K  N  L  249
GACGGCACTTGCACGTACGAGCGTGAAGGTCCGATTGGTTTCAAGGGGGTTGCGACTTCC  901
 D  G  T  C  T  Y  E  R  E  G  P  I  G  F  K  G  V  A  T  S  269
TAGACGGTCACATCGCACGTGATGTTATCAAAGTTAGTTACCCGACCATAGATTTGTTTC  961
 *
TCCTTTACCCCTTTATGGAATGTTAACGAGTCCCGCAATTGCTTTCCTCGTAGTGAAATC  1021
GCGGCTGGTCTGTGAAGGCTTCCAGTGCGGGAGAAGAATCCAATATGCTGGTATTTAGAG  1081
CAAAACCACATCAGCAAAAGCGGATCTTTCTATTTTCTCTTCTGCTAGAAATCATCACTA  1141
AAGGCCTCAAGGACATTCCTATTATCTATACCC                              1174
```

Figure 3. Nucleotide and deduced amino acid sequences of KSA-2, a lectin from *K. striatum* (registered as accession number LC057282 in GenBank). The stop codon TAG is shown as an asterisk. The *italicized* and *nonitalicized* numbers represent the positions of nucleotides and amino acids, respectively. .

```
EDA-2  (a)
GRYTVQNQWGGSSAPWNDAGLCILGSRGNQNVIAVDVTSSDGGANLGGTMTYSGEGPIGFKGARRGES  (1-68)
NVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVTALSVTSSDGGKTLTGTMTYEREGPIGFKGTQSGG  (69-135)
DTYNVENQWGGSSAPWNKAGIWALGDRNGQAMIAMDVSSSDGGQTLEGTMQYKGEGPIGFRGKLSGA  (136-202)
NNYAVENQWGGSSAPWNKAGDWLIGDRYNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS   (203-268)
KSA-2  (b)
GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAIDVNSSDGGANLNGTMTYSGEGPIGFKGARRGES  (1-68)
NVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGKTLTGTMTYEREGPIGFKGTQSGG  (69-135)
DTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSGA  (136-202)
NNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS   (203-268)
```

Figure 4. Comparison of the four tandem-repeated domains of lectin EDA-2 (a) and KSA-2 (b), respectively. Identical amino acids among the four repeated domains in the lectin molecules are indicated in red color.

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

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

EDA-2   135 GDTYNVENQWGGSSAPWNKAGIWALGDRNGQAMIAMDVSSSDGGQTLEGTMQYKGEGPIGFRGKLSG  201
KSA-2   135 GDTYNVENQWGGSSAPWNKAGIWALGDRSGQAMIAMDVSSSDGGKTLEGTMQYKGEGPIGFRGKLSG  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

EDA-2   202 ANNYAVENQWGGSSAPWNKAGDWLIGDRYNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  268
KSA-2   202 ANNYSVENQWGGSSAPWNKAGDWLIGDRHNQNITAVKVSSDNDGKNLDGTCTYEREGPIGFKGVATS-  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 5. Multiple alignments of EDA-2 and KSA-2 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).

The deduced amino acid sequence of lectins EDA-2 and KSA-2 shares similar structure and carbohydrate specificity. These lectins presents in lower organisms at various taxonomies and living

environments such as ESA-2, KAA-1 and KAA-2 from marine red algae, MBHA from a soil bacterium, OAA from a fresh water cyanobacterium, BOA and PFA from bacteria. The lectins in this family are commonly monomeric proteins composed of either two or four tandem repeats of homologous domains of about 67 conserved amino acids, depending on the organism. The significance difference in biological and evolutionary of the number of repeated domains is currently unknown. However, their sequence similarities suggest that these proteins may also share lectin domains that have a binding affinity for high-mannose *N*-glycans, as ESA-2 (Hori *et al*., 2007), KSA-2 (Hung *et al*., 2011), EDA-2 (Hung *et al.,* 2015), KAA-2 (Sato *et al.,* 2011a; Hirayama *et al.,* 2016), OAA (Sato *et al*., 2007), BOA (Whitley *et al*., 2013), PFA and MBHA (Sato *et al.,* 2012; Koharudin *et al*., 2012), which showed strong anti-HIV and anti-influenza virus activities. From sequence comparison, it is also speculated that the highly conserved segments in both N-terminal and C-terminal regions may be involved in the carbohydrate binding. Although the high-mannose *N*-glycan recognition profile of EDA-2 (Hung *et al*., 2015) and KSA-2 (Hung *et al*., 2011) was slightly differed from those of the high-mannose binding cyanobacterial lectins such as CV-N (Botos *et al*., 2002), MVL (Williams *et al*., 2005) and SVN (Bokesch *et al*., 2003), EDA-2 and KSA-2 has no sequence similarity with these cyanobacterial lectins, except for OAA (Sato *et al*., 2007). The strict specificity for high-mannose oligosaccharides with very high binding affinities seems to be a common feature among lectins from lower organisms such as cyanobacteria and algae.

CONCLUSION

cDNAs encoding for lectins EDA-2 and KSA-2 were successful isolated from *E. denticulatum* and *K. striatum* algae cultivated in Vietnam.. Analysis of deduced amino acid sequences of EDA-2 and KSA-2 DNA indicated that the tandem repeated homologous domains in their amino acid sequence resembled those of lectins presenting in lower organisms, which showed strong anti-HIV and anti-influenza virus activities. Therefore, these algae might be promising materials for production of anti-virus reagent or functional food that can prevent virus infection in future. Clarifying the structural basis of the recognition mode of the carrageenophyte lectins should contribute to demonstrate the biological function(s) as well as application of this lectin group.

Acknowledgements: *This research was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 106.05-2011.35 and 106-YS.06-2015.16.*

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.

Balzarini J (2007) Carbohydrate-binding agents: a potential future cornerstone for the chemotherapy of enveloped viruses? *Antivir Chem Chemother* 18: 1-11.

Barrientos LG, Gronenborn AM (2005) The highly specific carbohydrate-binding protein cyanovirin-N: structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini Rev Med Chem* 5: 21-31.

Barrientos LG, Matei E, Lasala F, Delgado R, Gronenborn AM (2006) Dissecting carbohydrate-cyanovirin-N binding by structureguided mutagenesis: functional implications for viral entry inhibition. *Protein Eng Des Sel* 19: 525-535.

Bokesch HR, O'Keefe BR, McKee TC, Pannell LK, Patterson GM, Gardella RS, Sowder RC, Turpin J, Watson K, Buckheit RW, Boyd MR (2003) A potent novel anti-HIV protein from the cultured cyanobacterium *Scytonema varium*. *Biochemistry* 42:2578-2584.

Botos I, O'Keefe BR, Shenoy SR, Cartner LK, Ratner DM, Seeberger PH, Boyd MR, Wlodawer A (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J Biol Chem* 277: 34336-34342.

Boyd MR, Gustafson KR, Mcmahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera MI, Laurencot CM, Currens MJ, Cardellina IIJH, Buckheit JrRW, Nara PL, Pannell LK, Sowder IIRC, Henderson LE (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother* 41: 1521-1530.

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 T, Shibata H, Hori K (2009) Biochemical comparison of lectins among three different color strains of the red alga *Kappaphycus alvarezii*. *Fish Sci* 75: 723-730.

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 (2015) Purification, primary structure and biological activity of high-mannose *N*-glycan-specific lectin from the cultivated *Eucheuma denticulatum*. *J Appl Phycol* 27: 1657-1669.

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

Koharudin LM, Gronenborn AM (2011) Structural basis of the anti-HIV activity of the cyanobacterial *Oscillatoria agardhii* agglutinin. *Structure* 19: 1170-1181.

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.

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.

Mori T, O'Keefe BR, Sowder RC, Bringans S, Gardella R, Berg S, Cochran P, Turpin JA, Buckheit RWJr, McMahon JB, Boyd MR (2005) Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga *Griffithsia* sp. *J Biol Chem* 280: 9345-9353.

O'Keefe BR, Shenoy SR, Xie D, Zhang W, Muschik JM, Currens MJ, Chaiken I, Boyd MR (2000) Analysis of the interaction between the HIV-inactivating protein cyanovirin-N and soluble forms of the envelope glycoproteins gp120 and gp41. *Mol Pharmacol* 58: 982-992.

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 T, Hori K (2009) Cloning, expression, and characterization of a novel anti-HIV lectin from the cultured cyanobacterium, *Oscillatoria agardhii*. *Fish Sci* 75: 743-753.

Sato Y, Morimoto K, Hirayama M, Hori K (2011a) 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, Hirayama M, Morimoto K, Yamamoto N, Okuyama S, Hori K (2011b) High mannose-binding lectin with preference for the cluster of α1–2-mannose from the green alga *Boodlea coacta* is a potent entry inhibitor of HIV-1 and influenza viruses. *J Biol Chem* 286: 19446-19458.

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.

Shenoy SR, Barrientos LG, Ratner DM, O'Keefe BR, Seeberger PH, Gronenborn AM, Boyd MR (2002) Multisite and multivalent binding between cyanovirin-N and branched oligomannosides: calorimetric and NMR characterization. *Chem Biol* 9: 1109-1118.

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.

Williams DCJr, Lee JY, Cai M, Bewley CA, Clore GM (2005) Crystal structures of the HIV-1 inhibitory cyanobacterial protein MVL free and bound to $Man_3GlcNAc_2$: structural basis for specificity and high affinity binding to the core pentasaccharide from *N*-linked oligomannoside. *J Biol Chem* 280: 29269-29276.

Ziólkowska NE, O'Keefe BR, Mori T, Zhu C, Giomarelli B, Vojdani F, Palmer KE, McMahon JB, Wlodawer A (2006) Domains wapped structure of the potent antiviral protein Griffithsin and its mode of carbohydrate binding. *Structure* 14: 1127-1135.

Ziólkowska NE, Wlodawer A (2006) Structural studies of algal lectins with anti-HIV activity. *Acta Biochim Pol* 53: 617-626.

NHÂN DÒNG VÀ NGHIÊN CỨU ĐẶC ĐIỂM CÁC cDNA MÃ HÓA CÁC LECTIN LIÊN KẾT N-GLYCAN DẠNG MANNOSE CAO TỪ HAI LOÀI TẢO ĐỎ *EUCHEUMA DENTICULATUM* VÀ *KAPPAPHYCUS STRIATUM*

Lê Đình Hùng[1,✉], Makoto Hirayama [2], Kanji Hori[2]

[1]*Viện Nghiên cứu và Ứng dụng công nghệ Nha Trang, Viện Hàn lâm Khoa học và Công nghệ Việt Nam*
[2]*Graduate School of Biosphere Science, Hiroshima University. Kagamiyama 1-4-4, Higashi–Hiroshima 739 - 8528, Japan*

TÓM TẮT

Tảo đỏ, *Eucheuma denticulatum* and *Kappaphycus striatum* đang được nuôi trồng rộng rãi ở Việt nam cũng như là một nguồn nguyên liệu để sản xuất carrageenan. Lectin từ mẫu tảo này đã được tinh chế và xác định các tính chất hóa sinh. Tuy nhiên, gen mã hóa lectin của các mẫu tảo hầu như chưa được nghiên cứu. Trong nghiên cứu này, trình tự cDNA chứa chiều dài hoàn chỉnh của gen mã hóa các lectin của carrageenophyte đã được phân lập sử dụng phương pháp khuyếc đại nhanh các đầu tận cùng cDNA (RACE) Phân tích trình tự toàn bộ. cDNA của EDA-2 và KSA-2 cho thấy các gen này đều mã hóa cho một polypeptide chứa 269 amino acid bao gồm trình tự khởi đầu mã hóa cho methionine, nhưng lại khác nhau về trong trình tự amino acid và trọng lượng phân tử. cDNA of EDA-2 có kích thước 1158 bp mang 103 bp của đoạn 5' không dịch mã, 248 bp của đoạn 3' không dịch mã và 807 bp của một khung đọc mở. cDNA of KSA-2 có kích thước 1174 bp bao gồm 94 bp của đoạn 5' không dịch mã, 273 bp của đoạn 3' không dịch mã và một khung đọc mở 807 bp. Trình tự amino acid suy diễn từ cDNA của cả hai lectin nêu trên đều chứa 4 vùng lặp lại nối tiếp nhau có khoảng 67 amino acid. Cấu trúc bậc 1 của EDA-2 và KSA-2 có sự tương đồng cao với các cấu trúc của lectin liên kết N-glycan dạng mannose cao bao gồm OAA từ vi tảo, BOA, MBHA và PFA từ vi khuẩn, và ESA-2, KAA-1 and KAA-2 từ tảo đỏ, các lectin này đã được chứng minh là có hoạt tính kháng virus HIV và virus cúm mạnh. Các kết quả nghiên cứu thu được đã cho thấy các mẫu tảo nuôi trồng có thể trở thành thuốc kháng virus mới hoặc làm thực phẩm chức năng tăng cường ngăn chặn sự lây nhiễm của virus trong tương lai.

Từ khóa:*., Tảo đỏ, Eucheuma denticulatum, Kappaphycus striatum, gen mã hóa lectin EDA-2; KSA-2*

✉*Author for correspondence:* Tel: *+84-583521133*; Fax: *+84-583521847*; E-mail: *ledinhhungims@yahoo.co.uk*