ARTICLE

# Partial gene sequence and characterization of EDEN-1, a major water-soluble protein from the economically important alga *Eucheuma denticulatum*

**By Paulina B. Suarez-Aspilla[1,2], Mailyn M. Terrado[1], Maria Cristina A. Dancel[3], Giuseppe C. Zuccarello[4], Nina Rosario L. Rojas[1]**

[1] Department of Chemistry, School of Science and Engineering, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines 1108
[2] Department of Chemistry, Silliman University, Dumaguete City, Philippines
[3] Mass Spectrometry Facility, Department of Chemistry, University of Florida, Gainesville, FL 32611-7200, U. S. A.
[4] School of Biological Sciences, Victoria University of Wellington, PO Box 600, Wellington, 6140 New Zealand

*Eucheuma denticulatum* is a major aquaculture product worldwide as a source for iota-carrageenan, but its protein chemistry is not well-known. This study reports the isolation, characterization, and partial gene and amino acid sequences of a major water-soluble protein from *E. denticulatum*, which we called EDEN-1. The protein was purified by ion-exchange chromatography on a strong anion exchanger. It is compact and monomeric by size exclusion chromatography, and has an intact molecular mass of 27.83 kDa by mass spectrometry. Its isoelectric point is 4.7, which is consistent with its amino acid composition. It is not glycosylated, based on periodic acid-Schiff staining. Peptide sequencing by tandem mass spectrometry of tryptic fragments yielded 5 short internal peptide sequences. These peptide sequences were back-translated to design degenerate primers to obtain a partial cDNA sequence for EDEN-1. The results of *de novo* peptide sequencing and cDNA sequencing were combined and verified with peptide mass fingerprinting of tryptic digests to yield partial gene and amino acid sequences of EDEN-1. This sequence accounts for 13,414 Da, or 48.2% of the intact mass of the protein. The amino acid sequence of EDEN-1 was found to be homologous to lectin ESA-2 from *E. serra*. This is the first known report of a cDNA sequence for this family of proteins. Consistent with its homology to lectin-like proteins, EDEN-1 is mainly beta-sheet based on its circular dichroism spectrum. EDEN-1 also shows weak glucoside hydrolase activity, suggesting that it may bind to carbohydrate moieties despite a lack of appreciable hemagglutinating activity.

**Email Address: nrojas@ateneo.edu**
Submitted:February 15, 2012
Revised: March 28, 2012
Accepted: April 1, 2012
Published: June 15, 2012
Editor-in-charge: Eduardo A. Padlan

**KEYWORDS**

Protein biochemistry; carrageenan-producing seaweed; red alga; cDNA sequence; peptide sequence; glycosidase activity

## INTRODUCTION

*Eucheuma denticulatum* (N.L. Burman) Collins and Hervey (also known as *Eucheuma spinosum*) is one of the most extensively farmed marine algae in the Philippines. Traditionally, these seaweeds were used as food and harvested from the wild, but today much of the supply comes from aquaculture for the production of iota-carrageenan. According to 2008 data, worldwide aquatic plant production from aquaculture was 15.8 million tonnes, with estimated worth of US$7.4 billion (FAO Fisheries and Aquaculture Department 2010). Of these aquatic plants, seaweeds accounted for 99.6% by quantity and 99.3% in value. Indonesia overtook the Philippines in 2007 as the world's largest producer next to China, but the Philippines still accounted for 10.6% of world seaweed production in 2008. *Kappaphycus alvarezii* and *Eucheuma* varieties together accounted for 3.8 million tonnes of cultured seaweed in the world in 2008 (FAO Fisheries and Aquaculture Department 2010).

*Kappaphycus alvarezii* and *Eucheuma* varieties are valuable as sources of carrageenan. *E. denticulatum* is the major source of iota-carrageenan, and the polysaccharide fraction of this alga is well-characterized (Trono 2005). The proximate composition content of *E. denticulatum* on a dry basis yields mostly ash (approximately 45%) and the remainder consists of ~25% carbohydrate, ~6% protein, ~23% fiber, and ~2% crude lipid (Msuya and Neori 2002; McDermid and Stuercke 2003).

There is not much known about specific proteins found in *Eucheuma*. A survey of protein sequence and function database UniProt (http://www.uniprot.org) shows that most of the protein sequence entries under *Eucheuma* species are derived from genes used for taxonomy, such as the plastid-encoded ribulose-1,3-bisphosphate carboxylase small and large subunits (*rbc*S and *rbc*L, respectively) and mitochondria-encoded cytochrome oxidase 1 and 2 (*cox*1 and *cox*2) (UniProt Consortium 2011; Jain et al. 2009).

Additional protein information can be inferred from mRNA data through expressed sequence tags, or ESTs. ESTs are short sequences derived from mRNA molecules expressed in the tissues, and thus represent the proteins expressed (Boguski et al.

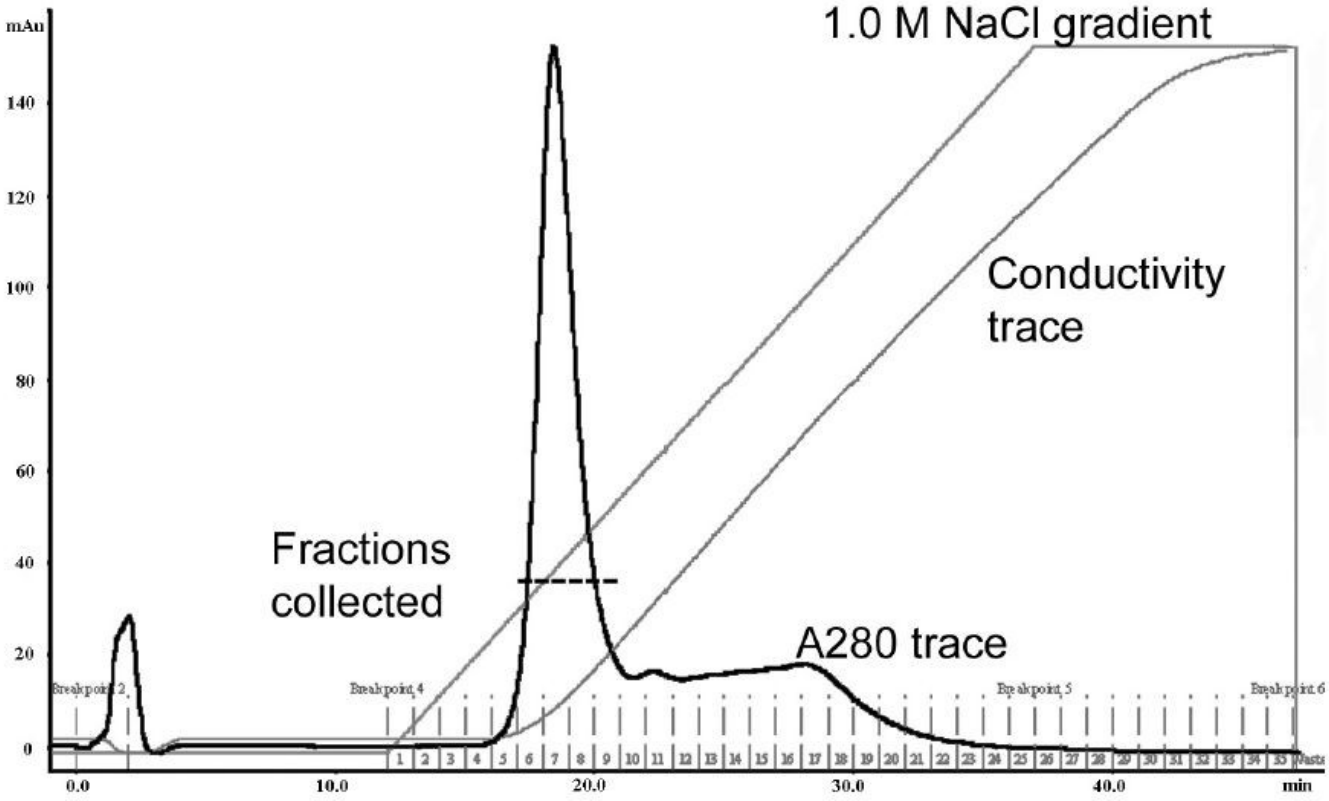


**Figure 1.** Chromatogram of *E. denticulatum* protein extract on Q-Sepharose (Amersham Biosciences) at 0.1 M Tris-HCl, pH 8.0, and gradient elution to 1.0 M NaCl. Fraction collection was 1.0 mL/min.

1993). A database that houses ESTs, the dbEST (http://www.ncbi.nlm.nih.gov/nucest), contains 9,368 *Eucheuma* sequences, of which 311 come from a small scale EST library of *E. denticulatum* (Aspilla et al. 2010) and 9,057 are derived from a large collection of over 10,000 EST sequences housed at the *E. denticulatum* EST Database (http://www.inbiosis.ukm.my/eudbase/) of R. Othman and colleagues (Hussein et al. 2011). Analysis of the small scale EST library reported a number of mRNA sequences in *E. denticulatum* that have no known homology to standard databases, suggesting that these may represent specific functions found in *E. denticulatum* and closely-related species (Aspilla et al. 2010).

Our group's preliminary protein profile of *E. denticulatum* showed that one water-soluble protein dominated our extracts, and this became the target of isolation and characterization. While we looked at *Eucheuma* proteins from the perspective of expression and abundance, investigators elsewhere had been exploring the functional aspects of *Eucheuma* extracts. Protein extracts from five different *Eucheuma* species were assayed for hemagglutinating activity to vertebrate erythrocytes by Kawakubo et al. (1997). They discovered that *E. denticulatum* showed the least hemagglutinating property, while *E. serra* had the highest activity. A protein from *E. serra* was further isolated and functionally characterized by the authors and named lectin ESA-2 (Kawakubo et al. 1999; Hori et al. 2007). In addition, the amino acid sequence of ESA-2 was determined directly from the protein (Hori et al. 2007). However, the protein extract from *E. denticulatum* was not characterized further by the authors.

This paper focuses on a major water-soluble protein from *E. denticulatum*, which we named EDEN-1. To characterize it, a two-pronged strategy was employed. First, mass spectrometric techniques were used for *de novo* sequencing (i.e., sequencing from scratch) of fragments derived from the protein. The amino acid sequences obtained were back-translated in order to design primers for reverse-transcriptase PCR amplification to obtain cDNA sequence information from EDEN-1's mRNA. Second, the protein itself was characterized in terms of size, charge, oligomeric state, and secondary structure. Also, given the importance of the polysaccharide produced by this alga, an initial attempt was made to explore the ability of EDEN-1 to hydrolyze glycosides.

## MATERIALS AND METHODS

### Sample Collection

*E. denticulatum* of the green variety was collected from a seaweed farm in Negros Oriental, Philippines. The fresh thalli (5-cm pieces taken from the tips) were washed and processed for protein extraction. Samples for RNA extraction were frozen in liquid nitrogen.

### Sample Preparation and Protein Isolation

The protein isolation method is a slight modification of total enzyme isolation of Rees (1961) for *Porphyra umbicalis*, a red alga that contains porphyran, an agar-like polysaccharide. Three kg of the algae were minced with 3.5 L of 0.5 % $Na_2CO_3$ solution (adjusted to pH 8.3) and allowed to stand for 24 hours. The mixture was then passed through three layers of gauze cloth and the filtrate adjusted to pH 6.6 with dilute acetic acid. The procedure of Rees was modified to use silica gel H 60 (100 g, Merck, Germany) instead of calcium phosphate gel to adsorb the proteins overnight. The silica gel with the adsorbed proteins was then centrifuged out and washed twice with one volume gel of 0.1 M acetate buffer (pH 6.5). The gel was then washed with 0.01 M phosphate buffer (pH 7.6), followed by 0.1 M phosphate buffer at the same pH. Final elution was with 0.2 M phosphate buffer, also at pH 7.6. Washings and eluates were combined, then treated with $(NH_4)_2SO_4$ at 80% saturation (i.e. 56.1 g $(NH_4)_2SO_4$ per 100 mL solution) (Horst 2000) and incubated at 4°C for three days. The mixture was then centrifuged. The precipitate was dissolved in water and dialyzed against water using a seamless cellulose tubing with MWCO 12000 Daltons (Sigma-Aldrich Corporation, St. Louis, MO, USA) at 4°C for

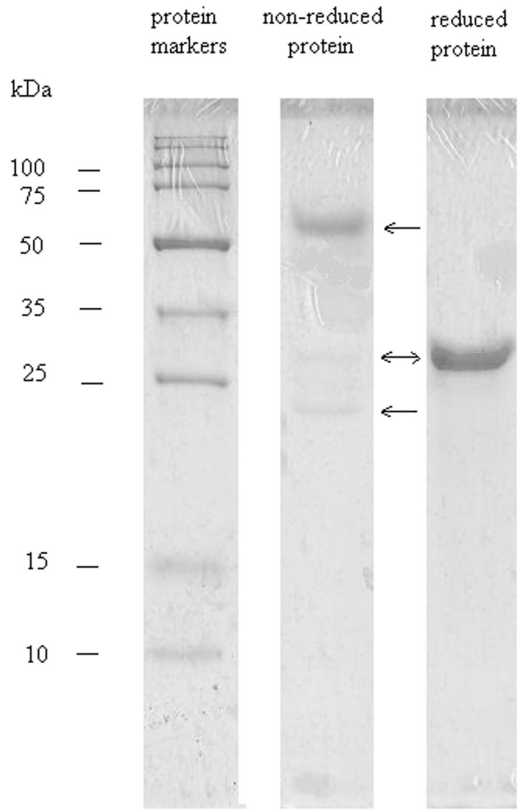


**Figure 2.** SDS-PAGE profile of EDEN-1 in the non-reduced form (DTT-free gel loading buffer) and reduced form (gel loading buffer with DTT).

three days at three times change of water.

**Protein Separation and Purification**

The crude EDEN-1 protein was purified by column chromatography using a strong anion exchanger. The crude protein was dissolved in 0.1 M Tris-HCl pH 8.0, loaded on Q Sepharose (Amersham Biosciences) and eluted via an NaCl gradient to 1.0 M on ÅKTAprime FPLC (Amersham Biosciences, Piscataway, NJ, USA). Protein elution was monitored by its UV absorbance at 280 nm. All solutions used for chromatography were filtered through 0.45 μm Durapore HV membrane filter and degassed.

**Protein Characterization**

**Size Determination.**

**Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).** SDS-PAGE was performed using standard procedures (Sambrook et al. 1989a) on a minigel format (Mini-Protean 3 system, Bio-Rad, Hercules, CA, USA). The resolving gel was 12% polyacrylamide. Reduced and non-reduced samples were loaded. For the reduced protein, the sample included 0.01M final concentration dithiothreitol (DTT) and was boiled for three minutes prior to loading, while for non-reduced, the DTT was omitted. Protein molecular weight (MW) markers (Cat # V8491, Promega Corp., Madison, WI, USA) were loaded on the gel to get apparent molecular weights of the protein samples. Running time for the proteins in the stacking gel was 10 min at 55V while resolution of the proteins was for one hour at 110V. Visualization of the protein was done by silver (Schevchenko et al. 1996) and Coomassie blue staining procedures (Sambrook et al. 1989a).

**Size-exclusion liquid chromatography.** A pre-packed Superdex™ 75 column (10 x 300 mm, 24 mL bed volume, Amersham Biosciences) was equilibrated with 200 mL pH 7 of 0.025 M phosphate buffer in 0.15 M NaCl. Five hundred microliters of EDEN-1 from the ion-exchange separation were introduced and eluted with the same buffer for three hours using the ÄKTAprime FPLC. Flow rate was 0.2 mL/min. A calibration curve was prepared by plotting the Stokes radii of the supplied proteins (Low Molecular Weight Gel Filtration kit, Amersham Biosciences) versus the square root of the negative log of the partition coefficient, $K_{av}$, where $K_{av}$ is $(V_e-V_o)/(V_t-V_o)$. The void volume, $V_o$, was determined by Blue Dextran 2000. $V_e$ and $V_t$ are elution and total bed volumes respectively.

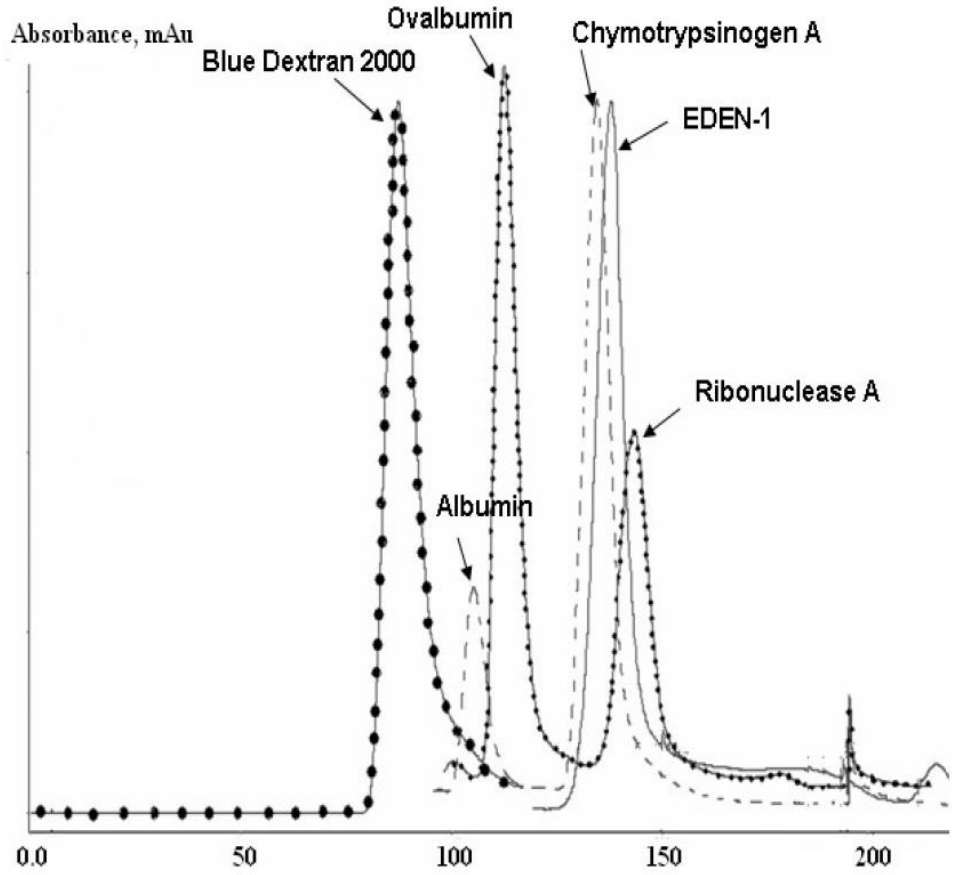


**Figure 3.** Size-exclusion chromatograms of EDEN-1 and size standards. The x-axis is in minutes.

**Mass spectrometry of intact protein.** The protein fraction from ion-exchange liquid chromatography was freeze-dried. The MALDI-TOF-MS was done using standard procedures. Briefly, the intact protein mass spectrum was acquired on a PerSeptive Biosystems Voyager-DE™ PRO Biospectrometry Workstation (PerSeptive Biosystems, Boston, MA, USA) at the University of Florida using sinapinic acid as matrix and 1% trifluoroacetic acid as proton donor.

. The sample protein spectrum was calibrated against protein standards of bovine insulin, equine myoglobin and bovine serum albumin (ProteoMass™ Protein MALDI-MS Calibration Kit, Sigma-Aldrich).

**Composition**

**Amino acid composition of the protein.** Purified protein was submitted to the Molecular Structure Facility (MSF) at University of California, Davis (USA) for amino acid composition analysis using standard procedures.

**Isoelectric point determination by 2-D gel electrophoresis.** For the 1st dimension, isoelectric focusing (IEF), was done using a linear 7 cm IPG strip pH 3-6 (Bio-Rad, Hercules, CA, USA). Sample loading onto the gel was combined with rehydration of the gel in 8.3 M urea, 0.5% CHAPS, 0.2% Bio-Lytes and 10 mM DTT at 20°C in a rehydration/equilibration tray for 12 hours. DTT was omitted for non-reducing conditions. After sample application and rehydration, isoelectric focusing was done by applying 50 µA/gel at 20°C in a preset linear voltage ramping method of 250 V for 15 min, ramping for two hours until the voltage reached 4000 volts, finally holding the voltage at 4000 V until a total of 20000 volt-hours was achieved. An overlay of mineral oil on the gel was applied in both rehydration and focusing steps to prevent drying out of the IPG strips.

Prior to the 2nd dimension (SDS-PAGE), the gel was saturated for 10-15 min with 6 M urea, 2% SDS, 0.375M Tris-HCl pH 8.8, and 20% glycerol. For reducing conditions, the buffer included 130 mM DTT, followed by a second 10 to 15 min incubation step in the same buffer but with 135 mM iodoacetamide instead of DTT.

For the 2nd dimension, the IPG strip was then placed horizontally on the wide sample well of the SDS-PAGE gel and held in place by 1% agarose gel. Molecular weight markers and electrophoresis conditions were the same as that of SDS-PAGE above.

**Glycosylation test using periodic acid-Schiff's (PAS) staining.** The method is adapted from Horst (2000). Reduced and non-reduced EDEN-1, molecular weight markers (also used as negative control), and positive controls (bovine submaxillary mucin (BSM) and horseradish peroxidase) were electrophoresed as with SDS-PAGE conditions as above. Two sets of each were placed in a single gel. After the run, the gel was incubated in 40% methanol/7% acetic acid overnight and fixed with 7.5% acetic acid for one hour. The gel was cut into two. One part was soaked with 1% periodic acid for one hour in the dark and washed 6 times with 7.5% acetic acid for 10 minutes each. It was then incubated in Schiff's reagent at 4°C in the dark for one hour, washed with 0.5% sodium metabisulfite and placed in 7.5% acetic acid. The other part was subjected to all conditions except periodic acid for comparison.

**Sequence information**

**Amino acid sequence**

**N-terminal determination of the protein.** Purified protein was submitted to the Molecular Structure Facility (MSF) at University of California, Davis (USA) for Edman degradation analysis of the N-terminus using standard procedures.

**Peptide sequencing by MS/MS.** EDEN-1 protein from ion-exchange liquid chromatography was freeze-dried. A

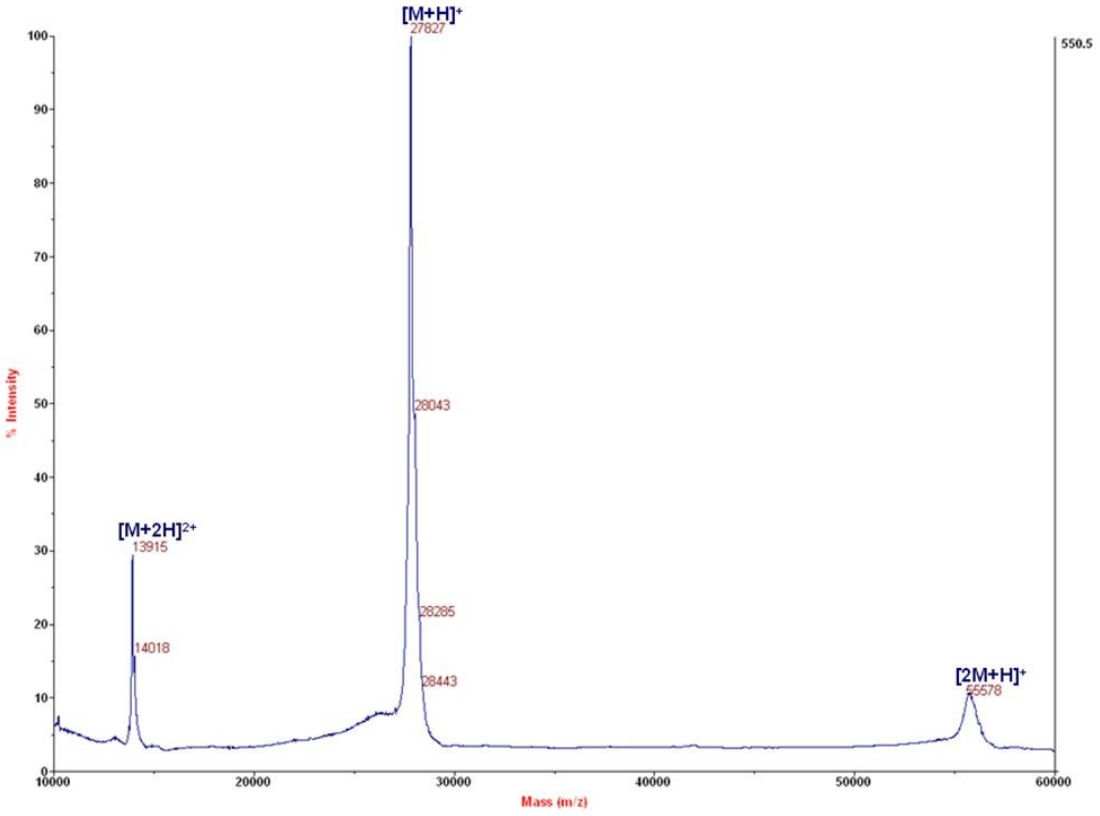


**Figure 4.** MALDI-TOF mass spectrum of the intact protein in positive mode. The spectrum shows the protonated molecular ion at *m/z* 27,827 that corresponds to a nominal mass of 27.83 kDa. This assignment is supported by the doubly-charged ion at *m/z* 13,915 and the protonated dimer at *m/z* 55,578 that indicate molecular weight of 28.23 kDa and 27.79 kDa, respectively

portion was reconstituted and subsequently dialyzed against water to lower the ionic strength (LD). The dialyzed sample was then freeze-dried. Both lyophilized (L) and lyophilized/dialyzed sample (LD) were subjected to SDS-PAGE, digested with trypsin, treated with iodoacetamide, dried, and submitted to the Molecular Structure Facility (MSF) at University of California, Davis (USA) for tandem MS analysis to obtain the sequence of peptide fragments using standard procedures.

**Peptide mass fingerprinting using MALDI-TOF-MS after in-gel digestion.** Protein purified by ion-exchange chromatography was digested using trypsin. Recovered peptides were analyzed by using a PerSeptive Biosystems Voyager-DE™ PRO Biospectrometry Workstation (PerSeptive Biosystems, Framingham, MA, USA) at the University of Florida. The observed peptide ions (*m/z*) obtained via MALDI-TOF-MS after tryptic digestion were compared with "in silico" generated digests of the protein sequence using MS-Digest of the Protein Prospector Suite (http://prospector.ucsf.edu/; Chalkley et al. 2005). The program takes into consideration possible missed cleavages and methionine oxidation.

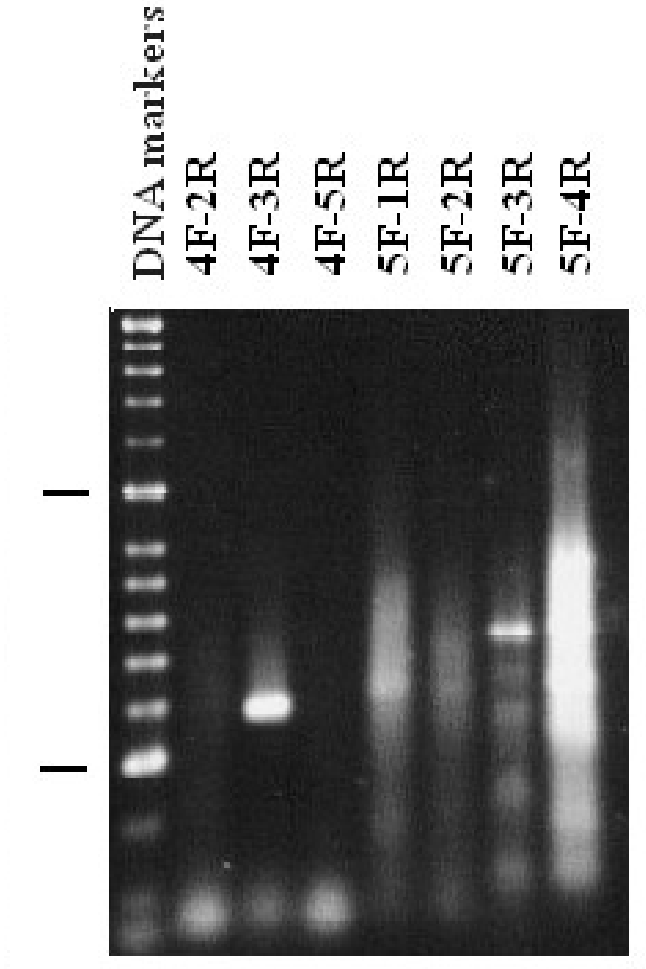


**Figure 5.** RT-PCR products of mRNA from *E. denticulatum*. These were amplified using different combination of primers designed from peptide sequences generated from tandem MS peptide sequencing as shown in Table 3. Template mRNA used was 100 ng. DNA marker is Hyperladder I (Bioline, London), with the upper mark corresponding to 1000 bp and the lower mark corresponding to 300 bp.

**Nucleotide sequence**

**Primer design.** From five peptide sequences generated by tandem MS, degenerate nucleotide primers were designed and sent to Invitrogen for synthesis (Invitrogen, Auckland, New Zealand). The primers are shown in Table 3, based on back translation of the peptide sequences.

**Isolation of total RNA and mRNA.** The isolation of total RNA and mRNA was performed at American Gene C.T. LLC (Cranston, Rhode Island). In brief, total RNA was isolated using the Tris-LiCl method of Wang et al. (2000), which is suitable for samples with high polysaccharide content. The procedure calls for multiple steps of precipitation with acetate and resolubilization in Tris buffer with LiCl precipitation towards the end. An extra phenol: chloroform extraction step to the LiCl RNA precipitate was added to the procedure. Approximately 5 g of algae was used. mRNA was isolated from the total RNA by the Oligotex mRNA Mini Kit (Qiagen, Hilden, Germany) modified by a second Oligotex binding step and addition of glycogen during precipitation of the mRNA.

**RT-PCR of isolated m-RNA.** RT-PCR of isolated mRNA was according to the recommended protocol of SuperScript™ III One-Step PCR System with Platinum® *Taq* High Fidelity (Invitrogen Life Technologies, Carlsbad, CA, USA) with some modifications. All steps were done on ice except where indicated. The thermal cycler (TPersonal, Biometra, Gottingen, Germany) program included cDNA synthesis for 1 h at 43°C, followed by pre-denaturation at 94°C for 2 min; 80 cycles of PCR amplification with denaturation at 94°C for 30 sec, annealing at 50°C for 30 sec, and extension at 72°C for 2 min; and a final extension step at 72°C for 10 min.

A master mix of 250 µL of 2X Reaction mix (0.4 mM of each dNTP, 2.4 mM $MgSO_4$), 10 µL of RNA template (about 50 ng/µL), and 10 µL of SuperScript™ Platinum Taq Hi-Fidelity Enzyme Mix was prepared. The mixture was made up to 2.5 mM $Mg^{2+}$ by adding 21 µL of 15 mM $MgSO_4$. One µL of RNAse inhibitor (RNAsin, Sigma) at 40 u/µL was also added. Total volume was 292 µL. Sixteen microliters of the primer mix were placed into 18 PCR tubes. A combination of forward primer and reverse primer at 2 µL each (10 mM) were added to each tube. Each PCR reaction then contained about 34 ng mRNA template.

The components were gently mixed and spun briefly to make sure that all the components are at the bottom of the PCR tube. The tubes were then placed in the pre-heated (43°C) thermal cycler and run. The cDNA products were then run on 1.5% agarose gel in 0.05X TBE buffer together with DNA markers (Hyperladder I, Bioline, London, UK).

**Purification of PCR products.** PCR products in solution (recommended for single bands) and PCR products run on agarose gels were purified using the Roche High Pure PCR Product Purification kit (Roche Diagnostics, Mannheim, Germany) following the manufacturer's specifications.

**Ligation of DNA into a vector and cloning into *E. coli* competent cells.** Ligation protocol was taken from Promega (Fitchburg, WI, USA). Seven µL of purified PCR product were mixed with one µL each of 10X Ligation Buffer, pGEM-T Easy vector and T4 DNA ligase (3 Weiss units/µL), all from Promega. The mixture was then placed in a water bath at 4ºC overnight. Two µL of ligation product were added to 50 µL of *E. coli* competent cells (JM109 High Efficiency Competent Cells, Promega) and the mixture stirred gently with a pipe tip. The mixture was incubated on ice for 20 min, warmed for 45 seconds at 42ºC, and immediately chilled on ice for 2 minutes. Pre-warmed Luria-Bertani (LB) broth (500 µL) was added and vigorously shaken for one hour at 37ºC. One hundred microliters were plated out on an LB-Amp-XT-gal agar plate, sealed with parafilm and incubated upside down for 16 hours but not more than 20 hours at 37ºC. The plates were placed at 4ºC for at least an hour to distinguish white colonies from blue colonies.

**Minipreparation of plasmid DNA – Lysis by Alkali.** Plasmid DNA was obtained using the method of Sambrook et al. (1989b). In a culture tube, a white colony was inoculated in 2 mL LB containing 4 µL of Ampicillin (50 mg/mL). The bacteria were grown overnight at 37°C in a shaker. The bacteria were transferred into a 1.5 µL microcentrifuge tube and pelleted out by centrifugation at 13000 rpm for five minutes. The bacterial

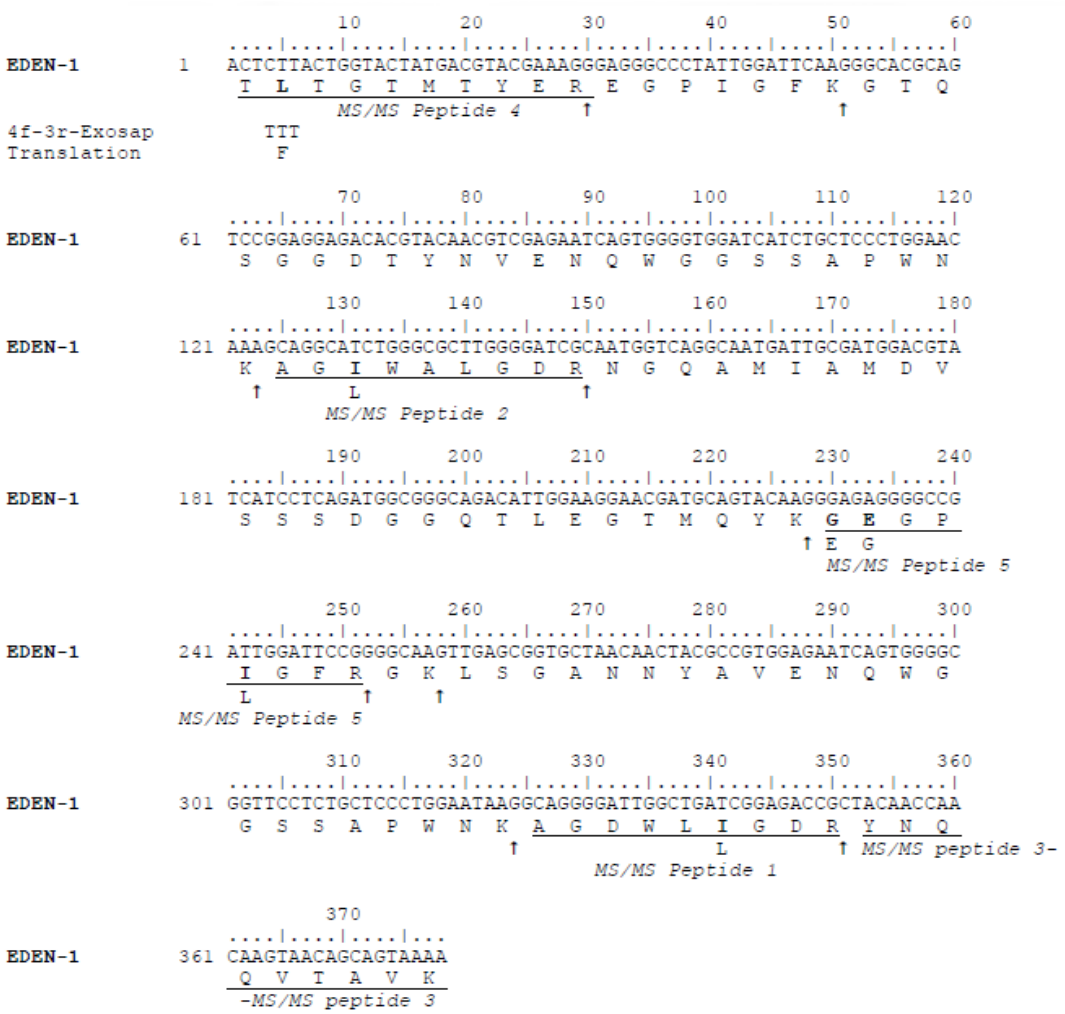


**Figure 6.** Partial nucleotide and amino acid sequence of EDEN-1. The underscored amino acid sequences refer to the peptide sequences obtained from tandem MS. Residues in boldface show inconsistencies in the data, and alternative sequences are noted below for comparison. Note that the MS data show leucines, but isoleucines indicated by the nucleotide sequence are also consistent with the mass data. Arrows indicate trypsin cleavage sites. The cDNA and amino acid sequences are deposited in GenBank under accession number JX021449.

pellet was resuspended in 100 µL of Solution I (sterile 50 mM glucose in pH 8.0 25 mM Tris-HCl and pH 8.0 10mM EDTA) by vortexing. Two hundred microliters of freshly prepared Solution II (0.2% NaOH and 1% SDS) were then added. The contents were mixed thoroughly and incubated on ice for five minutes. Finally, 150 µL of Solution III (29.4 g KOAc and 11.5 mL glacial acetic acid in 100 mL) were added and dispersed through the viscous bacterial lysate. The tube was centrifuged for 10 min at maximum speed and the supernatant transferred to a fresh tube. The supernatant was then treated with equal volumes of phenol and chloroform, respectively. The DNA was then precipitated out of the aqueous layer by adding 2.5 volumes of cold absolute ethanol. The precipitate was then washed with 70% EtOH and air dried. The DNA was then redissolved in sterile distilled water or 0.1X TE buffer (10 mM Tris, 0.1M EDTA) and stored at -20ºC.

**Sequencing.** Purified plasmid DNA from above was submitted for sequencing to Macrogen (Seoul, Korea) using M13 forward and reverse primers. In the case of single band PCR products, a simplified clean-up procedure for treating the PCR products with ExoSap-IT® (a combination of endonuclease I and shrimp alkaline phosphatase) was employed following manufacturer's specifications (Amersham Biosciences, Piscataway, NJ, USA). Sequencing primers were the degenerate primers used in RT-PCR.

**Sequence similarity searches of the amino acid sequence.** Amino acid sequences were submitted online for similarities using protein BLAST (Altschul et al. 1997). The amino acid sequence of the best matches in the database were compared and aligned with EDEN-1 using CLUSTALW (Larkin et al. 2007). BioEdit Sequence Alignment Editor (Hall 1999) was also used to view the data from different sequencing analyses.

**Other protein properties**

**Secondary structure of the protein by circular dichroism.** Protein purified from ion-exchange chromatography was lyophilized and subsequently dialyzed against 0.025M phosphate buffer, pH 7.0, 0.1 M NaCl. The sample was then submitted to the Analytical Services Laboratory, University of the Philippines-Diliman for analysis. Circular dichroism was determined using a Jasco J-715 CD/ORD Spectropolarimeter and scanned from 200 to 250 nm.

**Screening for glycoside hydrolysis activity.** A protein concentration of 300 µg/mL (as determined by bicinchoninic acid assay) was tested for glycosidase activity. Substrates tested (at concentrations of 4 ng/µL) were *p*-nitrophenyl derivatives (PNGs) of β-D-xylopyranoside, β-D-glucopyranoside, β-L-arabinopyranoside, β-D-galactopyranoside, α-L-rhamnopyranoside and α-D-glucopyranoside (all from Sigma, St. Louis, MO, USA). EDEN-1 protein and each substrate were allowed to react in 0.025M phosphate buffer pH 7 at a volume

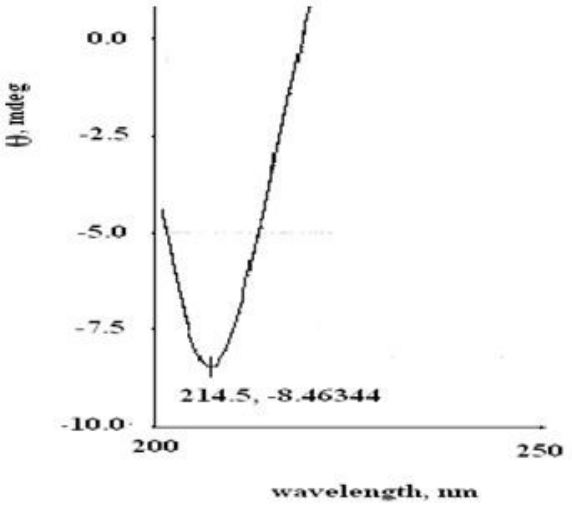


**Figure 7.** Circular dichroism spectrum of EDEN-1in 0.025 M phosphate buffer, pH 7.0. Spectrum was taken on a Jasco CD-ORD instrument (Institute of Chemistry, University of the Philippines-Diliman). Path length used was 0.2 cm.

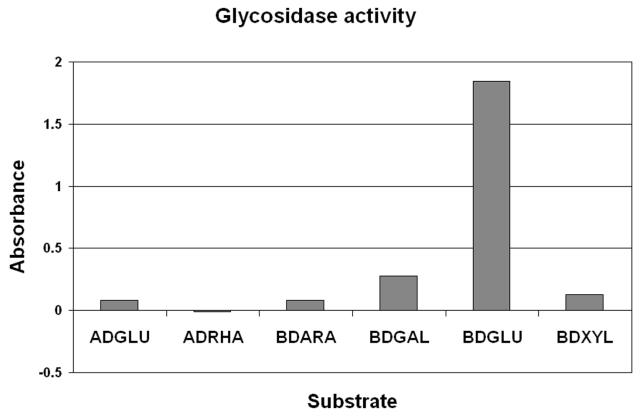


**Figure 8.** Hydrolase activity of EDEN-1 on p-nitrophenyl glycosides. Final readings were taken after incubation at room temperature for three days. ADGLU = alpha-D-glucoside, ADRHA = alpha-D-rhamnoside, BDARA = beta-D-arabinoside, BDGAL = beta-D-galactoside, BDGLU = beta-D-glucoside, and BDXYL = beta-D-xyloside.

ratio of 1 enzyme to 9 of substrate for a total volume of 100 µL. The absorbance of the resulting solution was measured using a Shimadzu CS 9301 Dual Wavelength Flying Spot Scanning Densitometer at 400 nm (Shimadzu, Kyoto, Japan).

## RESULTS

**Protein isolation and purification.** The protein extract from *E. denticulatum,* upon ion exchange chromatography on a strong anion exchanger, showed a major peak that eluted at about 20% of the 1.0 M NaCl gradient, as shown in Figure 1. This indicates that at pH 8.0, the protein was negatively charged overall. The collected fractions from this major peak were pooled for further characterization.

**Protein size by SDS-PAGE.** The fraction collected from ion-exchange chromatography yielded a single peak of approximately 27 kDa on SDS-PAGE. Under non-reducing conditions (without DTT), a major band at 55 kDa and two less intense bands at 27 kDa and 22 kDa respectively were revealed, suggesting the presence of cysteine residues that could form a disulfide-bridged dimer (the 55-kDa band) or a more compact structure (the 22-kDa band), as shown in Figure 2. Excision of the bands from non-reducing conditions, followed by electroelution and SDS-PAGE, confirmed that the 55-kDa and 22-kDa bands are alternative conformations of the 27-kDa SDS-PAGE band, and not other proteins.

**Protein size by size exclusion chromatography.** On a size exclusion or gel-filtration column, under near physiological conditions (pH 7.0, 0.025 M phosphate buffer, 0.15 M NaCl), the protein elutes between chymotrypsinogen (MW 25.0 Da) and ribonuclease A (MW 13.7 kDa), but has an elution volume closer to that of chymotrypsinogen (Figure 3 and Table 1). Calculation of the molecular weight based on a calibration curve of log of the molecular weight of size standards vs. partition coefficient or elution volume yields a size 18.7 kDa. For comparison, using the Stokes radius, Rs, and a calibration curve of the square root of the negative logarithm of the partition coefficient vs. the Stokes radius, yields a size of 18.9 Å for EDEN-1. The gel filtration chromatography results, together with the SDS-PAGE results, indicate that the protein is monomeric and compact under physiological conditions.

**Protein intact mass by mass spectrometry.** The molecular weight of the unmodified intact protein by MALDI-TOF mass spectrometry is 27.81 kDa (± 0.02 kDa) based on the protonated ion at *m/z* 27,827, the doubly-charged ion at *m/z* 13,915 and the dimer at *m/z* 55,578 (Figure 4). This is more accurate than the 27 kDa (±1 kDa) that was calculated from SDS-PAGE.

**Amino acid composition.** Table 2 shows the results of amino acid composition analysis. The results suggest that the protein is probably acidic, due to the large number of Asx and Glx. Although some of these residues may be found as Asn and Gln, respectively, some may be in the acid form. Two cysteines

```
                                10        20        30        40        50        60
                           ....|....|....|....|....|....|....|....|....|....|....|....|
EDEN-1, partial       1    ------------------------------------------------------------
ESA-2 (Hori et al.)   1    GRYTVQNQWGGSSAPWNDAGLWILGSRGNQNVMAVDVNSSDGGANLNGTMTYSGEGPIGF

                                70        80        90       100       110       120
                           ....|....|....|....|....|....|....|....|....|....|....|....|
EDEN-1, partial       1    ----------------------------------------------------TLTGTMTY
ESA-2 (Hori et al.)   61   KGARRGESNVYDVENQWGGSSAPWHAGGQFVIGSRSGQGVLAVNITSSDGGK........

                               130       140       150       160       170       180
                           ....|....|....|....|....|....|....|....|....|....|....|....|
EDEN-1, partial       9    EREGPIGFKGTQSGGDTYNVENQWGGSSAPWNKAGIWALGDRNGQAMIAMDVSSSDGGQT
ESA-2 (Hori et al.)   112  ........................................S...............K.

                               190       200       210       220       230       240
                           ....|....|....|....|....|....|....|....|....|....|....|....|
EDEN-1, partial       69   LEGTMQYKGEGPIGFRGKLSGANNYAVENQWGGSSAPWNKAGDWLIGDRYNQQVTAVK--
ESA-2 (Hori et al.)   114  ........................S.............A.........H..NI....VS

                               250       260       270
                           ....|....|....|....|....|....|..
EDEN-1, partial       126  -----------------------------------
ESA-2 (Hori et al.)   122  SDNDGKNLDGTCTYEREGPIGFKGVATS
```

**Figure 9.** Comparison of the partial amino acid sequence of EDEN-1 with the sequence of lectin ESA-2 from *Eucheuma serra* (Hori et al., 2007). Dashes indicate gaps, while dots in the sequence indicate identity.

are likely to be present. This is consistent with the difference in behavior between the reduced and non-reduced forms during electrophoresis.

**Isoelectric point determination.** The isoelectric point or pI of EDEN 1 was approximately 4.7 from isoelectric focusing experiments. This acidic isoelectric point is consistent with the amino acid composition, as well as with the ion-exchange chromatography behavior of the protein.

**Glycosylation status.** Glycoprotein staining experiments indicate that the protein is not glycosylated.

**N-terminal sequence.** N-terminal analysis yielded GR(W,Y,A)TVQN, although there was some uncertainty in position 1 also. Due to the ambiguity in position 3, this amino acid sequence was not used to design primers for cDNA synthesis and amplification.

**Amino acid sequencing of fragments by mass spectrometry.** Additional amino acid sequence information was obtained by digestion of EDEN-1 with trypsin followed by tandem mass spectrometry. The five peptides that yielded sequences through MS/MS are listed in Table 3, along with the degenerate primers that were derived from them by back translation.

**cDNA amplification and partial gene sequencing.** The degenerate primers derived from back translation of peptide fragment MS/MS sequences were used to generate a number of PCR products (Figure 5). The combination of primers 4F and 3R yielded an intense band of 400-500 bp on electrophoresis. This band was excised and sequenced.

**Partial gene sequencing.** The 4F-3R cDNA fragments amplified from mRNA were sent for DNA sequencing directly or cloned into a plasmid vector and then sequenced. Comparison and alignment of the results of the different DNA sequencing trials give us the EDEN-1sequence shown in Figure 6, with the associated translation into amino acid residues.

The amino acid sequence from translation of the nucleotide sequence is also shown in Figure 6. The translated amino acid sequence is also compared to the results obtained directly from tandem MS. Note that in most cases, the difference is between

**Table 1.** Size exclusion chromatography results of EDEN-1 in comparison to molecular weight size standards. Void volume, $V_0$, of Blue Dextran 2000 is 8.0mL, while bed volume, $V_0$, is 24.0 mL. The partition coefficient, $K_{av} = (V_e-V_0)/(V_t-V_0)$, is also shown.

| Protein | MW (kDa) | Stokes radius, $R_S$ (Å) | Elution volume, $V_e$ (mL) | Calculated partition coefficient, $K_{av}$ |
|---|---|---|---|---|
| Albumin | 67.0 | 35.5 | 10.5 | 0.156 |
| Ovalbumin | 43.0 | 30.5 | 11.3 | 0.206 |
| Chymotrypsinogen | 25.0 | 20.9 | 13.5 | 0.344 |
| Ribonuclease A | 13.7 | 16.4 | 14.3 | 0.394 |
| EDEN-1 | *18.7 (calc)* | *18.9 (calc)* | 13.8 | 0.363 |

**Table 2.** Amino acid composition of EDEN-1. The minimum molecular weight consistent with this composition is 24,874 Da, using Glx as glutamic acid only and Asx as aspartic acid only in the calculations

| **Amino acid** | **nmol** | **Molar ratio vs. His** | **Relative Number of Residues** | **Number in EDEN-1 partial sequence** |
|---|---|---|---|---|
| Ala | 1.759 | 16.75 | 17 | 10 |
| Arg | 0.810 | 7.71 | 8 | 4 |
| Asx (Asp + Asn) | 2.791 | 26.58 | 27 | 6 + 9 |
| Cys (Cysteic acid) | 0.211 | 2.01 | 2 | 0 |
| Glx (Glu + Gln) | 2.289 | 21.80 | 22 | 6 + 8 |
| Gly | 4.189 | 39.90 | 40 | 23 |
| His | 0.105 | 1.00 | 1 | 0 |
| Ile | 0.900 | 8.57 | 9 | 5 |
| Leu | 0.933 | 8.89 | 9 | 5 |
| Lys | 0.870 | 8.29 | 8 | 6 |
| Met (as MetSO2) | 0.420 | 4.00 | 4 | 4 |
| Phe | 0.431 | 4.10 | 4 | 2 |
| Pro | 0.773 | 7.36 | 7 | 4 |
| Ser | 1.935 | 18.43 | 18 | 9 |
| Thr | 1.357 | 12.92 | 13 | 9 |
| Trp | not determined | | | 6 |
| Tyr | 0.646 | 6.15 | 6 | 5 |
| Val | 1.258 | 11.98 | 12 | 5 |

leucine and isoleucine, which have identical mass. One DNA sequence result showed a TTT for the second codon, while others indicated a CTT. Glutamate and glycine are interchanged at the beginning of MS/MS peptide 5 in the middle of the EDEN-1 sequence.

**Peptide mass fingerprinting.** Peptide mass fingerprinting using MALDI-TOF-MS after trypsin digestion was used to clarify discrepancies among the translated nucleotide sequences and the peptide sequences derived from *de novo* tandem mass spectrometry. In particular, different cDNA sequences were obtained for codon 2 from 2 different cDNA clones, leading to a difference of leucine vs. phenylalanine. This was resolved by comparing the experimental peaks with the tryptic digests predicted from the sequence, as shown in Table 4. The presence of *m/z* = 1,172.6 confirms that leucine is the amino acid residue at position 2, rather than phenylalanine, which would have produced a peak at *m/z* 1,206.5 instead. Peptide products from missed cleavages were also observed in the mass spectrum, yielding longer fragments consistent with the sequence.

**BLAST results of the sequence.** Comparison of the partial protein sequence via BLAST with entries in UniProt (http://www.uniprot.org; UniProt Consortium 2011) and GenBank yields a strong similarity with Lectin ESA-2 (UniProt P84331, entry nameLEC1_EUCSE) from *Eucheuma serra*, with an e-value of $1\times10^{-67}$. Sequence alignment of EDEN-1 and ESA-2 (Figure 9) shows that the EDEN-1 sequence fragment corresponds to residues 113-238 of ESA-2, that the two differ in identity at only 7 of the 126 sequenced positions (6 %), and that the differences are generally conservative, such as glutamine in EDEN-1 compared to asparagine in ESA-2 position 233 or lysine in ESA-2 position 179.

**Secondary structure.** The circular dichroism spectrum showed a negative band at 214 nm (Figure 7), which is characteristic of β-sheets. This finding is consistent with the similarity of the sequence to those of lectins known to be predominantly beta sheet in structure.

**Screening for glycoside hydrolysis activity.** Glycosidic bond cleavage activity of the EDEN-1 protein was extremely slow, but selective for *p*-nitrophenyl-β-D-glucopyranoside. Reaction was observed with *p*-nitrophenyl-β-D-glucopyranoside alone after overnight incubation. After three days, the color intensity of the reaction mixtures was measured in comparison. The order of reactivity among the six *p*-nitrophenyl glycosides tested were as follows: *p*-nitrophenyl-β-D-glucopyranoside >> *p*-nitrophenyl-β-D-galactopyranoside > *p*-nitrophenyl-β-D-xylopyranoside > *p*-nitrophenyl-α-D-glucopyranoside and *p*-nitrophenyl-β-D-arabinopyranoside >> *p*-nitrophenyl-α-D-rhamnopyranoside (Figure 8). The last substrate did not show any color change at all.

## DISCUSSION

In this study, we report the isolation and characterization of a major water soluble protein from *E. denticulatum*. Through a combination of peptide and nucleotide sequencing strategies, we were able to determine a partial sequence for the protein, which we called EDEN-1, both at the nucleotide level and at the protein level.

The final sequence shown in Figure 6 best matches the data gathered from DNA sequencing and tandem MS/MS sequencing of the protein, verified by peptide mass fingerprinting. This partial amino acid sequence of EDEN-1 consists of 126 amino acids and has a molecular weight of 13,414 Da, corresponding to 48.2% of the intact mass of the protein by mass spectrometry. All five peptides that yielded *de novo* amino acid sequence data are included in this partial sequence, with some corrections based on the combined data sets.

The sequence of EDEN-1 is very similar to that of the lectin-like protein ESA-2 from the related alga *Eucheuma serra*, whose amino acid sequence is known but whose nucleotide sequence is not (Hori et al. 2007). ESA-2 has a calculated molecular mass of 27,950.3 Da and an intact mass of 27,949.0

**Table 3.** Peptide fragment sequences and the PCR primers generated from them. The peptide sequences were derived from *de novo* sequencing of tryptic peptides using tandem mass spectrometry at the University of California-Davis Molecular Structure Facility.

| | SEQUENCE | FORWARD PRIMER* (5' to 3', designated as F) | REVERSE PRIMER* (5' to 3', designated as R) |
|---|---|---|---|
| 1 | AGDWLLGDR | GCN GGN GAY TGG YTN YTN GGN GA | CKR TCN CCN ARN ARC CAR TCN CCN GC |
| 2 | AGLWALGDR | GCN GGN YTN TGG GCN YTN GGN GA | CKR TCN CCN ARN GCC CAN ARN CCN GC |
| 3 | YNQQVTAVK | CKR TCN CCN ARN GCC CAN ARN CCN GC | TAY AAY CAR CAR GTN ACN GCN GTN AA |
| 4 | TLTGTMTYER | ACN YTN ACN GGN ACN ATG ACN TAY GA | CKY TCR TAN GTC ATN GTN CCN GTN ARN GT |
| 5 | EGGPLGFR | GAR GGN GGN CCN YTN GGN TT | CKR AAN CCN ARN GGN CCN CCY TC |

*Standard code for degenerate nucleotides: R=GA, Y=TC, K=GT, M=AC, S=GC, W=AT, E=GTC, D=GAT, H=ACT, V=GCA, N=AGCT.

Da, similar to EDEN-1. ESA-2 is one of several lectins isolated from marine red alga (Kawakubo et al. 1997; Kawakubo et al. 1999). These molecules share similar N-terminal sequences. Revisiting the ambiguous N-terminal analysis for EDEN-1 in comparison with these lectins suggests that the N-terminal sequence for EDEN-1 is most likely to be GRYTVQN, with Y at the ambiguous 3rd position. This is similar to the N-terminal sequences from lectins isolated from *E. serra* and *E. cottonii* (also known as *Kappaphycus alvarezii)* (Kawakubo et al. 1999).

The lectins described in detail by Kawakubo et al., especially ESA-2, were found to have strong hemagglutinating activity (Kawakubo et al. 1997; Kawakubo et al. 1999). The carbohydrate-binding profile of ESA-2 was described in detail by Hori et al. (2007).

On the other hand, extracts from *E. denticulatum* did not exhibit strong hemagglutinating activity despite their similarity in sequence (Kawakubo et al. 1997; Suarez-Aspilla, unpublished results). We hypothesized that if EDEN-1 had some carbohydrate-binding property, it may be possible to get a hint of this by screening for glycoside hydrolase activity. Our initial results, shown in Figure 8, suggest that EDEN-1 may be able to hydrolyze certain carbohydrates slowly, with a preference for glucose moieties.

The composition, beta-sheet structure, and monomeric nature of EDEN-1 are consistent with *Eucheuma* lectin-like proteins or lectins in general, despite the apparent lack of hemagglutinating properties (Hermann et al. 1998).

This is the first report of a nucleotide sequence, albeit partial, for the *Eucheuma* lectin-like proteins. MEGABLAST search of the partial nucleotide sequence of EDEN-1 against the non-redundant nucleotide database at GenBank (Benson et al. 2009) does not yield any matches. Nor does a nucleotide BLAST search yield matches against the dbEST database (Boguski et al, 1993), which contains 9,368 entries of sequence fragments derived from mRNA of *E. denticulatum*.

The lack of hits in dbEST suggests that the high abundance of EDEN-1 in *E. denticulatum* is not necessarily reflected in the mRNA expression patterns of this organism, even though the nucleotide sequences described in this study were amplified from the same mRNA preparation that yielded the small scale EST library (Aspilla et al. 2010). This result is not entirely a surprise since protein abundance is not necessarily correlated with abundance of its corresponding mRNA (Gygi et al. 1999).

**Table 4.** Comparison of experimental MALDI-TOF-MS of tryptic peptides with masses of predicted tryptic digestion products of the EDEN-1 sequence shown in Figure 6. The predictions, made using MS Digest (http://prospector.ucsf.edu/cgi-bin/msform=msdigest) also account for missed cleavages and methionine oxidation.

| *m/z* | Modifications | Start | End | Missed Cleavages | Sequence | Intensity[1] |
|---|---|---|---|---|---|---|
| 747.4 | | 11 | 17 | 0 | (R)EGPIGFK(G) | ND |
| 832.4 | | 77 | 84 | 0 | (K)GEGPIGFR(G) | S |
| 958.5 | | 42 | 50 | 0 | (K)AGIWALGDR(N) | S |
| 1002.5 | | 109 | 117 | 0 | (K)AGDWLIGDR(Y) | S |
| 1050.6 | | 118 | 126 | 0 | (R)YNQQVTAVK(-) | ND |
| 1172.6 | | 1 | 10 | 0 | (-)TLTGTMTYER(E) | S |
| 2034.0 | | 109 | 126 | 1 | (K)AGDWLIGDRYNQQVTAVK(-) | S |
| 2350.1 | | 87 | 108 | 0 | (K)LSGANNYAVENQWGGSSAPWNK(A) | S |
| 2535.2 | | 85 | 108 | 1 | (R)GKLSGANNYAVENQWGGSSAPWNK(A) | W |
| 2540.1 | | 18 | 41 | 0 | (K)GTQSGGDTYNVENQWGGSSAPWNK(A) | S |
| 2719.2 | | 51 | 76 | 0 | (R)NGQAMIAMDVSSSDGGQTLEGTMQYK(G) | ND |
| 2735.2 | 1 Met-ox | 51 | 76 | 0 | (R)NGQAMIAMDVSSSDGGQTLEGTMQYK(G) | S |
| 2751.2 | 2 Met-ox | 51 | 76 | 0 | (R)NGQAMIAMDVSSSDGGQTLEGTMQYK(G) | W |
| 2767.2 | 3 Met-ox | 51 | 76 | 0 | (R)NGQAMIAMDVSSSDGGQTLEGTMQYK(G) | W |

[1]ND = not detected; S = strong to moderate; W = weak

## SUMMARY AND CONCLUSION

We have isolated a major water-soluble protein called EDEN-1 from *E. denticulatum*. It is related by composition, sequence, size, and monomeric state to ESA-2 from *E. serra* and other hemagglutinating lectins from other *Eucheuma* species, but unlike them, EDEN-1 does not exhibit appreciable hemagglutinating activity. However, it has a beta-sheet secondary structure profile that is consistent with many lectins, and preliminary evidence suggests that it may

bind to and hydrolyze carbohydrates, particularly glucose moieties. About half of the amino acid sequence of the protein has been determined. More important, the nucleotide sequence coding for this portion of the protein has been determined.

These specific findings may be used in the future to (1) obtain complete gene information for this class of proteins; and (2) understand the subtle differences in sequence and structure among these algal proteins that lead to large differences in hemagglutinating properties and carbohydrate-binding specificities. Furthermore, the natural function of this abundant protein in *Eucheuma* remains an open question inviting future studies.

## NOTE ADDED IN PROOF

The partial cDNA and amino acid sequences of EDEN-1 that are described in this paper (see Figure 6) are deposited in GenBank under accession number JX021449.

## ACKNOWLEDGMENTS

The authors thank Dr. Milady R. Niñonuevo for trypsin digestion and sample preparation for the peptide sequencing, and for coordinating the sample submissions to the University of California-Davis Molecular Structure Facility. We thank Professor Carlito B. Lebrilla for supporting the analyses performed at University of California-Davis. We also thank Dr. David Cheng of American Gene C.T. LLC (Cranston, RI, USA) for mRNA preparation and related advice. This work was supported by grants from the Commission on Higher Education (CHED) and the Department of Science and Technology-Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST-PCIEERD), formerly Philippine Council for Advanced Science and Technology Research and Development (DOST-PCASTRD), of the Republic of the Philippines, and the United Board for Christian Higher Education in Asia, Silliman University, and Ateneo de Manila University.

## CONFLICTS OF INTEREST:

The authors declare no conflict of interest.

## CONTRIBUTIONS OF INDIVIDUAL AUTHORS

PBSA performed the experiments, except where noted, and co-wrote the manuscript. MMT assisted in the protein experiments, especially the purification and electrophoretic analyses. MCAD performed and led the analysis of the intact mass and peptide mass fingerprinting experiments. GCZ guided the cDNA experiments, provided insight on algal biology, hosted PBSA's work in his laboratory, and edited the manuscript. NRLR guided the overall project and co-wrote the manuscript.

## REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25:3389-3402.

Aspilla PS, Antonio AACB, Zuccarello GC, Rojas NRL. A partial expressed sequence tag (EST) library of the economically important red alga *Eucheuma denticulatum* (N. L. Burman) Collins and Hervey. Phil Sci Lett 2010; 3: 109-120; as corrected 12 August 2011.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. Nucleic Acids Res 2009; 36(Database issue):D25-30.

Boguski MS, Lowe TM, Tolstoshev CM. dbEST--database for "expressed sequence tags". Nature Genetics 1993; 4(4):332-333.

Chalkley RJ, Hansen KC, Baldwin MA. Bioinformatic methods to exploit mass spectrometric data for proteomic applications. Methods in Enzymology 2005; 402:289-312.

FAO Fisheries and Aquaculture Department (2010). The State of World Fisheries and Aquaculture. Rome: Food and Agriculture Organization of the United Nations, 19-20. http://www.fao.org/docrep/013/i1820e/i1820e.pdf.

Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Molecular and Cellular Biology 1999; 19:1720-1730.

Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 1999; 41:95-98.

Hermann MS, Richardson CE, Seltzer LM, Behnke WD. A circular dichroic investigation of the secondary structure of lectins. Biopolymers 1998; 17:2107-2120.

Hori K, Sato Y, Ito K, Fujiwara Y, Iwamoto Y, Makino H, Kawakubo A. Strict specificity for high-mannose type N-glycans and primary structure of a red alga *Eucheuma serra* lectin. Glycobiology 2007; 17:479-491. doi:10.1093/glycob/cwm007.

Horst M. Molecular approaches to glycobiology. (Manual prepared for Interdisciplinary Center for Biotechnology Research Workshop at University of Florida, Gainesville, USA, May 8-12, 2000).

Hussein ZAM, Loke KK, Abidin RAZ, Othman R. EuDBase: An online resource for automated EST analysis pipeline (ESTFrontier) and database for red seaweed *Eucheuma denticulatum*. Bioinformation 2011; 7: 157-162.

Jain E, Bairoch A, Duvaud S, Phan I, Redaschi N, Suzek BE, Martin MJ, McGarvey P, Gasteiger E. Infrastructure for the life sciences: design and implementation of the UniProt website. BMC Bioinformatics 2009; 10:136. http://www.uniprot.org.

Kawakubo A, Makino H, Hirohara H, Hori K. The marine red alga *Eucheuma serra* J. Agardh, a high yielding source of two isolectins. J Appl Phycol 1997; 9:331-338.

Kawakubo A, Makino H, Ohnishi J-I, Hirohara H, Hori K. Occurrence of highly yielded lectins homologous within the genus *Eucheuma*. J Appl Phycol 1999; 11: 149–156.

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

McDermid J, Stuercke B. Nutritional composition of edible Hawaiian seaweeds. J Appl Phycol 2003; 15:513-524.

Msuya M, Neori A. *Ulva reticulata* and *Gracilaria crassa*: macroalgae that can biofilter effluent from tidal fishponds in Tanzania. Western Indian Ocean J Mar Sci 2002; 1:117-126.

Rees DA. Enzymic synthesis of 3,6-anhydrogalactose within porphyran from L-galactose 6-sulfate units. Biochem J 1961; 81:347.

Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: A laboratory manual. 2nd ed. Plainview, New York: Cold Spring Harbor Press, 1989a; 18.47-18.59.

Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: A laboratory manual. 2nd ed. Plainview, New York: Cold Spring Harbor Press, 1989b; 1.25-1.28.

Schevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal Chem 1996; 68:850-858.

Trono GC. Cultured Aquatic Species Information Programme. Eucheuma spp. FAO Fisheries and Agriculture Department, Rome 2005; http://www.fao.org/fishery/culturedspecies/Eucheuma_ spp/en Accessed on May 15 2011.

Wang, S., Hunter W, Plant A. Isolation and purification of functional total RNA from woody branches and needles of Sitka and White spruce. BioTechniques 2000; 28:292-296.

UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res 2011; 39: D214-D219.