Available online at www.sciencedirect.com
SCIENCE DIRECT
Biochemical and Biophysical Research Communications 344 (2006) 173–180
BBRC
www.elsevier.com/locate/ybbrc

Precursor structure of egg proteins in the coral *Galaxea fascicularis*

Hideki Hayakawa [a,*], Tadashi Andoh [b], Toshiki Watanabe [a]

[a] *Division of Molecular Marine Biology, Department of Marine Bioscience, Ocean Research Institute, The University of Tokyo, 1-15-1 Minamidai, Nakano, Tokyo 164-8639, Japan*
[b] *Coastal Fisheries and Aquaculture Division, Hokkaido National Fisheries Research Institute, Fisheries Research Agency, 116 Katsurakoi, Kushiro, Hokkaido 085-0802, Japan*

Received 20 February 2006

Abstract

In the egg of the reef coral *Galaxea fascicularis*, four proteins (named GfEP-1 to -4) are stored in high abundance. In the present study, a cDNA containing a full-length open reading frame for GfEP-1 was cloned, and the translated protein sequence was compared to the N-terminal sequences of GfEP-2, -3, and -4. GfEP-1 and -2 were shown to be generated by processing of a precursor of 1439 amino acids, and GfEP-3 turned out to be a partial fragment of GfEP-2. The precursor protein contained regions which exhibited similarities to vitellogenins (Vgs) in bilaterian animals (oviparous vertebrates and invertebrates including nematodes, arthropods, and molluscs). This study reports the first cloning and characterization of a full-length cDNA encoding a Vg in a non-bilaterian animal, and argues that the emergence of Vg as a precursor of egg yolk proteins predated the divergence of the cnidarian and bilaterian lineages.

Keywords: Vitellogenesis; Vitellogenin; Egg protein; Yolk; Coral reef; Scleractinia; Hermatypic coral; cDNA cloning; Precursor structure; Processing

Stony anthozoan and hydrozoan corals which are in obligatory symbiosis with the symbiotic dinoflagellate *Symbiodinium* spp. (generally known as zooxanthellae) contribute to the accretion of coral reefs and are thereby called hermatypic (reef-building) corals [1]. In the last few decades, significant decline of hermatypic coral populations has been observed in many areas of the world. The mass mortality of corals has been mainly attributed to: (i) global warming due to increase in the CO_2 level, and (ii) outbreaks of coral-eating animals such as the crown-of-thorns starfish [2,3]. Elevated sea temperature causes 'bleaching' (or loss of the symbionts) which is fatal to the host [4].

Sub-lethal effects of these changes of the environment include disturbance of the reproduction in corals. Gametogenesis was shown to be adversely affected in coral colonies that had experienced bleaching [5]. Eutrophication of reef waters is also likely to have negative impacts on gamete formation in corals [6]. Effects of other stressors (e.g., chemical pollutants) are yet to be examined.

Extensive histological analyses of gametogenesis have been performed in a number of coral species [7]. However, very few studies have been reported on molecular aspects of gametogenesis. We initiated molecular analysis of proteins accumulated in eggs of hermatypic corals (*Favites chinensis* (hermaphrodite) [8] and *Galaxea fascicularis* (gonochorist) [9]), with the purpose of utilizing the expression of their mRNAs to monitor the vitellogenic activity in assessing effects of stressors.

In the egg of *G. fascicularis*, four major protein species with the apparent molecular masses of approx. 88, 74, 70, and 32 kDa (named GfEP-1 to -4) were found [9]. We cloned a partial cDNA encoding GfEP-1 and showed that the expression level of its mRNA is significantly higher in females than in males [9]. A cDNA encoding a major egg protein (FcEP-1) in *F. chinensis* was also cloned [8]. No homology was found between the FcEP-1 and partial GfEP-1 sequences.

* Corresponding author. Fax: +81 3 5351 6534.
E-mail address: hayakawa@ori.u-tokyo.ac.jp (H. Hayakawa).

0006-291X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2006.03.116

In this report, we have isolated a cDNA which encodes a large precursor (named GfVg) for two of the major egg proteins (GfEP-1 and -2) in *G. fascicularis*, and show that GfEP-2 is homologous to FcEP-1. The amino acid sequence of the precursor exhibits similarities to those of vitellogenins (Vgs) in bilaterian animals.

Vgs are the precursors of proteins that are accumulated in the egg yolk and utilized as nutrients in embryonic development of non-mammalian vertebrates and some bilaterian invertebrates [10–12]. Comparison of Vg sequences in vertebrates, rhabditid nematodes, bivalve mollusks, crustaceans, and some insects suggested common evolutionary origin of the Vg genes in these diverse animal groups [13–16]. In some invertebrates such as dipteran insects and sea urchins, yolk proteins unrelated to Vgs are accumulated in eggs [17,18]. There was no report on a Vg in non-bilaterian animals (Porifera, Cnidaria, and Ctenophora).

This study reports the first cloning and characterization of a full-length cDNA encoding a Vg in a non-bilaterian animal.

Materials and methods

Cloning of the GfVg cDNA by screening of a cDNA library. Female colonies of *G. fascicularis* were collected 0–1 month before spawning on July, 2004 near Zampa Beach, Okinawa, Japan where the majority of the individuals had the 'hard' morphotype [19,20]. Internal soft tissues, which consisted mostly of the mesentery and mesenterial filament, were dissected, and total RNA was prepared as previously described [21], except that the homogenization with glass beads was omitted, and a cDNA library was constructed using a SMART cDNA library construction kit (Clontech) and Gigapack III Gold Packaging Extract (Stratagene). In order to identify cDNA clones containing coding sequences for GfEP-1 [9], approx. 300,000 plaques were screened using digoxygenin-labeled probes corresponding to the following two sequences: G^{4}–T^{219} (probe 1) and A^{1691}–A^{1905} (probe 2) of the partial *GfEP-1* cDNA sequence in Fig. 3 in [9]. The inserts in positive clones (*GfVg-1* and -2 shown in Fig. 1, see below) were sequenced using an ABI PRISM 3130 automatic sequencer (Applied Biosystems) or DSQ2000L automatic DNA sequencer (Shimadzu). To determine the sequence of the *GfVg-2* clone on both strands, six subfragments (2a–f) were generated (Fig. 1). Fragments 2a, b, c, and f were generated using restriction enzymes indicated in Fig. 1. Fragments 2d and 2e were generated by PCR using the following primer sets: 5′-ATTC AGTTTAAGGCGGTGACAC-3′ and 5′-GCGGTCATTGGTGACG TGAG-3′ (fragment 2d), and 5′-AGGCGATGGACTTAGATTGG-3′, and 5′-TAATACGACTCACTATAGGG-3′ (fragment 2e) (the latter primer sequence was derived from the pTriplEx2 vector).

Sequence analysis. Conceptual translation, calculation of molecular, mass and identification of potential N-glycosylation and phosphorylation sites were performed using the GENETYX software (ver. 7.0.11, GENETYX Co.). Identification of a signal peptide in GfVg and prediction of a cleavage site were performed using the SignalP program (ver. 3.0, [22]). A homology search in the UniProt + PRF + PDB database was conducted using the FASTA program (ver. 3.4t21) on the DNA Databank of Japan (DDBJ) website. A search of the Pfam databases [23] was done through a 'Motif search' program on GenomeNet website (Bioinformatics Center, Institute for chemical Research, Kyoto University, http://www.genome.ad.jp/).

N-terminal sequence analysis of GfEP-2, -3, and -4. SDS–PAGE of proteins extracted from eggs, and electrical blotting to PVDF membranes were performed as previously described [8,9], except that 8% polyacrylamide gels were used in SDS–PAGE to separate GfEP-2 (74 kDa) and -3 (70 kDa), and a 10% gel was used to separate GfEP-4 (32 kDa). Molecular weight standards (LMW Calibration Kit for SDS Electrophoresis (Amarsham) or Prestained SDS–PAGE standards, Broad Range (Bio-Rad)) were used for estimation of molecular masses of egg proteins. The 74, 70, and 32 kDa bands were visualized by Coomassie brilliant blue (CBB) staining, excised from a membrane, and applied to a PPSQ-21 Protein Sequencer (Shimadzu) to determine the N-terminal sequences of GfEP-2, -3, and -4, respectively.

Antibody generation and Western blot analysis. Synthesis of a peptide antigen and production of an antiserum were ordered to Operon Biotechnologies (Tokyo, Japan). Briefly, a peptide of 18 amino acids near the C-terminus of GfVg (CGDADGEQWNEYKDPQGR) was synthesized and coupled to bovine serum albumin. A rabbit was injected with the antigen, and the antiserum was collected. Proteins extracted from eggs and pseudo-eggs were separated by SDS–PAGE (a 8% polyacrylamide gel). Electrical blotting, incubation of the blot in antibody solutions, and signal detection were carried out as previously described [24] with the following modifications: the antiserum was diluted 1/1000 in the blocking solution, and goat anti-rabbit IgG alkaline phosphatase-conjugate (Chemicon, diluted 1/1000) was used as the secondary antibody.

Immunohistochemistry. Fixation, decalcification, embedment in paraffin, sectioning, and immuno-staining were carried out according to [25] with the following modifications. Dried skim milk (1%) was added to the saturating medium to reduce non-specific binding of the antibodies to tissues. The antiserum and the secondary antibody (see above) were diluted 1:500 in the saturating medium and signal detection was performed as described in [24].

Phylogenetic analysis. The amino acid sequence of GfVg was aligned with those of FcEP-1 ([8]) and vitellogenins of two vertebrates and three invertebrates (see Fig. 6) using Clustal W (ver.1.83 [26]) allowing gaps. The amino acid sequences of the two regions that were conserved among all of the seven proteins (M^{801}–Q^{1013} and D^{1369}–W^{1435} in GfVg and homologous regions in the other six proteins (Fig. 6B)) were merged, and the dataset (7 taxa, 293 positions) was subjected to maximum likelihood (ML) analysis with PhyML [27]. An input tree generated by BIONJ with the WAG model [28] of amino acid substitution was used, incorporating invariable

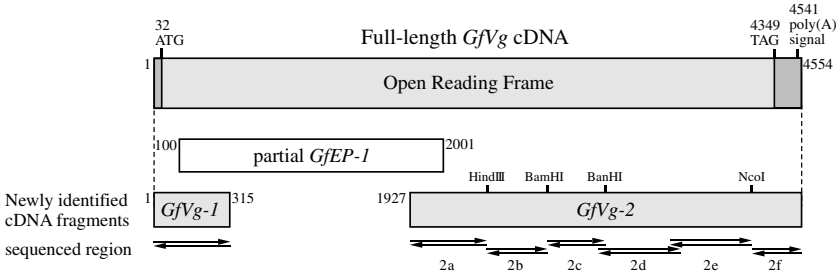

Fig. 1. Cloning of the *GfVg* cDNA. The previously reported partial *GfEP-1* sequence [9] is indicated by an open box, and the full-length *GfVg* cDNA and two newly identified fragments (*GfVg-1* and -2) by shaded boxes. Nucleotide positions (according to the numbering in Accession No. AB179781 in the GenBank/EMBL/DDBJ database) are indicated on the left and right of the boxes. The boundaries of the ORF are shown above the top box. Arrows below the *GfVg-1* and -2 boxes indicate nucleotide sequencing of *GfVg-1* and sub-fragments of *GfVg-2* (named 2a–f).

sites and a discrete γ-distribution (eight categories) (WAG + I + Γ). The proportion of invariable sites, a discrete γ-distribution, and base frequencies of the model were estimated from the dataset. PhyML bootstrap trees (100 replicates) were constructed using the same parameters as the individual ML trees (Fig. 6B).

Results and discussion

Identification of the GfVg cDNA and sequence analysis

A cDNA library of internal soft tissues of *G. fascicularis* was screened to determine the nucleotide sequence of the full-length open reading frame (ORF) containing the coding sequence for GfEP-1. Two probes (probes 1 and 2) that were derived from 5′ and 3′-regions of the partial GfEP-1 cDNA sequence, respectively, were used. Two clones (*GfVg-1* and -2) which contain sequences 5′ and 3′ to the partial cDNA sequence, respectively, were thereby identified, and their nucleotide sequences were determined (Fig. 1). Six sub-fragments (2a–f) of *GfVg-2* were generated by restriction digestion or PCR and sequenced, and their nucleotide sequences were merged. A 3′ part of the *GfVg-1* (216 bp) and a 5′ part (95 bp) of *GfVg-2* overlapped with the partial *GfEP-1* sequence [9] with 100% identity (Fig. 1). The three sequences were thereby merged into a single sequence of 4554 bp (Accession No. AB179781 in the GenBank/EMBL/DDBJ databases).

The merged cDNA sequence contained an open reading frame (ORF) for 1439 amino acids, as well as putative untranslated regions (UTRs) of 31 and 203 bp on the 5′ and 3′-sides, respectively. To address the possibility whether the merged ORF was a chimera consisting of sequences derived from different splicing products or related genes, total RNA from internal soft tissues was subjected to RT-PCR with a pair of primers derived from the putative 5′ and 3′-UTRs (G^{1}–C^{21} and C^{4357}–A^{4376}). The size of an amplified product (approx. 4.3 kb) was close to the size (4376 bp) expected from the sequence shown in Fig. 2 (data not shown). To show that the product contained the *GfVg* ORF sequence, it was used as the template in a second PCR with primers derived from the ORF (G^{100}–A^{119} and C^{3680}–C^{3699}), and a band (approx. 3.6 kb) of the expected size (3600 bp) was observed (data not shown). This result was in

1 *MNPFWVILLSLGAVSS***LHLGYQ**K**EQRY**TYSYQTEASQAIPGEAKKSLGLRIKCQVHIDVLKKMEMQATLDNCQVHQVQND
GfEP-1 N-terminus

81 SKEEQETPADHQEAMRRKLAKPVRFSFKVQDGSLKEISAHPNDSEWSLNIKRGLVNLFQISYQAKQQENVAYSISRNKSP

161 QQSVTAKVVYVINRDNSLPANNEQVVKLTKVMNYEKCSKRPEFSQSNYYGQQCQECQQSVGYAPYHRSAGEIEHTLTGSE

241 DAGYVINRTEALEQHVITPLSQSAGHITVTTRQILQFQKKSVAGAASLELTKRIGQAFTAESHEKCGRDDTQRQKQIAQQ

321 AETSINEMCKAFKDPASRHAAEKFAVAVQQLRKCNQDTMERLTEIMLARQDQTERKVCLDALSFVGTREAFQVLQQKVQQ

401 KKIKSSDDLQRISLGLASAPRPTADHIDAAWKLCEEVRKDSSCRRQCLLSLGVLIYKGSRYQQDTKTVVACQQRVREMLE

481 KLSNTATSQEDRLTFIKTLGNAGSPDAQEQLQKILKDRQQPLHTRVECVWALRRITRQAKEKTYPCLISIFADRQENPEL

561 RMATFVQLLNTEPNFVILQALTNTVRREINDPRPGRRSNQLASFVISHLSALAYHNNVLTKKRSMQARMALRLLPKMSFG

641 LSYSKGMRLVQHSEKLQAGFEFEANKIDVPESMLPRNLNARLQVNLFGYKMNSLEFGARIENMDDVVDDIISQVRERHKR

721 SLWGSVVSWFSKQQSDKDTKVTEPPTKRSTGRSVRNTGATSRPLCQSSFQPTVSRRKSTRPPPTFQAS**DEQTKQRHMSAF**
GfEP-2 N-terminus

801 **MK**LYGNEVKFMEVKQDVVDQLARDVSDLLLGKVNQLEYPGEGITIS**KGGVEIKQTLEKAFMDANAQ**NVLPTLAGIPLNIQ
GfEP-3 N-terminus

881 FRSAASVRITAGANLDVHPSLKTFWQYQSMTGNVEVKPSINTHVHAVVAIHTPFLRVGVQMQANGNSNTSLKIEVQAEAE

961 KSCNFGYSIPQEQKDIIQFKAVTQGFSQQQDPQTCEQKEKQISLDLSKDHLTQLKKSCAGQALFGVQLCAEGQGPDLASL

1041 RLRQVPVFSAICQAEMRVSLAPASDSPLLIQWNNNIEKNDDKEKDVSGEIKVSSNKITRRLPYRITYKRDQQQLEIGSTS

1121 LVAQGYEDAKMKLIAKGDGLRLEFGRKGKPQYSVTVSGQIENQGKSLKVQANWTEVPEQWKTFFRMWEPQILYALQQFAW

1201 VRRTDQQTKQVALEFQLTSPMTASSTVKTPNAEAKRTKLSLPICVERLPTSIQEIKDHLYAKCEVQEQSIKVFDQLRYQH

1281 NIKGGCPYVLAQEYREGKQSRFELTVKTDEQGQKTLIASFQQQSQQHKETVEIKPDLTILIDGQKQTCAQQPCRSKQGDF

1361 TVRTVQTTDGKTEIQLSTKLGLHVTMEGQRIHVYVAPFLCGRVRGLCGDADGEQWNEYKDPQGRVQTLQKFIQSWQQKC* 1439
C-terminal antigen

Fig. 2. Deduced amino acid sequence of GfVg. A putative signal peptide sequence is in italics. The N-termini of GfEP-1 [9], -2, and -3 (this study) are indicated in the figure, and the amino acid sequences consistent with the result of N-terminal sequence analyses are in bold. Two consensus cleavage signal sequences (R–X–X–R) recognized by subtilisin family endoproteases are boxed (black, closed). Putative N-glycosylation sites are boxed (open). A peptide sequence of the antigen for the C-terminal antiserum is boxed in broken lines. The partial GfEP-1 sequence in our previous report [9] corresponded to Leu^{17}-Glu^{671}.

support of the possibility that the ORF sequence was derived from a single mRNA.

A computer-aided analysis of the GfVg sequence predicted a signal cleavage site between Ser^{16} and Leu^{17}. This prediction was consistent with the observation that Leu^{17} corresponded to the N-terminus of GfEP-1 ([9] and Fig. 2 of this study). There were five potential N-glycosylation sites (Fig. 2) and 54 phosphorylation sites (data not shown) in the GfVg sequence. As described below, the translated protein sequence exhibited similarities to Vg sequences in higher animals, and was therefore named GfVg (*G. fascicularis* Vg).

N-terminal sequencing of GfEP-2, -3, and -4

The predicted molecular mass for Leu^{17}-Cys^{1439} was 161,304 Da—a value much larger than the apparent molecular mass of GfEP-1 (approx. 88 kDa on SDS–PAGE) [9]. This observation raised the possibility that the ORF encoded a large precursor containing GfEP-1 and an additional protein(s) that was separated from GfEP-1 by processing. To address the possibility that one or more of GfEP-2, -3, and -4 were encoded by the same precursor as GfEP-1, the N-terminal sequences of the three proteins were determined and compared to the GfVg sequence (Fig. 3).

Edman degradation of GfEP-2 yielded mixed signals at 14 positions (Fig. 3), suggesting that an additional protein(s) was present in the slice of the membrane containing the 74-kDa band. In the mixed sequence, however, a stretch of amino acids that was identical to Asp^{789}-Lys^{802} of GfVg could be found (Figs. 2 and 3). At most of the positions, the amino acids identical to the partial GfVg sequence represented the strongest signal. This result argued for the possibility that GfEP-1 and -2 were separated by cleavage of the GfVg precursor on the N-terminal side of Asn^{789}. Assuming processing at this site, the calculated molecular masses of the two resulting polypeptides, Leu^{17}-Ser^{788} and Asp^{789}-Cys^{1439}, were 87,888 and 73,433 Da (Fig. 4), and very close to the apparent molecular masses of GfEP-1 and -2 (88 and 74 kDa), respectively.

Similarly, in analyzing GfEP-3 a sequence identical to Lys^{847}-Gln^{866} of GfVg was found (Figs. 2 and 3). Cleavage on the N-terminal side of Lys^{847} would generate two polypeptides of 94,514 Da (Leu^{17}-Ser^{846}) and 66,808 Da (Lys^{847}-Cys^{1439}) (Fig. 4). Thus, GfEP-3 was likely to be a partial segment of GfEP-2.

The GfEP-4 protein (approx. 32 kDa) exhibited a broad banding pattern [9], suggesting the possibility that the band consisted of more than one protein. Therefore, the band blotted onto a membrane was cut into three slices, each of which was subjected to sequence analysis. All of the slices contained an identical sequence (Fig. 3), which did not show similarity to the GfVg sequence, indicating that GfEP-4 was not encoded by the *GfVg* mRNA.

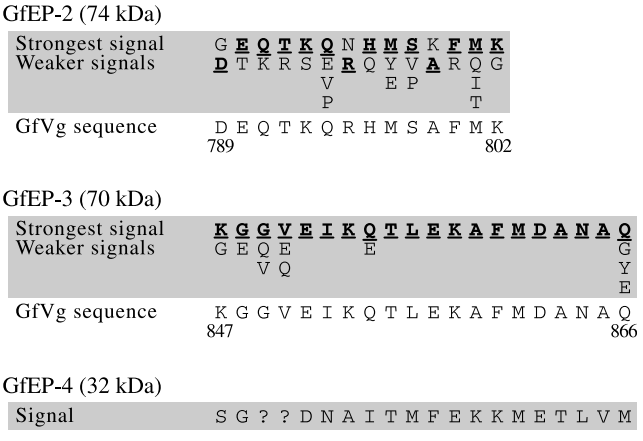

Fig. 3. N-terminal amino acid sequences of GfEP-2, -3, and -4. Mixed amino acid signals at the same positions are shown vertically (shaded gray). The strongest signals are shown in the top rows and weaker signals in lower rows. Amino acid residues in underlined bold fonts are consistent with sequences deduced from the *GfVg* cDNA (the corresponding GfVg sequences with the position numbers are shown at the bottom). A question mark indicates a position where the significant amino acid signal was not detected.

Western analysis and immunohistochemistry

To further address the possibility that GfEP-2 and -3 are derived from C-terminal parts of GfVg, an antiserum was raised against a peptide antigen which was designed based on a sequence near the C-terminus of GfVg (Fig. 2), and its reaction with GfEP-2 and -3 was examined in a Western blot analysis. As shown in Fig. 4B, immunoreactive bands were observed at 74 and 70 kDa. This observation supported the above possibility that GfEP-2 and -3 were encoded by the *GfVg* mRNA.

Next, to study distribution of GfEP-2 and -3 in developing oocytes, female mesenteries containing ovaries were sectioned and subjected to an immunohistochemical analysis using the antiserum. Immunological signals were detected throughout the cytoplasm of oocytes but not in the germinal vesicle (Fig. 4C). The staining in the ooplasm was not observed in a control experiment using a preimmune serum as the primary antibody (data not shown). This result indicated broad distribution of GfEP-2 and/or -3 in the ooplasm, and was consistent with the possibility that these proteins were associated with yolk granules.

Processing of GfVg

Based on the above results, we propose that the GfVg polypeptide constitutes a precursor that is processed at one or two sites (in addition to the signal cleavage site) to generate GfEP-1, -2, and -3 (Fig. 5). Processing at site 1 separates GfEP-1 (88 kDa) and -2 (74 kDa), and cleavage at site two generates GfEP-3 (70 kDa). It is likely that when GfEP-3 is generated, processing at site 1 also takes place, since no band was detected around 94.5 kDa by CBB staining [9].

Currently, it is unknown what processing enzymes cleave the GfVg precursor protein. Vgs in vertebrates are

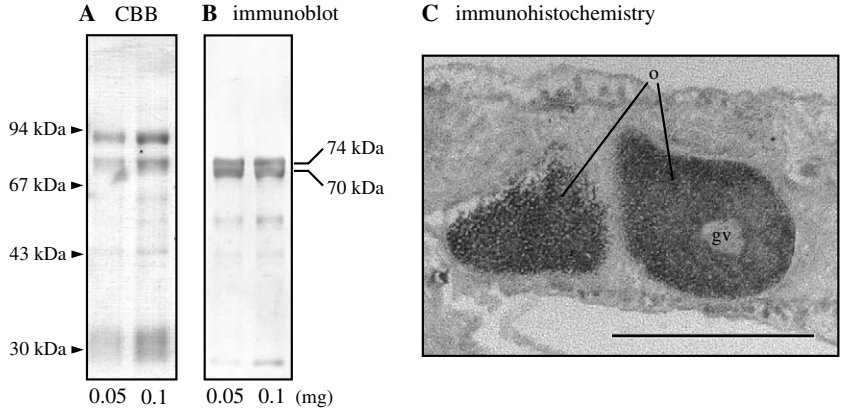

Fig. 4. Western analysis and immunohistochemistry of GfEP-2 and -3. (A) Egg proteins extracted from 0.05 and 0.1 mg of eggs were separated using SDS–PAGE with an 8% polyacrylamide gel and visualized with CBB. Molecular masses of markers are indicated on the left. (B) Immunodetection of GfEP-2 (74 kDa) and -3 (70 kDa) on a Western blot. (C) Immunostaining of a tissue section through a mesentery in a female approx. 1 month before spawning. Signals were detected in the cytoplasm, but not in the germinal vesicle (gv), of oocytes (o). Scale bar = 200 μm.

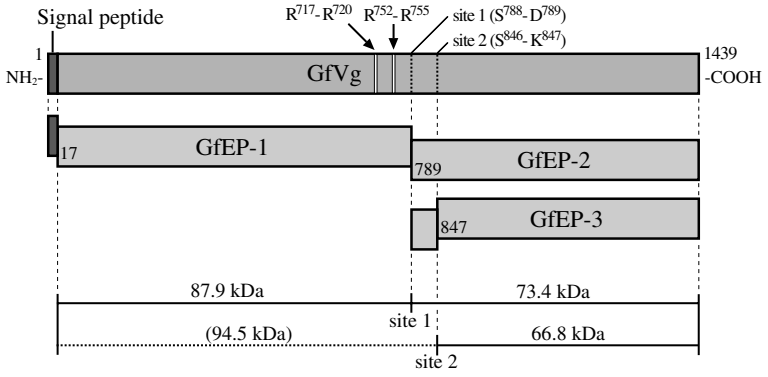

Fig. 5. Proposed structure and processing of the GfVg precursor. The GfVg precursor, the putative signal peptide, GfEP-1, -2, and -3 are indicated with boxes. The two R–X–X–R motifs are indicated by arrows. The calculated molecular masses of proteins assuming processing at site 1, 2 and the predicted signal cleavage site are shown at the bottom. Numbers on both sides of the top box are the amino acid positions.

processed by cathepsin D [29]. In insects and crustaceans cleavage sites in Vgs are located C-terminally to R–X–X–R sequences, strongly suggesting that the processing enzymes belong to the family of subtilisin-like proprotein convertases [30]. The GfVg sequence contained two R–X–X–R sequences (Figs. 2 and 5), but those sites were more than 35 amino acids upstream from the cleavage sites. No similarity could be found between the amino acid sequences surrounding the two cleavage sites in GfVg, except that the sites are located C-terminally to serine.

Homology search and molecular phylogenetic analysis of GfVg

In a database search for proteins related to GfVg, the closest match was FcEP-1, one of the two abundant egg proteins in *F. chinensis* [8]. FcEP-1 was aligned with the C-terminal 653 amino acids of GfVg with 52% identity, and the N-terminus of FcEP-1 corresponded to that of GfEP-2 (data not shown). Sequence similarities were also found between GfEP-2 and Vgs in vertebrates and bilaterian invertebrates. In aligning GfEP-2 with FcEP-1 and Vg sequences of vertebrates and bilaterian invertebrates (a nematode, bivalve, and shrimp), two conserved regions were found (Fig. 6A). Using the sequences in these regions, a molecular phylogenetic analysis was performed to infer the evolutionary relationship (Fig. 6B). GfVg and FcEP-1 clustered with the shrimp Vg. In our previous study, sequence similarities were also found between the partial GfEP-1 sequence and the N-terminal regions of Vgs in higher animals [9]. These observations argue that GfVg is a member of the superfamily consisting of Vgs of oviparous vertebrates and some bilaterian invertebrates [13], and an ancestral Vg as a precursor of yolk proteins emerged before the divergence of the cnidarian and bilaterian lineages.

To date, a full-length cDNA encoding FcEP-1 has not been isolated, and therefore it has not been determined whether the two abundant egg proteins in *F. chinensis* (FcEP-1 and -2) are generated by processing of a GfVg-like precursor. It also remains unknown whether FcEP-2 and GfEP-1 are homologous, because no similarity could be

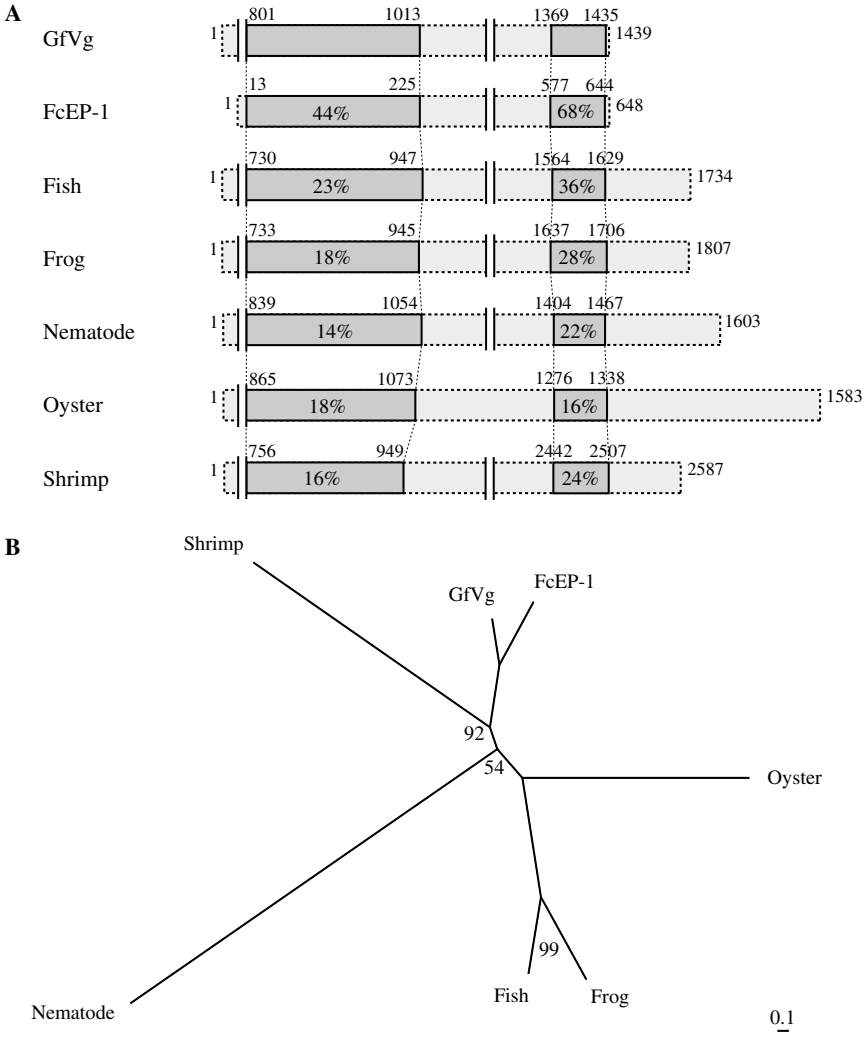

Fig. 6. Phylogenetic analysis of GfVg and Vgs in vertebrates and invertebrates. (A) The following proteins were aligned with a C-terminal part of GfVg: FcEP-1 (Accession No. AB109564 in the GenBank/EMBL/DDBJ database [8]), and Vgs in the Japanese eel *Anguilla japonica* (Vg I, AY423445), the African clawed frog *Xenopus laevis* (Vg A2, Y00354 [39]), the Pacific oyster *Crassostrea gigas* (AB084783 [15]), the nematode *Caenorhabditis elegans* (Vg 5, X03044 [37]), and the European white shrimp *Litopenaeus vannamei* (AY321153 [36]). Shaded parts of the bars indicate the two conserved regions (see text). Numerals above and right side of the boxes indicate the amino acid positions in the respective proteins at the boundaries of the conserved regions and the lengths of amino acid sequences, respectively. Values inside the boxes indicate levels of amino acid identity (in percentage). (B) An unrooted phylogram based on the sequences of shaded regions in (A) using the maximum likelihood method. The numbers at nodes indicate the bootstrap values that were calculated from 100 replications. The scale bar represents 0.1 mutation per site.

found between the N-terminal sequences of the two proteins [8].

To further address the issue of the homology between the cnidarian Vg and bilaterian Vgs, a search of the Pfam database based on hidden Markov model profiles [23] was also performed on GfVg. Two regions (Y^{21}–S^{610}, and C^{1263}–R^{1424}) of GfVg exhibited highly significant matches to the following two families, respectively: lipoprotein amino terminal region (Accession No. PF01347, $E = 3.3e^{-47}$) and von Willebrand factor type D (vWD) domain (PF00094, $E = 8.2e^{-13}$). The former family contains N-terminal regions of proteins involved in lipid transport such as Vg, microsomal triglyceride transfer protein, and apolipoprotein B-100 [31]. The vWD domain has also been found in vertebrate, nematode, and arthropod Vgs near the C-termini [31]. Thus, the result of the Pfam search was consistent with the proposed homology between GfVg and the bilaterian Vgs.

In the GfVg sequence, no serine-rich regions could be found. Vgs of oviparous vertebrates are known to contain the phosvitin domain which is highly rich in Ser residues [32,33]. Due to the high degree of phosphorylation of Ser residues, this domain is capable of binding large amounts of calcium ion, and speculated to carry Ca^{2+} to oocytes to support formation of calcified hard tissues (e.g., bones) in embryonic development. Similar polyserine regions are found in Vgs of some insects, but not in Vgs of crustaceans that were characterized to date [13,34–36]. The Ser-rich domains/regions in vertebrates and insects are likely to have evolved independently, because their positions are

not conserved [13]. No such regions were found in Vgs of nematodes [14,37,38]. Thus, the absence of a Ser-rich region in GfVg does not necessarily argue against the homology between the coral protein and the Vgs in higher animals. The requirement for storage of Ca^{2+} in coral eggs should be small, since their larvae are planktonic and lack calcified structures.

In summary, the precursor structure of major egg proteins was elucidated and their amino acid sequences inferred for the first time in a cnidarian (or non-bilaterian animal). Results of the homology and motif searches, as well as molecular phylogenetic analysis, argue for emergence of the Vg as an egg protein precursor before the cnidarian-bilateria divergence.

Acknowledgments

The authors are grateful to Dr. N. Ozaki (University of Tsukuba) for assistance in construction of the cDNA library. This work was supported in part by Grants-in-Aid (14042208) from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] J.E.N. Veron, Corals of the World, Australian Institute of Marine Science, Townsville, Queensland, 2000.
[2] J.B.C. Jackson, M.X. Kirby, W.H. Berger, K.A. Bjorndal, L.W. Botsford, B.J. Bourque, R.H. Bradbury, R. Cooke, J. Erlandson, J.A. Estes, T.P. Hughes, S. Kidwell, C.B. Lange, H.S. Lenihan, J.M. Pandolfi, C.H. Peterson, R.S. Steneck, M.J. Tegner, R.R. Warner, Historical overfishing and the recent collapse of coastal ecosystems, Science 293 (2001) 629–638.
[3] T.P. Hughes, A.H. Baird, D.R. Bellwood, M. Card, S.R. Connolly, C. Folke, R. Grosberg, O. Hoegh-Guldberg, J.B.C. Jackson, J. Kleypas, J.M. Lough, P. Marshall, M. Nystrom, S.R. Palumbi, J.M. Pandolfi, B. Rosen, J. Roughgarden, Climate change, human impacts, and the resilience of coral reefs, Science 301 (2003) 929–933.
[4] B.E. Brown, Coral bleaching: causes and consequences, Coral Reefs 16 (1997) S129–S138.
[5] J.M. Mendes, J.D. Woodley, Effect of the 1995–1996 bleaching event on polyp tissue depth, growth, reproduction and skeletal band formation in *Montastraea annularis*, Mar. Ecol. Prog. Ser. 235 (2002) 93–102.
[6] K.E. Fabricius, Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis, Mar. Poll. Bull. 50 (2005) 125–146.
[7] P.L. Harrison, C.C. Wallace, Reproduction, dispersal and recruitment of scleractinian corals, in: Z. Dubinsky (Ed.), Coral Reefs: Ecosystems of the World, vol. 25, Elsevier, Amsterdam, 1990, pp. 133–207.
[8] S. Imagawa, Y. Nakano, T. Watanabe, Molecular analysis of a major soluble egg protein in the scleractinian coral *Favites chinensis*, Comp. Biochem. Physiol. B. 137 (2004) 11–19.
[9] H. Hayakawa, Y. Nakano, T. Andoh, T. Watanabe, Sex-dependent expression of mRNA encoding a major egg protein in the gonochoric coral *Galaxea fascicularis*, Coral Reefs 24 (2005) 488–494.
[10] J.L. Specker, C.V. Sullivan, Vitellogenesis in fishes: status and perspectives, in: K.G. Davey, R.G. Peter, S.S. Tobe (Eds.), Perspectives in Comparative Endocrinology, National Research Council of Canada, 1994, pp. 304–315.
[11] T.S. Adams, Arthropoda: insecta, in: Progress in Reproductive Endocrinology, in: T.S. Adams (Ed.), Reproductive Biology of Invertebrates, vol. 8, Wiley, Chichester, 1997, pp. 277–338.
[12] T.W. Sappington, A.S. Raikhel, Molecular characteristics of insect vitellogenins and vitellogenin receptors, Insect Biochem. Mol. Biol. 28 (1998) 277–300.
[13] J.S. Chen, T.W. Sappington, A.S. Raikhel, Extensive sequence conservation among insect, nematode, and vertebrate vitellogenins reveals ancient common ancestry, J. Mol. Evol. 44 (1997) 440–451.
[14] C.E. Winter, C. Penha, T. Blumenthal, Comparison of a vitellogenin gene between two distantly related rhabditid nematode species, Mol. Biol. Evol. 13 (1996) 674–684.
[15] T. Matsumoto, A.M. Nakamura, K. Mori, T. Kayano, Molecular characterization of a cDNA encoding putative vitellogenin from the Pacific oyster *Crassostrea gigas*, Zool. Sci. 20 (2003) 37–42.
[16] M. Osada, M. Harata, M. Kishida, A. Kijima, Molecular cloning and expression analysis of vitellogenin in scallop, *Patinopecten yessoensis* (Bivalvia, Mollusca), Mol. Reprod. Dev. 67 (2004) 273–281.
[17] M. Bownes, Why is there sequence similarity between insect yolk proteins and vertebrate lipases? J. Lipid Res. 33 (1992) 777–790.
[18] T. Unuma, H. Okamoto, K. Konishi, H. Ohta, K. Mori, Cloning of cDNA encoding vitellogenin and its expression in red sea urchin, *Pseudocentrotus depresses*, Zool. Sci. 18 (2001) 559–565.
[19] M. Hidaka, Use of nematocyst morphology for taxonomy of some related species of scleractinian corals, Galaxea 11 (1992) 21–28.
[20] T. Watanabe, M. Nishida, K. Watanabe, D.S. Wewengkang, M. Hidaka, Polymorphism in the nucleotide sequence of a mitochondrial intergenic region in the scleractinian coral *Galaxea fascicularis*, Mar. Biotechnol. 7 (2005) 33–39.
[21] I. Yuyama, H. Hayakawa, H. Endo, K. Iwao, H. Takeyama, T. Maruyama, T. Watanabe, Identification of symbiotically expressed coral mRNAs using a model infection system, Biochem. Biophys. Res. Commun. 336 (2005) 793–798.
[22] J.D. Bendtsen, H. Nielsen, G. Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, J. Mol. Biol. 340 (2004) 783–795.
[23] E.L.L. Sonnhammer, S.R. Eddy, R. Durbin, Pfam: a comprehensive database of protein domain families based on seed alignments, Proteins 28 (1997) 405–420.
[24] H. Endo, Y. Takagi, N. Ozaki, T. Kogure, T. Watanabe, A crustacean Ca^{2+}-binding protein with a glutamate-rich sequence promotes $CaCO_3$ crystallization, Biochem. J. 384 (2004) 159–167.
[25] S. Puverel, E. Tambutté, D. Zoccola, I. Domart-Coulon, A. Bouchot, S. Lotto, D. Allemand, S. Tambutté, Antibodies against the organic matrix in scleractinians: a new tool to study coral biomineralization, Coral Reefs 24 (2005) 149–156.
[26] J.D. Thomopson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[27] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (2003) 696–704.
[28] S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach, Mol. Biol. Evol. 18 (2001) 691–699.
[29] H. Retzek, E. Steyrer, E.J. Sanders, J. Nimpf, W.J. Schneider, Molecular cloning and functional characterization of chicken cathepsin D, a key enzyme for yolk formation, DNA Cell Biol. 11 (1992) 661–672.
[30] Y. Rouille, S.J. Duguay, K. Lund, M. Furuta, Q.M. Gong, G. Lipkind, A.A. Oliva, S.J. Chen, D.F. Steiner, Proteolytic processing mechanisms in the biosynthesis of neuroendocrine peptides—the subtilisin-like proprotein convertases, Front. Neuroendocrinol. 16 (1995) 322–361.
[31] P.J. Babin, J. Bogerd, F.P. Kooiman, W.J.A. van Marrewijk, D.J. van der Horst, Apolipophorin II/I, apolipoprotein B, vitellogenin, and microsomal triglyceride transfer protein genes are derived from a common ancestor, J. Mol. Evol. 49 (1999) 150–160.

[32] W. Wahli, Evolution and expression of vitellogenin genes, Trends Genet. 4 (1988) 227–232.
[33] B.M. Byrne, M. Gruber, G. AB, The evolution of egg yolk proteins, Prog. Biophys. Mol. Biol. 53 (1989) 33–69.
[34] A. Okuno, W.J. Yang, V. Jayasankar, H. Saido-Sakanaka, T.T. do Huong, S. Jasmani, M. Atmomarsono, T. Subramoniam, N. Tsutsui, T. Ohira, I. Kawazoe, K. Aida, M.N. Wilder, Deduced primary structure of vitellogenin in the giant freshwater prawn, *Macrobrachium rosenbergii*, and yolk processing during ovarian maturation, J. Exp. Zool. 292 (2002) 417–429.
[35] N. Tsutsui, I. Kawazoe, T. Ohira, S. Jasmani, W.J. Yang, M.N. Wilder, K. Aida, Molecular characterization of a cDNA encoding vitellogenin and its expression in the hepatopancreas and ovary during vitellogenesis in the kuruma prawn, *Penaeus japonicus*, Zool. Sci. 17 (2000) 651–660.
[36] S. Raviv, S. Parnes, C. Segall, C. Davis, A. Sagi, Complete sequence of *Litopenaeus vannamei* (Crustacea: Decapoda) vitellogenin cDNA and its expression in endocrinologically induced sub-adult females, Gen. Comp. Endocrinol. 145 (2006) 39–50.
[37] J. Spieth, K. Denison, E. Zucker, T. Blumenthal, The nucleotide sequence of a nematode vitellogenin gene, Nucleic Acids Res. 13 (1985) 7129–7138.
[38] J. Spieth, M. Nettleton, E. Zucker-Aprison, K. Lea, T. Blumenthal, Vitellogenin motifs conserved in nematodes and vertebrates, J. Mol. Evol. 32 (1991) 429–438.
[39] S. Gerber-Huber, D. Nardelli, J.A. Haefliger, D.N. Cooper, F. Givel, J.E. Germond, J. Engel, N.M. Green, W. Wahli, Precursor-product relationship between vitellogenin and the yolk proteins as derived from the complete sequence of a *Xenopus* vitellogenin gene, Nucleic Acids Res. 15 (1987) 4737–4760.