Developmental and Comparative Immunology 34 (2010) 465–473

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

In search of the origin of FREPs: Characterization of *Aplysia californica* fibrinogen-related proteins

A.M. Gorbushin [a,*], Y.V. Panchin [b,c], N.V. Iakovleva [a]

[a] *Institute of Evolutionary Biochemistry and Physiology RAS, pr. Torez 44, Saint-Petersburg 194223, Russia*
[b] *Institute for Information Transmission Problems RAS, 127994 Moscow, Russia*
[c] *Belozersky Institute for Physicochemical Biology, Moscow State University, Russia*

ARTICLE INFO

Article history:
Received 31 October 2009
Received in revised form 14 December 2009
Accepted 14 December 2009
Available online 28 December 2009

Keywords:
Innate immunity
Lectins
Phylogeny
Somatic mutations
Retrotransposition
Retrogenes
Lottia gigantea
Mollusca
Heterobranchia

ABSTRACT

All haemolymph lectins with uniquely juxtaposed N-terminal domain similar to the immunoglobulin superfamily (IgSF) and C-terminal fibrinogen (FBG) termed FBG-related proteins (FREP) are documented till now only in the pulmonate mollusc *Biomphalaria glabrata*. Using genomic WGS database we have found two FREP genes from marine opistobranch *Aplysia californica* named AcFREP1 and AcFREP2. The AcFREP1 and AcFREP2 mRNA molecules have been subsequently isolated from cDNA of sea hare larvae as well as adult mollusc tissues. These genes encode proteins (504 and 510 aa respectively) with domain architecture typical for FREPs with two N-terminal IgSF domains and C-terminal FBG domain. Although cDNA sequences of AcFREP1 and AcFREP2 are 81% identical, their genomic structure is entirely different: AcFREP1 is intronless and AcFREP2 is encoded in four exons. These genes are paralogous pair in which AcFREP2 is a parental gene and AcFREP1 is the new transposed copy that has lost the introns (retrogene). Using RT-PCR analysis, expression of AcFREP1 and AcFREP2 was shown to be developmentally and tissue-specific and no constitutive expression in haemocytes was found. The overall frequency of nucleotide substitutions in genomic DNA trace sequences of coding region of the AcFREP1 and AcFREP2 is not higher than in the sequences of control conserved genes (actin, FMRFamide). Thus, previously reported high diversification of *Biomphalaria* FREP gene, BgFREP3, is not detected in *Aplysia* FREPs. A search for FREP homologs in other available complete genome of mollusc, *Lottia gigantea* (Patellogastropoda), a representative of the evolutionary earliest gastropod clade, did not reveal any DNA sequences coding for similar lectins. We suggest that unique domain architecture of FREPs is an evolutionary novelty that appeared and evolved only within one branch of Protostomata species, exclusively in heterobranch molluscs (Pulmonata and Opistobranchia).

1. Introduction

Core immune mechanisms, shared among Metazoa, are triggered by the receptors termed pattern recognition proteins (PRPs) – molecules that are competent in recognition of pathogen-associated molecular structures. In innate immunity, lectins are among the evolutionary conserved PRPs. These molecules are involved in self-/non-self-recognition by more or less specific binding of glycoproteins with the appropriate carbohydrate moieties on the surface of potential pathogen.

Several lectins containing fibrinogen-related domain (FreD) have been reported as PRPs from invertebrate animals, in Arthropoda such as tachilectins from the horseshoe crab, *Tachypleus tridentatus* [1], fibrinogen-related proteins OMFREP, Ixoderin A, B from ticks *Ornithodoros moubata* and *Ixodes ricinus* [2], fibrinogen-related proteins from mosquitoes *Anopheles*, *Armigeres* and *Aedes* [3–6] in Echinodermata, FreD-containing sequence from sea cucumber *Parastichopus parvimensis* [7] in Urochordata, ficolins from the ascidian *Halocynthia roretzii* [8]. All these lectins contain a typical C-terminal domain with high sequence similarity to that of the human fibrinogen β/γ chain (called FBG domain), but differ in their N-terminus.

A unique juxtaposition of N-terminal domain with similarity to the immunoglobulin superfamily (IgSF) and C-terminal FBG was shown in mollusc haemolymph lectins termed FREPs. Since their discovery and the eventual cloning in 1997 [9] FREPs have been a target for immunobiological research owing to three distinct characteristics: upregulation following relatively specific immunostimulation [10], binding to pathogens and precipitation antigens present in the excretory–secretory products of specific parasite [9,11], relatively high diversity and somatic diversification [12–14]. Considerable progress has since been made in terms of revealing

* Corresponding author. Fax: +7 812 5523012.
E-mail address: agorbushin@gmail.com (A.M. Gorbushin).

0145-305X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2009.12.007

gene structure, mRNA expression, diversification and functional roles of 14 currently known IgSF-containing FREP genes. However, one of the gaps in our current knowledge of FREPs is the absence of data on their evolutionary history. When did the IgSF and fibrinogen domains first juxtapose to form the single polypeptide that was identified as a FREP? To help overcome this gap, we have searched genome databases of two gastopod molluscs *Lottia gigantea* and *A. californica* (http://www.jgi.doe.gov and http://www.ncbi.nlm.nih.gov). Both species are marine and phylogenetically distant from *Biomphalaria glabrata*, the only species with cloned and characterized FREP lectins. Placing these data in their phylogenetic context is important to understand the FREP genes evolution.

We were also interested in determining whether or not elevated rate of experimentally observed nucleotide substitutions reported for the FREP3 gene in *B. glabrata* is true for other molluscan species. Characterization of the genes involved in mollusc immunity is essential for elucidating the mechanisms of innate immunity formation and developing new approaches to control pathogens that are transmitted by molluscs.

2. Materials and methods

2.1. A. californica and L. gigantea genome searching

Amino acid sequences containing several IgSF and FBG domains from *B. glabrata* FREPs were used as a tBLASTn query against NCBI *A. californica* (taxid:6500) unpublished whole genome shotgun reads (WGS; 7× coverage). Joint Genome Institute (JGI) interactive *L. gigantea* genome BLAST (Lotgi1 release; 8.87× coverage: http://genome.jgi-psf.org/) searches were conducted to retrieve potential FREP-like sequences from the genome, predicted proteins and EST base. Matching contigs having even weak similarity to juxtaposed IgSF and FreD domains were retrieved from the nucleotide database and the coding region was preliminary identified by several invertebrate gene prediction models using FGeneSH software (http://linux1.softberry.com/berry.phtml). Additionally, 70 *L. gigantea* predicted proteins, containing fibrinogen C-terminal domain-like superfamily domain were retrieved from structural and functional protein annotations database SUPERFAMILY (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/).

2.2. Cloning of cDNA encoding A. californica FREPs

Standard molecular biology procedures were used for cDNA cloning and sequencing as described by Moroz et al. [15]. RACE procedure is also described elsewhere [16].

In the *A. californica* WGS database we identified two contigs that had significant similarity to *B. glabrata* FREPs. Gene-specific primers were designed for amplification of 951 bp fragment of AcFREP1 (AcFREP1f—CGGCCTCAGTGAATGTAAGTAACT; AcFREP1r—CCCATGCGATACTCTTGAATCC) and 1140 bp fragment of AcFREP2 (AcFREP2f—CGATCCGAGCCAAACAGGAACGTA; AcFREP2r—TCGGCCCACCAGATACCTTCAA). The control samples for each tissue were run to amplify the fragment of actin cDNA (Acf—TGAGCGTGAAATTGTCCGTGA; Acr—GCCTGGGAACATGGTGGAAC) to check the RNA integrity and quantity. To study the AcFREP mRNA expression and tissue distribution *A. californica* total RNA was isolated from post-metamorphic early juveniles (8–9 larval stages according to [17]) and adult mollusc organ parts and haemocytes. The samples (1 μg) were subjected to DNAse I treatment (Fermentas, Lithuania) to avoid false-positive results due to genomic DNA contamination. DNAse-treated RNA samples were tested for absence of gDNA remains in PCR without reverse transcription and used for cDNA synthesis for "adult" samples using M-MuLV reverse transcriptase (Fermentas, Lithuania) and for "larval" samples using the Marathon cDNA amplification kit (Clontech, USA) according to manufacturer's instructions. To ascertain the full size of mRNA transcript the rapid amplification of cDNA ends (RACE) using standard primers for adapter sequence and gene-specific primer was performed. The cycling conditions were as follows 95 °C for 3 min (95 °C for 15 s, 60 °C for 30 s, 68 °C for 2 min) 35 cycles, 68 °C for 10 min. All PCR reactions were performed on Eppendorf Mastercycler (Eppendorf, Germany). Amplified products corresponding to the 3′- and 5′-ends of the cDNA were cloned into pTZ57R/T vector (InsTAclone PCR cloning kit, Fermentas, Lithuania) and then sequenced using standard M13 primers with the GenomeLab DTCS kit on a CEQ 8000 DNA sequencer (Beckman Coulter, Fullerton, USA). Six clones of every gene cDNA were isolated and sequenced. The synthetic oligonucleotides were purchased from Evrogen (Moscow, Russia).

2.3. Sequence analysis

Intron–exon structure was obtained through cDNA/gDNA comparisons and identification of splice site consensus sequences. Interspersed repeats were searched by screening gDNA sequences against a library of repetitive elements RepeatMasker (http://www.repeatmasker.org/).

Using a combination of methods, the predicted FREP protein sequences were screened to confirm the presence of fibrinogen C-terminal domain-like superfamily domain (FreD) and at least one IgSF domain: (1) blast search against the NCBI protein database resulted in a significant hit to a FreD-containing sequence ($E < 10^{-5}$) and (2) protein domain search against SMART database [18], PFAM [19] and SUPERFAMILY hidden Markov models (SCOP domains: http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/hmm.html) resulted in a IgSF profile hit from more than one database. Putative signal sequences were identified by SMART and confirmed by hydropathy plots. Potential *N*-glycosylation sites were identified using the PROSITE database [20].

The phylogenetic relationships of FREP protein sequences were determined by using Neighbor-Joining method with *Gamma* parameter 1.0 in Mega4 [21]. Bootstrap support values after 1000 replicates were obtained for the NJ trees.

2.4. Analysis of nucleotide variability

In order to estimate frequency and distribution of point nucleotide variability or substitutions in both AcFREPs, genomic trace sequences corresponding to these genes were analyzed together with *Aplysia* actin (mRNA: U01352; gDNA: AASC02017377) and FMRFamide (mRNA: M14958; gDNA: AASC02048406) for control comparisons. The last gene is known to exist in the *Aplysia* genome as a single copy [22]. The trace sequences of all three genes were recovered from public NCBI TRACE archives and assembled in contigs using SEQMAN PRO 7.1 software. Before assembling, all traces were screened for vector contamination and quality end-trimming was performed. After assembling, for dual-end sequenced clones, if overlapping, one of the pair sequences was removed from the analysis and all trim coordinates were adjusted manually. No other edits were applied. Estimation of variation in sequence reads was performed using the "single-nucleotide polymorphism (SNP) discovery" algorithm of SEQMAN PRO 7.1. A reference gDNA sequence of corresponding gene was used for comparison with sample sequences and only the primary trace peak has been considered as a possible modified base. All allelic variants were excluded from the analysis. A central premise of this method is that sequencing errors are equiprobable events and minimum editing of raw trace sequences minimizes subjective factor in confirming or rejecting mismatch. It is important to note that single animal was used as a DNA source in *Aplysia* genomic project (NCBI). This feature allows distinguishing allelic differences in particular genes.