Developmental and Comparative Immunology 35 (2011) 635–643

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

The C1q domain containing proteins of the Mediterranean mussel *Mytilus galloprovincialis*: A widespread and diverse family of immune-related molecules

Marco Gerdol[a], Chiara Manfrin[a], Gianluca De Moro[a], Antonio Figueras[b], Beatriz Novoa[b], Paola Venier[c], Alberto Pallavicini[a,*]

[a] Department of Life Sciences, University of Trieste, Trieste, Italy
[b] Instituto de Investigaciones Marinas (CSIC), Eduardo Cabello 6, 36208 Vigo, Spain
[c] Department of Biology, CRIBI Biotechnology Center, University of Padova, Padova, Italy

ARTICLE INFO

Article history:
Received 10 November 2010
Received in revised form 26 January 2011
Accepted 27 January 2011
Available online 2 February 2011

Keywords:
M. galloprovincialis
MgC1q
Immune gene
Bacterial infection
Expression level

ABSTRACT

The key component of the classical complement pathway C1q is regarded as a major connecting link between innate and acquired immunity due to the highly adaptive binding properties of its trimeric globular domain gC1q. The gC1q domain also characterizes many non-complement proteins involved in a broad range of biological processes including apoptosis, inflammation, cell adhesion and cell differentiation. In molluscs and many other invertebrates lacking of adaptive immunity, C1q domain containing (C1qDC) proteins are abundant, they most probably emerged as lectins and subsequently evolved in a specialized class of pattern recognition molecules through the expanding interaction properties of gC1q.

Here we report the identification of 168 C1qDC transcript sequences of *Mytilus galloprovincialis*. The remarkable abundance of C1qDC transcripts in the Mediterranean mussel suggests an evolutionary strategy of gene duplication, functional diversification and selection of many specific C1qDC variants.

A comprehensive transcript sequence survey in Protostomia also revealed that the C1qDC family expansion observed in mussel could have occurred in some specific taxa independently from the events leading to the establishment of a large complement of C1qDC genes in the Chordates lineage.

1. Introduction

The gC1q is a globular domain which was first identified in the A, B and C chains of the C1q complement C1 complex subcomponent (Kishore and Reid, 2000). In addition to its fundamental role in the classical complement pathway, C1q provides a major link between innate and adaptive immunity, being involved in a wide range of immunological processes such as apoptotic cells clearance, bacteria and retrovirus recognition, cell adhesion and cell growth modulation (Kishore et al., 2004). Such an extreme versatility is granted by the ligand binding properties of the gC1q domain (Gaboriaud et al., 2003; Kishore and Reid, 1999).

The remarkable similarity of the gC1q and tumor necrosis factor (TNF) domains supports a common evolutionary origin for these two gene families (Shapiro and Scherer, 1998). Decisive amino acid changes and association to other functional domains can explain the wide variety of non-complement proteins: globally referred to C1qDC proteins they consist of an optional leading signal peptide, a central collagen-like region of variable length, acting as oligomerization domain and sometimes missing, and a C-terminal C1q domain (Ghai et al., 2007). Depending on the presence or the absence of the collagen-like region, C1qDC proteins are classified as C1q-like proteins or ghC1q proteins, respectively (Carland and Gerwick, 2010).

C1qDC proteins are probably essential in the innate immune system of early animals, as in the agnathan lamprey, having a still primitive adaptive immunity, C1q was shown to act as a lectin. Actually, lectin-like C1q proteins emerged before the immunoglobulins and expanded through the great flexibility and modulability of the gC1q domain in ligand binding (Fujita et al., 2004; Matsushita et al., 2004). Many C1qDC proteins can be regarded as specialized pattern recognition proteins (PRPs), able to bind pathogens directly through pathogen-associated molecular patterns (PAMPs) and to trigger phagocytosis (Bohlson et al., 2007; Medzhitov and Janeway, 2002).

Despite widespread in animal species, both retention or loss of C1q genes have apparently occurred in the evolution of Metazoa. Seven C1q gene models have been identified in the sea urchin *Strongylocentrotus purpuratus* (Hibino et al., 2006), only two in the Ascidian *Ciona intestinalis* (Azumi et al., 2003) and their number starts growing in ancestral Chordates: 50 C1q gene models in the

Abbreviations: C1qDC, C1q domain containing; PRPs, pattern recognition proteins; PAMPs, pathogen-associated molecular patterns; ghC1q, globular head C1q; sghC1q, secreted globular head C1q.

* Corresponding author. Tel.: +39 040812237; fax: +39 040810860.
E-mail address: pallavic@units.it (A. Pallavicini).

0145-305X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2011.01.018

Cephalochordate *Branchiostoma floridae* which is considered as the most primitive extant of the chordate lineage (Huang et al., 2008; Yu et al., 2008), 52 gene models in zebrafish (Mei and Gui, 2008) and 29 in humans (Tom Tang et al., 2005). On the contrary, C1qDC genes seem to be completely absent in Fungi and Plantae (Yuzaki, 2008).

Some C1qDC proteins with specific ligand recognition properties have been described and characterized also in molluscs. In particular, a sialic acid-binding lectin has been identified in the snail *Cepaea hortensis* (Gerlach et al., 2004) and an LPS-binding protein has been described in the scallop *Chlamys farreri* (Zhang et al., 2008). Other two C1qDC proteins, the major extrapallial fluid protein of *Mytilus edulis* (Hattan et al., 2001; Yin et al., 2005) and a protein highly expressed in the mantle tissue of *Pinctada fucata* (Liu et al., 2007) may be somehow involved also in the process of nacre biomineralization. The role of C1qDC proteins in specific pathogen recognition has been investigated in molluscs only recently: up-regulation of C1qDC proteins has been linked to infections with bacterial and metazoan parasites in molluscs such as *Ruditapes decussatus* (Prado-Alvarez et al., 2009), *Biomphalaria glabrata* (Adema et al., 2010), *Crassostrea gigas* (Taris et al., 2009) and *Mercenaria mercenaria* (Perrigault et al., 2009). AiC1qDC-1, a novel C1q domain containing protein recently characterized in the scallop *Argopecten irradians*, displays a fungi-agglutinating activity, and highlights, once again, the surprising ability of the gC1q domain to interact with many different PAMPs (Kong et al., 2010). In *Mytilus galloprovincialis*, the expression of MgC1q has been thoroughly examined in different tissues and larval stages (Gestal et al., 2010): MgC1q RNAs are abundant in hemocytes and increase rapidly and strongly in response to the injection of Gram+ and Gram− bacteria. Despite these facts point to an involvement of molluscan C1qDC proteins in pathogen recognition and innate immune response, to date the available data do not clarify the expansion and multifaceted functions of C1qDC proteins in this phylum and, more in general, in the Protostomia.

Recently, Carland and Gerwick (2010) reviewed the distribution of C1qDC proteins in animals, revealing the ancient origin of the gC1q domain and concluding that ghC1q genes became prevalent starting with Protostomia and radiated in the vertebrate animals. Here, we report and discuss for the first time the presence of a large family of C1qDC sequences almost exclusively coding for ghC1q proteins in the transcriptome of a non-Chordate organism, *M. galloprovincialis*. Despite the lack of genomic sequences, such an abundance and diversity of transcripts is suggestive of a similar over-representation of the C1qDC genes in the nuclear DNA. Mining publicly available transcriptomic and genomic data we also show that this astounding gene family expansion is restricted to Bivalvia and possibly to a few other unrelated Protostomia classes, and we raise the hypothesis that multiple events of C1qDC gene family expansion can have occurred in few taxonomic groups independently from the events leading to the acquisition of a large complement of C1qDC genes in the Chordates lineage.

2. Materials and methods

2.1. Sequence analysis

We used InterProScan (Zdobnov and Apweiler, 2001) to identify the InterPro signature IPR001073 for the C1q domain in the 7112 independent sequences of Mytibase, the annotated EST database of *M. galloprovincialis* (Venier et al., 2009). We selected consensus sequences having a significant score for at least one of the four PRINTS, PROFILE, SMART and PFAM signatures for complement C1q, and the related clustered ESTs were individually checked for possible sequencing errors. To provide a conservative estimate of the C1q gene models present in Mytibase, an ESTs collection derived from many mussels, we collapsed in a single consensus both highly similar clusters, possibly originated by noisy chromatograms or sequencing errors, and clusters coding for peptides with an identity percentage greater than 75%, assuming they could refer to the same gene. All the resulting clusters were translated in putative proteins using the Expasy Translate tool (http://www.expasy.ch/tools/dna.html) and only the full length sequences were retained for subsequent analysis. The combined tools for transmembrane topology and signal peptide prediction Phobius (Kall et al., 2004) and SPOCTOPUS (Viklund et al., 2008) were both used to avoid misclassification of these two classes of hydrophobic regions. Coiled coiled domains were predicted with COILS (Lupas et al., 1991) considering true only the cases predicted with a probability higher than 0.7 in at least two out of the three given window sizes. The coiled coil domain containing sequences were then scanned for the presence of leucine zipper motifs using 2ZIP (Bornberg-Bauer et al., 1998). InterProScan supported the identification of additional domains other than C1q in the same proteins.

The full length mRNAs described in this manuscript have been submitted to the EMBL database under the accession numbers from FR715581 to FR715677.

2.2. Mussel samples

Mussels of 6.5–7 cm shell length were collected from a farming site of the Venice Lagoon, Italy. To evaluate the tissue-specific expression of different C1qDC transcripts, total RNA was individually purified from hemolymph and from digestive gland, gill, gonads and posterior abductor muscle, previously homogenized in Tri reagent® (Sigma–Aldrich, St. Louis, MO).

Bacterial challenges were performed on adult mussels from Riá de Vigo, Spain, kept in tanks under controlled conditions (filtered seawater at 15 °C with aeration) and fed daily with *Isochrysis galbana*, *Tetraselmis suecica* and *Skeletonema costatum*. After an acclimatization time of 10 days, three groups of 60 mussels were challenged by injection into the adductor muscle with *Vibrio anguillarum* or *Micrococcus lysodeikticus* (100 μl of 10^7 live bacteria in filtered sea water). Controls were injected with 100 μl of filtered sea water. *M. lysodeikticus* was grown in LB medium at 37 °C and *V. anguillarum* in TSA supplemented with NaCl 1% at 20 °C. All individuals were maintained out of water for 20–30 min before and after the injection. At 3, 6 and 24 h post-injection, the hemolymph was collected and pooled from 20 mussels per sampling time and treatment.

Following extraction, the RNA quality was assessed by electrophoresis on denaturing agarose gel and its quantity was estimated using a spectrophotometer. Complementary DNA was prepared by retro-transcription with the iScript™ cDNA Synthesis Kit (Bio-Rad) from the pooled RNA samples representing five or 20 individuals.

2.3. Quantitative PCR expression analysis

The expression levels of eight selected C1q transcripts, namely MgC1q1, MgC1q2, MgC1q3, MgC1q4, MgC1q5, MgC1q6, MgC1q7 and MgC1q8, were assessed in samples representing the hemocytes, digestive gland, gills, gonads and posterior abductor muscle of five adult mussels. Primer pairs were designed (Table 1) and used to obtain specific PCR amplicons, finally checking the reaction specificity by Sanger sequencing (ABI3130 Genetic Analyzer).

The expression of transcripts classified as hemocyte-specific, according to their relative abundance in the selected tissues, was also analyzed in the hemolymph sampled at 3, 6 and 24 h post-challenge from mussels injected with Gram+ (*M. lysodeikticus*) or Gram− (*V. anguillarum*) bacterial cells.