Gene 431 (2009) 47–54

Contents lists available at ScienceDirect

Gene

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

Isolation and characterization of two cytoplasmic *hsp90s* from *Mytilus galloprovincialis* (Mollusca: Bivalvia) that contain a complex promoter with a p53 binding site

Chrysoula N. Pantzartzi [a], Antonis Kourtidis [b], Elena Drosopoulou [a], Minas Yiangou [a], Zacharias G. Scouras [a,*]

[a] Department of Genetics, Development and Molecular Biology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, Thessaloniki, GR-54124 Greece
[b] GenNYSis Center for Excellence in Cancer Genomics, State University of New York at Albany, Rensselaer, NY 12144, USA

ARTICLE INFO

Article history:
Received 4 August 2008
Received in revised form 29 October 2008
Accepted 31 October 2008
Available online 13 November 2008

Keywords:
HSP90 family
HSE
Genomic organization
ESTs
Mytilus

ABSTRACT

The commercially important marine bivalve *Mytilus galloprovincialis* (Mediterranean mussel) is considered a valuable bioindicator, due to its exposure to various pollutants and extreme environmental conditions. Environmental responsive genes, such as the *hsp90*s, protect the structure and function of cells and accomplish a significant task in cellular homeostasis. To study the *hsp90*s in *M. galloprovincialis* a genomic library was screened and two *hsp90*s were isolated. Sequence analysis revealed that the two genes exhibit great similarities in both the 5′ non-coding and the coding region but differ in the 3′ non-coding region, as well as in three introns, due to the presence of repeated sequences. Few synonymous substitutions in the coding region of the genes result to an identical predicted polypeptide, which belongs to the cytoplasmic HSP90 subfamily. The 5′ non-coding region contains a non-translated exon and multiple binding sites for various transcription factors. The presence of a p53 binding site in the promoter of the isolated genes raises questions about the possible implication of *hsp90*s in the molluscan leukemia.

1. Introduction

The Mediterranean mussel (*Mytilus galloprovincialis*; Lamark, 1819) belongs to Mollusca, the second largest and the most diverse metazoan phylum (Koehn, 1991). It is an edible marine bivalve with significant commercial impact. In addition, its abundance in coastal areas, exposed to heavy metals and other chemical pollutants, renders this species an important bioindicator (Viarengo et al., 1997; Dondero et al., 2006). The unique capability of *M. galloprovincialis* to adapt to unstable and extreme environmental conditions draws attention to its environmental responsive genes, such as the heat shock genes.

Heat shock proteins (Hsps) protect cells from the potentially fatal consequences of adverse environmental, physical, or chemical stresses (Feder and Hofmann, 1999). They are highly conserved proteins grouped into families according to their size. The HSP90 family consists of mitochondrial, chloroplastic, endoplasmic and cytoplasmic homologues, according to their sub-cellular localization (Gupta, 1995; Chen et al., 2006). In vertebrates, there are two cytoplasmic isoforms, namely the α (heat-inducible) and the β (constitutively expressed or cognate) (Krone and Sass, 1994; Gupta, 1995; Chen et al., 2006). In invertebrates, a unique cytoplasmic Hsp90 is encoded either by a single or by two gene copies (Blackman and Meselson 1986; Benedict et al., 1996; Birnby et al., 2000; Chen et al., 2006). An inducible and a cognate isoform exist in *Saccharomyces cerevisiae* (Gupta, 1995; Erkine et al., 1995, 1999), whereas multiple identical copies code for the same Hsp90 in protists (Mottram et al., 1989). Hsp90s have been found in all organisms studied so far, except Archaea (Laksanalamai et al., 2004).

Binding of Heat Shock Factors (HSFs) to promoter sequences called Heat Shock Elements (HSEs) activates *hsp90* transcription (Morimoto 1998). Expression of *hsp90* genes can be regulated by one or several members of the HSF family in response to environmental stressors (e.g. heat or cold shock, salinity, heavy metals etc.), or during various stages of development (see Pirkkala et al., 2001 as review).

Hsp90s participate in multicomponent chaperone machineries with Hsp70 and other chaperones, contributing to various cellular processes such as protein folding, degradation, protection from protein unfolding and aggregation during stress, cytoskeleton organization, protein translocation and signal transduction (Richter and Buchner, 2001; Young et al., 2001; Pratt and Toft, 2003; Wegele et al., 2004; Brown et al., 2007). Proteins involved in cell cycle control and signal transduction constitute a large proportion of Hsp90's substrates. Depending on the substrate, Hsp90 may either associate with substrates and maintain them inactive in the absence of a stimulatory signal or assist in final folding, complex formation and/or maintenance, after signalling (Picard, 2002).

Abbreviations: aa, amino acid(s); bp, base pair(s); Da, Dalton; ESTs, Expressed Sequence Tags; GAF, GAGA Factor; HSE, Heat Shock Element; HSF, Heat Shock Factor; Hsp, Heat shock protein; *hsp*, heat shock gene; PCR, polymerase chain reaction; TIS, Transcription Initiation Site.

* Corresponding author. Tel.: +302310998324; fax: +302310998329.
E-mail address: scouras@bio.auth.gr (Z.G. Scouras).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.10.028

The role of Hsp90s in many types of cancers, e.g. breast cancer, pancreatic carcinoma and leukemia, intensified the research on this multifunctional protein family and the use of its inhibitors as potential anti-cancer drugs (see Csermely et al., 1998; Whitesell and Lindquist, 2005 for reviews and references therein). Hsp90 is essential for the stabilization and maturation of nuclear hormone receptors, transcription factors, and protein kinases that are commonly misregulated during tumorigenesis. One of the Hsp90 client proteins is the transcription factor p53, a tumor suppressor protein that has several functions, including cell cycle regulation, DNA binding, transcriptional activation and repression (Whitesell and Lindquist, 2005). Homologues of the p53 gene family have been isolated and characterized in several molluscan species, including three *Mytilus* species (*M. edulis*, *M. trossulus* and *M. galloprovincialis*). In addition, a number of studies have demonstrated that p53 is implicated in the onset of molluscan leukemia (Barker et al., 1997; Kelley et al., 2001; Muttray et al., 2005; Walker et al., 2006).

In order to enrich the limited dataset on *hsp* genes in Mollusca, previous reports from our laboratory have focused on the study of *hsp70* of *M. galloprovincialis*, where multiple cognate and inducible cytoplasmic copies have been isolated and characterized (Kourtidis and Scouras, 2005; Kourtidis et al., 2006a,b). We report here on the isolation, characterization and complete genomic organization of two *hsp90s* from *M. galloprovincialis* that encode for the same putative protein and contain a rather complex 5′ non-coding region in which a p53 binding site is included. In addition, using these genes as queries, we identified, grouped and mapped non-characterized ESTs from *M. galloprovincialis*, *M. californianus* and *M. edulis* that bear highly similar *hsp90* sequences.

2. Materials and methods

2.1. Screening of a Mytilus galloprovincialis genomic library — clone mapping

A *M. galloprovincialis* genomic library (Kourtidis et al., 2006a) was screened using a homologous probe obtained via polymerase chain reaction (PCR). Degenerate primers were designed using CODEHOP strategy (Rose et al., 1998) and aminoacid alignments of various HSP90 family members from the BLOCKS database (Henikoff et al., 2000). Genomic DNA from *M. galloprovincialis*, prepared according to Kourtidis et al. (2006a), was used as template. Sequence analysis showed that the PCR product used as probe was 307 bp in length. Probe was labelled with ^{32}P-dCTP (ICN Biochemicals, Irvine, CA) as a precursor, using the random priming method (Feinberg and Vogelstein, 1983). Plaque hybridizations were performed at 61 °C in aqueous solution, several positive clones were selected and lambda DNA was isolated according to Sambrook et al. (1989). The clones were mapped with single and double digestions using three restriction endonucleases and Southern hybridizations at the same conditions as above. Autoradiography was performed using Kodak X-OMAT 100 film.

2.2. Subcloning and sequence analysis

Selected restriction fragments from the isolated genomic clones were subcloned using the pGEM-3Zf (-) plasmid vector (Promega, Madison, WI) according to the manufacturer manual, and sequenced in an automated sequencer (Lark Technologies, Essex, UK).

Similarity searches were performed using the BLAST algorithm (Altschul et al., 1990) from the NCBI site. The coding regions of the *hsp90s* in *M. galloprovincialis* were predicted using the Eukaryotic GeneMark.hmm (Lomsadze et al., 2005) and the ORF Finder tools (NCBI site). The first non-coding exon was predicted through the FirstEF program (Davuluri et al., 2001). Exon–intron boundaries were cross-checked with SpliceView (Rogozin and Milanesi, 1997). Regulatory elements, in both the 5′ and 3′ non-coding regions (the sequenced regions upstream the ATG and downstream the stop codon, respectively) were identified with the Alibaba2, MATCH, PolyAScan tools (http://www.gene-regulation.com) and TESS (Transcription Element Search System, http://www.cbil.upenn.edu/cgi-bin/tess/tess). The protein's molecular weight was calculated using ProtParam tool (Gasteiger et al., 2005) from the Expasy server. The conserved HSP90 protein family signatures were identified using the program PSCAN from the EMBOSS suite of programs (Rice et al., 2000) and data from PRINTS database. Pairwise alignments, similarity and identity scores were derived through the EMBOSS-ALIGN program. Multiple alignments were performed using ClustalW (Thompson et al., 1994). Synonymous (p_S) and non-synonymous (p_N) substitutions were calculated using MEGA version 3.1 (Kumar et al., 2004) and applying the modified Nei-Gobojori method, considered to be more accurate when the ratio of transitions to transversions (R) is higher than 0.5 (in our case $R = 1.333$) (Zhang et al., 1998).

3. Results

3.1. Identification of two hsp90 transcriptional units in Mytilus galloprovincialis

Screening a *M. galloprovincialis* genomic library resulted in the isolation of several clones bearing *hsp90* sequences. Five of these clones were subjected to mapping and were divided into two groups, according to their restriction and hybridization patterns. These two groups, namely λMgHsp90.1 (Fig. 1A) and λMgHsp90.2 (Fig. 1B), cover genomic regions of ~17.5 and 18.5 kb in length, respectively. Selected fragments of one representative clone from each group were subcloned and sequenced. Total sequence of ~8.5 kb concerning group λMgHsp90.1 (GenBank accession no. AM236589) and of ~12.5 kb concerning group λMgHsp90.2 (GenBank accession no. AJ586906) was obtained.

Sequence analyses showed that each representative clone contains a complete *hsp90* transcriptional unit that we designated *Mghsp90-1* and *Mghsp90-2*, respectively. Both units exhibit extensive similarities in structure and sequence (Fig. 1, Table 1 and Supplemental Fig. 1). Each gene consists of 9 exons, the 1st of which is non-coding with a size of 187 bp, according to the FirstEF program prediction (Figs. 1 and 2, Table 1 and Supplemental Fig. 1). The coding region of each gene is 2169 bp long, exhibiting 99.7% sequence identity. A high degree of size and sequence similarity (98–99%) is also evident in the intronic regions of the two units with the exception of introns 5, 6 and 7 (Table 1). Introns 5 and 7 of *Mghsp90-2* are significantly longer than the respective ones of *Mghsp90-1*, due to the presence of repeated sequences, while intron 6 of *Mghsp90-1* contains an addition of approximately 70 bp compared to the respective intron of *Mghsp90-2* (Supplemental Fig. 1). Exclusion of the repeated sequences reveals great similarity (>96.5%) of the above introns of the two genes (Table 1). Repeated sequences were also observed in introns of the *hsp70* gene family in *M. galloprovincialis* (Kourtidis et al., 2006b).

A typical polyA signal (AATAAA) has been identified 250 bp downstream the stop codon in *Mghsp90-2*, while a non-typical one (AACAAA) was found 161 bp downstream the stop codon in *Mghsp90-1* (Supplemental Fig. 1). High sequence similarity between the two units is extended in the first 120 bp downstream the stop codon (95%) but it is significantly reduced (<50%) in the following bases (Supplemental Fig. 1).

3.2. The 5′ regulatory region of the Mghsp90 genes

The 5′ upstream region of the two *M. galloprovincialis hsp90s* is almost identical containing a large number of regulatory elements. In both genes, putative CAAT and TATA boxes were detected, at a distance of 271 and 49 bp upstream the transcription initiation site (TIS), respectively. In addition, CAAT and TATA boxes as well as a putative TIS were also detected within the 1st intron (Fig. 2 and Supplemental Fig. 1).