IDENTIFICATION OF HOMOLOGUES OF THE EMH-3 HOMEBOX-CONTAINING GENE IN DEMOSPONGES

EVELYN RICHELLE-MAURER, GONTRAN SONET & GISELE VAN DE VYVER

Laboratoire de Physiologie cellulaire, IBMM CP 300, Université Libre de Bruxelles, 50 av. F. Roosevelt, 1050 Brussels, Belgium
E-mail: emaurer@ulb.ac.be

ABSTRACT

In the present study, we have identified and characterized homologues of the homeobox-containing gene EmH-3 from Ephydatia muelleri in three additional freshwater sponges: Eunapius fragilis, Spongilla lacustris and Trochospongilla horrida. They were designated EmH-3Efr, EmH-3Sl and EmH-3Th. They exhibited a great conservation of genomic and amino acid sequences, exon/intron structure and expression patterns with the EmH-3 gene, which strongly suggested that they have the same function in sponge development. Phylogenetic analyses have shown a grouping of E. muelleri and E. fluviatilis, E. fragilis and S. lacustris, T. horrida being outside both groups. They also showed that EmH-3 and its homologues were most closely related to the Tlx/Hox11 family of higher metazoans. They are probably representatives of this family as they display the family-specific signature and, as are the members of this family, are involved in cell multiplication and differentiation.

KEY WORDS

Demosponges, homeobox-containing genes, sequencing, expression, phylogeny.

INTRODUCTION

In recent years, efforts have been made to understand the molecular mechanisms underlying animal development in relation to different metazoans’ body plans and developmental features. It has been shown that a family of homologue genes, the homeobox-containing genes play a key role in morphogenesis, particularly in the control of spatio-temporal organization, cell fate decisions and differentiation (MEYER, 1996; FINNERTY, 1998; GELLON & MCGINNIS, 1998; MARTINEZ et al., 1999; GALLIOT & MILLER, 2000; MILLER & BALL, 2000; MORRIS, 2000). In this context the study of the structure and function of these genes in basal metazoans, e.g. Porifera has appeared to be fundamental to understanding the evolution of these genes throughout the animal kingdom (MARTINDALE & KOURAKIS, 1999; GAUCHAT et al., 2000; MANUEL & LE PARCO, 2000; FINNERTY, 2001). Their study is particularly crucial as sponges and higher metazoan phyla share different metabolic pathways, structural elements and molecules suggesting that the basis of pluricellularity and the general regulation mechanisms have been established before the divergence of sponges and eumetazoans (EXPOSITO & GARRONE, 1990; DEGNAN et al., 1993; MÜLLER & MÜLLER, 1999; BORCHIELLINI et al., 2000; MÜLLER, 2001). In this context, we have investigated the expression of different homeobox-containing genes in freshwater sponges and determined the role the EmH-3 gene plays in their morphogenesis. This gene, isolated from the genomic
library of the freshwater sponge *Ephydatia muelleri* (RICHELLE-MAURER et al., 1998) presents a high similarity to *EfH-1/prox2* (98 %) isolated from the freshwater sponge *Ephydatia fluviatilis* (COUTINHO et al., 1994; SEIMIYA et al., 1994) and *Spox TA1* (96 %) isolated from the marine sponge *Tethya aurantium* (DEGNAN et al., 1995). It is probably a representative of the bilaterian non-Hox Tlx/Hox11 family (COUTINHO et al., 1998; RICHELLE-MAURER et al., 1998; GAUCHAT et al., 2000). Expression analyses have shown that *EmH-3* expression is time and cell-specific. The levels of *EmH-3* expression increase from almost undetectable levels in resting gemmules to high levels at the moment of hatching and throughout the sponge's life (RICHELLE-MAURER et al., 1998; RICHELLE-MAURER & VAN DE VYVER, 1999a). *EmH-3* is strongly expressed in the undifferentiated and pluripotent archaeocytes whereas in choanocytes or pinacocytes its expression is very weak, similar to that found in the resting gemmules (RICHELLE-MAURER & VAN DE VYVER, 1999b). In contrast, other genes such as the *E. fluviatilis* homeobox-containing gene *prox1* and the *E. muelleri* actin gene are expressed at almost the same level at all stages of development and in all cell fractions (RICHELLE-MAURER & VAN DE VYVER, 1999b). Regarding the function of *EmH-3*, the use of inhibitors of sponge development, such as retinoic acid, have shown that *EmH-3* is necessary for the differentiation of archaeocytes into choanocytes and hence for the completion of a functional sponge (NIKKO et al., 2001). It is worth noting that the *EmH-3* promoter is functionally active in mammalian cells indicating that the regulatory programmes in Porifera are similar to those found in higher metazoans (COUTINHO et al., 1998).

In the present work homologues of the *EmH-3* gene have been identified and characterized in three additional freshwater sponge species, *Spongilla lacustris*, *Eunapius fragilis*, and *Trochospongilla borrida*, by PCR and RT-PCR with specific primers and sequencing.

**MATERIALS AND METHODS**

**Sponge culture**

Sponges were grown in laboratory conditions at 20° C, in Petri dishes containing mineral medium, from pure stocks of gemmules as described before (RICHELLE-MAURER et al., 1998). Five species of freshwater sponges were used: *Ephydatia muelleri*, *Ephydatia fluviatilis*, *Eunapius fragilis*, *Spongilla lacustris*, and *Trochospongilla borrida*.

**DNA extraction and PCR conditions**

Genomic DNA was extracted from gemmules using DNAzol reagent (Invitrogen) following the manufacturer's instructions. The gemmules were ground in a Potter homogeniser in the presence of DNAzol, the empty shells were discarded and the homogenates centrifuged at 10000 x g for 10 min prior to DNA precipitation with ethanol. The quantity and quality of DNA were evaluated respectively by spectrometry and on a 0.7 % agarose gel.

PCR amplifications were performed in 50 µl using 45 µl of PCR SuperMix (Invitrogen), 400 nM primers and 10 ng DNA template. The typical reaction profile was 35 cycles (2 min 94° C, 1 min 55° C, 1 min 30 s 72° C) preceded by 1 min at 94° C and followed by 10 min at 72° C (Techne Unit Genius). However, in the case of *S. lacustris* and *T. borrida*, the annealing temperature had to be lowered from 55° C to 45° C to get amplification. The primers used were specific of the *EmH-3* gene, lying at the beginning of the first exon and in the homeobox (Fig. 2). About 10 % of the PCR products were electrophoresed on a 1.5 %
agarose gel. *E. muelleri* and *E. fluviatilis* were used as controls: their amplified PCR products were sequenced and compared to the published *EmH-3* and *prox2* sequences (RICHELLE-MAURER et al., 1998, SEIMIYA et al., 1994).

The sequences reported in this study have been deposited in the GenBank database under the following accession numbers AY300029 (*EmH-3Efr*), AY300050 (*EmH-3Sl*) and AY300031 (*EmH-3Th*).

**RNA extraction and RT-PCR conditions**

The expression of *EmH-3* homologue genes was analysed by RT-PCR using the same set of primers as those used for the PCR experiments. Total RNA extraction and PCR experiments were carried out as described elsewhere (RICHELLE-MAURER et al., 1998, RICHELLE-MAURER & VAN DE VYVER, 1999a,b). Briefly total RNA was extracted at different stages of development from gemmules to the completion of functional sponges, using TRIzol reagent (Invitrogen) and was amplified using the Promega Access RT-PCR one tube System. The cycling conditions were those described in the manufacturer’s protocol using 10 ng template and an annealing temperature of 55° C. Expressions patterns were compared to those obtained with *EmH-3* in *E. muelleri* and with the *EmA1* actin housekeeping gene in the different sponge species. Ten percent of the RT-PCR products were analysed on a 1.5 % agarose gel. Quantitative evaluation of the bands was made by using program ImageMaster 1D Elite Version 3.0 for gel analysis (Amersham Pharmacia Biotech). The results are presented in the form of histograms.

**Sequencing and Phylogenetic analysis**

The (RT)-PCR products of the expected size were either directly purified on QIAquick columns QIAGEN or extracted from the agarose gel and purified on QIAquick extraction kit columns QIAGEN. Sequencing was performed according to Sanger’s method (SANGER et al., 1977). PCR fragments were sequenced on both strands using an ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kit, version 2.0 (Applied Biosystems 2000). Sequence chromatograph data were analysed and edited with the Sequencing analysis Program (ABI). The sequences of each strand were aligned with the Seq Pup program (GILBERT, 1996).

Multiple sequence alignments of nucleotide (nt) and amino acid (aa) sequences were constructed with the ClustalX program (THOMPSON et al., 1997). They were optimized by testing various gap weights by the Soap Program (LOYTNOJA & MILINKOVITCH, 2001). The Paup software (SWOFFORD, 1998) was used for distance computation, Neighbour-joining (NJ) and Maximum Parsimony (MP) tree-buildings, and bootstrapping. The reliability of the nodes was assessed by using 500 bootstrap replicates (PAGE & HOLMES, 1998). Searches for related sequences in higher metazoans were performed in the GenBank databases by BLAST.

**RESULTS**

**Gene structure and sequencing**

Amplified PCR products within the expected range size were recovered from *Spongilla lacustris*, *Eunapius fragilis* and *Trochospongilla borrida* with the *EmH-3* specific primers. Their size varied according to the species, from 736 bp (*E. fragilis*) to 905 bp (*T. borrida*) (Tab. I). The amplified products from *Ephydatia muelleri* and *Ephydatia fluviatilis* had the estimated size as deduced from the published sequences of *EmH-3* (RICHELLE-MAURER et al., 1998) and *prox2* genes (SEIMIYA et al., 1994) i.e. 753 bp and 738 bp respectively.
On the other hand all transcripts had the same size (444 bp) as \textit{E}m\textit{H}-3 transcripts except \textit{E. fragilis} (441 bp) (Tab. I). They were 54 bp (55 bp) longer than \textit{E. fluviatilis} transcripts.

**Tab. I.** Size (bp) of the amplified PCR and RT-PCR products

<table>
<thead>
<tr>
<th></th>
<th>\textit{E. fragilis}</th>
<th>\textit{S. lacustris}</th>
<th>\textit{T. horrida}</th>
<th>\textit{E. muelleri}</th>
<th>\textit{E. fluviatilis}</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>736</td>
<td>804</td>
<td>905</td>
<td>753</td>
<td>738</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>441</td>
<td>444</td>
<td>444</td>
<td>444</td>
<td>390</td>
</tr>
</tbody>
</table>

The alignment of the genomic and cDNA sequences allowed the exons-introns boundaries to be distinguished. The genomic sequences presented the same structural organization in the three species, \textit{i.e.} three exons separated by two introns, the homeobox lying in the third exon (Fig. 1). The exons had similar lengths in the three species, (Figs 1, 2) but the length of the introns varied, particularly the second one, from 232 bp in \textit{E. fragilis} to 414 bp in \textit{T. horrida}, accounting for the difference of length of the PCR products. Splicing sites were identical to those found in \textit{E. muelleri} but differed from those found in \textit{E. fluviatilis}, the first exon being 54 bp shorter in the latter species (Figs 1, 2). It has to be noted that the first codon of the third exon is formed by the last nucleotide of exon 2 and the two first nucleotides of exon 3.

The genomic sequences obtained from \textit{E. fragilis}, \textit{S. lacustris}, and \textit{T. horrida} displayed a high degree of identity with the \textit{E. muelleri} sequence (Fig. 2). The identity was not restricted to the homeobox but extended to the three exons (Tab. II). Moreover, two thirds of the nucleotide substitutions were located at the level of the degenerated base of the codon and consequently had no influence on the deduced amino acid sequence. Identity was much lower in the untranslated regions (Tab. II).
It is worth noting that the genomic sequences obtained from *E. muelleri* and *E. fluviatilis* were identical to the published sequences from these species, respectively, *EmH-3* (Richelle-Maurer et al., 1998) and *prox2* (Seimiya et al., 1994) except that in the latter species two substitutions were observed at positions 194 and 487 in the introns (Fig. 2, highlighted).

**Tab. II.** Percentage identity between the nucleotide (nt) and deduced amino acid (aa) sequences from *Ephydatia muelleri* (*EmH-3*) and from *Eunapius fragilis*, *Spongilla lacustris* and *Trochospongilla borrada*.

<table>
<thead>
<tr>
<th></th>
<th><strong>E. muelleri (EmH-3)</strong></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nt sequences</td>
<td>aa sequences</td>
<td>Introns</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homeobox (exons)</td>
<td>Homeodomain (exons)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. fragilis</em></td>
<td>87.0 (84.9)</td>
<td>100.0 (87.6)</td>
<td>57.7</td>
<td></td>
</tr>
<tr>
<td><em>S. lacustris</em></td>
<td>87.0 (86.5)</td>
<td>98.2 (84.7)</td>
<td>52.7</td>
<td></td>
</tr>
<tr>
<td><em>T. borrada</em></td>
<td>91.4 (87.0)</td>
<td>100.0 (87.8)</td>
<td>56.0</td>
<td></td>
</tr>
</tbody>
</table>

The deduced amino acid sequences, presented in Fig. 3, showed a very high identity with the *EmH-3* sequence, up to 100% in the homeodomain for *E. fragilis* and *T. borrada* (Tab. II). Identity remained very high outside the homeodomain (87.8%), particularly at the level of the third exon (Fig. 3). Most of the differences were located in the last 18 amino acids of the first exon corresponding to the region, which was shorter in *E. fluviatilis*. Identity with *EmH-3* in this region dropped to 52.9% (*S. lacustris*). All the homeodomains possessed 6 of the 7 conserved residues (*R*<sub>1</sub>, *R*<sub>10</sub>, *K*<sub>24</sub>, *S*<sub>28</sub>, *K*<sub>39</sub>, and *T*<sub>47</sub>) of the bilaterian Tlx/Hox11 family-specific signature (Gauchat et al., 2000).

**Expression analysis**

Expression patterns were very similar in *E. fragilis*, *S. lacustris*, and *T. borrada* (only *E. fragilis* is shown in Fig. 4) and resembled those found in *E. muelleri*. When the *EmH-3* specific primers were used, transcripts were hardly discernible in the resting gemmules (Fig. 4a) but their level increased drastically at hatching and this high level was maintained in the functional sponges. On the contrary, with the *EmA1* actin primers, the levels of transcripts were almost the same in the gemmules and during sponge development (Fig. 4b).

**Phylogenetic analyses**

The Neighbour-joining (NJ) and Maximum Parsimony (MP) trees based on the genomic sequences and rooted with *Spox TA1* from the marine sponge *T. aurantium* presented the same topology: a grouping of *E. muelleri* and *E. fluviatilis* on one hand, of *S. lacustris* and *E. fragilis* on the other hand, with *T. borrada* being outside both groups (Fig. 5). Such grouping was supported by high bootstrap values. However, the position of *T. borrada* to the other species varied according to the method used. In the NJ tree, *T. borrada* was related to the *Sl-Efr* cluster (59%) while in the MP tree this species was found to be related the *Em-Efl* cluster but this was weakly supported by a much lower bootstrap value (46%, not shown).
Fig. 2. Sequences of PCR amplified products from E. muelleri (Em), E. fragilis (Ef), S. lacustris (Sl), T. horrida (Th) and E. fluviatilis (Ef). Sequences are aligned to E. muelleri (Em) allowing deletions/insertions (using the ClustalX software). The exons are in bold and the introns in italic; the homeobox region is boxed; nucleotide identities are indicated by hyphens; gaps by =; non-sequenced regions by dots, the primers used for (RT)-PCR are underlined; differences from the published prox2 sequence (SEIMIYA et al., 1994) are highlighted.
Fig. 3. Deduced amino acid sequences from *Ephydatia muelleri* (Em), *Eunapius fragilis* (Efr), *Spongilla lacustris* (Sl), *Trochospongilla horrida* (Th) and *Ephydatia fluviatilis* (Efl). Sequences are aligned to *E. muelleri* allowing deletions/insertions (using the ClustalX software). Identities are indicated by hyphens, gaps by =; the homodomain is boxed; the residues of the Tlx/Hox11 family-specific signature are highlighted (inverted colours); the boundaries between the tree exons are shown by a space.

<table>
<thead>
<tr>
<th>Exon 1</th>
<th>Exon 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Em</td>
<td>ISDTEITRQGR</td>
</tr>
<tr>
<td>Efr</td>
<td>------------</td>
</tr>
<tr>
<td>Sl</td>
<td>------------</td>
</tr>
<tr>
<td>Th</td>
<td>------------</td>
</tr>
<tr>
<td>Efl</td>
<td>------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exon 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Em</td>
</tr>
<tr>
<td>Efr</td>
</tr>
<tr>
<td>Sl</td>
</tr>
<tr>
<td>Th</td>
</tr>
<tr>
<td>Efl</td>
</tr>
</tbody>
</table>
Fig. 4. Expression patterns in the course of development. Levels of the transcripts (pixel intensities in arbitrary units) in the three species (only Eunapius fragilis is shown as example) with the specific primers for EmH-3 (a) and EmA1 (b) at different stages of development: resting gemmules (0d); 2d incubation, hatching (H) and in functional sponges (H+4d). Incubation times are expressed in days (d) at 20°C in mineral medium.

Fig. 5. Rooted Neighbour-joining tree computed from the genomic sequences from Ephydatia muelleri (Em), Ephydatia fluviatilis (Ef), Eunapius fragilis (Efr), Spongilla lacustris (Sl) and Trochospongilla borrida (Th), with Spox TA1 from the marine sponge Tethya aurantium (Ta) as an outgroup. The bootstrap percentages (500 replicates) are shown at the nodes.
Fig. 6. Phylogenetic relationships between sponge homeodomains and the most closely related bilaterian homeodomains inferred by Neighbor-joining. The tree was constructed from the alignment of the full-length homeodomains and was rooted with the divergent *Mata* yeast homeodomain. The family to which the genes belong and was named with the divergent *Mata* yeast homeodomain followed by the organisms from which the sequences were identified and the accession numbers in the protein database. Numbers at the nodes indicate percentages of 500 bootstrap replicates; values under 50% are omitted, except for significantly important nodes.
The evolutionary relationships between the homeodomains of freshwater sponges and bilaterian animals were assessed by homology searches by Blast in the GenBank database. They revealed that the freshwater sponge homeodomains displayed similarities with those of members of the non-Hox families, Tlx, Hox11, Gbx, Barx, and Lbx of higher metazoans. In particular, up to 77% identity was found with members of the Tlx/Hox11 family. The NJ tree constructed from these homeodomain sequences showed that the sponge homeodomains were most closely related with members of the Tlx/Hox11 and Gbx families (Fig. 6) although this relationship was not well supported, the bootstrap values being less than 50%.

DISCUSSION AND CONCLUSIONS

The genes identified and characterized in Eunapius fragilis, Spongilla lacustris and Trochospongilla borrida by PCR with the EmH-3 specific primers can be considered as true homologues of the EmH-3 homeobox-containing gene from Ephydatia muelleri. Indeed, they satisfy the criterions for homology i.e. a high identity of sequence, the same gene structure and expressions patterns (Schierwater & Kuhn, 1998). They were designated EmH-3Efr, EmH-3Sl and EmH-3Th. Sequence identity is not restricted to the conserved homeobox but extends outside the homeobox to the third exon and to a lesser extent to the other translated regions contrary to what is observed in other organisms where little or no similarity is found outside the homeobox (Schierwater & Kuhn, 1998). Identity is much lower in the untranslated regions, which showed also a greater variability in length. The three genes possess the same fundamental 3 exons-2 introns structure as the EmH-3 gene. The position of the splicing sites confirmed those assumed in EmH-3 (Richelle-Maurer et al., 1998). Interestingly, of the five freshwater sponge species studied, only Ephydatia fluviatilis displayed an alternative splicing site, which resulted in a difference of length of the first exon.

As to their expression, EmH-3Efr, EmH-3Sl and EmH-3Th are differentially expressed during sponge development, from almost undetectable levels in gemmules to high levels at hatching and in functional sponges, which strongly suggests a role in cell multiplication and differentiation during sponge development as demonstrated for EmH-3 (Nikko et al., 2001).

Phylogenetic analyses showed that sponge homeodomains were found to be most closely related to the non-Hox Tlx/Hox11 family in agreement with previous research (Coutinho et al., 1998; Richelle-Maurer et al., 1998; Gauchat et al., 2000; Manuel, 2001; Coutinho et al., 2003). Although this relationship is only weakly supported by low bootstrap values, EmH-3, EmH-3Efr, EmH-3Sl and EmH-3Th are probably representatives of the Tlx/Hox11 family. In fact, they display the family-specific signature and as determined for EmH-3 they are involved in cell
multiplication and differentiation like the members of the Tlx/Hox11 family (Kennedy et al., 1991; Liu et al., 1991)

It is interesting to note that homeodomains from different taxonomic groups of animals are found in this family. This corresponds to the evolutionary scenario of homeodomains suggesting an ancient origin of the non-Hox families, the Hox/paraHox families being considered as more recent (Kappen et al., 1993; Valentine et al., 1996; Brooke et al., 1998; Gauchat et al., 2000).

Work is in progress to identify EmH-3 homologues in marine sponges and to determine if, in these sponges, the same conservation of sequence and structure is observed as in freshwater sponges. As mentioned before the presence of an EmH-3-like gene in a primitive marine sponge *T. aurantium* (Degnan et al., 1995) suggests that the *EmH-3* gene may be widespread among Porifera and may constitute one of the metazoan stem homeobox-containing genes.

ACKNOWLEDGEMENTS

This work was funded by a research grant of the FNRS (Fonds National de la Recherche Scientifique de Belgique, convention 1.5.058.02).

REFERENCES


