The Characterization of SmedHSP90 Gene Using Methods of Bioinformatics

Xingzi Xi[1], Keshi Ma[2]

[1]. Department of Education Sciences, Xinxiang University, Xinxiang 453003, China,
[2]. College of Life Sciences, Zhoukou Normal University, Zhoukou 466000, China
xingzixi2003@sina.com

Abstract: Bioinformatics is a very powerful tool in the field of genomics study. Firstly, we obtained the heat shock protein 90 gene structure of planarian Schmitdea mediterranea (designated SmedHSP90) by using Smedgenome database. The ORF of SmedHSP90 is 2148 bp encoding a polypeptide of 715 amino acids with all five conserved motifs of HSP90 protein family signature. In addition, the gene structure of SmedHSP90 contains a 48 bp intron, and the 5'- and 3'- splicing site follows the typical "GT-AG" rule. Secondly, we constructed the HSP90 phylogenetic tree using the software Clustal 1.83 and Mega 3.1, and discussed the evolutionary position of SmedHSP90. Thirdly, three-dimensional domain structure of SmedHSP90 was predicted by SWISS-MODEL Server. Bioinformatic analysis of SmedHSP90 gene provides basic data for the study of stress response in planarians.
[Xingzi Xi, Keshi Ma. **The Characterization of SmedHSP90 Gene Using Methods of Bioinformatics.** *Life Sci J* 2013;10(1):2836-2839] (ISSN:1097-8135). http://www.lifesciencesite.com. 341

Keywords: bioinformatics, planarian, HSP90

1. Introduction

The Heat shock proteins (HSPs) are highly conserved molecular chaperone belonging to multigenic families. According to their molecular weight, they can be classfied into several major categories, e.g., HSP100, HSP90, HSP70, HSP60 and low molecular weight (Jackson, 2013). Particularly, HSP90 proteins are ubiquitously expressed chaperones accounting for 1-2% of all cellular proteins in most cells. It has been demonstrated to play crucial roles in protein folding, protein degradation and signal transduction (Picard, 2002). The most well-known function of HSP90 concerns the maintenance of key proteins such as steroid receptors and protein kinases by forming specific complexes (Csermely et al., 1998). The literature shows that HSP90 can be regulated by a range of stressors such as heat or cold shock, food deprivation, heavy metals and diseases (Gao et al., 2007; Wang et al., 2012).

Planarians *Schmitdea mediterranea* are well-famous animal models for the studying stem cell regulation and central nervous system regeneration (Saló and Agata, 2012). Following the cut, a tiny fragment of planarians can regenerate a entire animal in 1-2 weeks. Another striking feature of planrians is that they show strong tolerance to prolonged starvation (Bowen et al., 1976). An adult planarian can survive several months without eating any food. As aquatic animals, planarians are easily threatened by water pollution. Nowadays, the characterization and functions of HSP90 in other vertebrates and invertebrates have been well discussed. However, there is no more information of HSP90 in planarian. In this paper, we used the methods of bioinformatics to study the gene structure of HSP90 from planarians *Schmitdea mediterranea* (designated SmedHSP90), to constructed phylogenetic tree of HSP90s across species and to analyse the three dimensional structure. The aim of this study was to appraise these methods for candidate gene struture/function analysis in the light of established laboratory-based knowledge of HSP90s.

2. Material and Methods

2.1 Idenfification of the SmedHSP90 nucleotide sequence and determination of the coding sequence

SmedHSP90 genomic and mRNA sequences were identified from the National Centre for Biotechnology Information (NCBI) internet site (www.ncbi.nlm.nih.gov) and the Schmidtea mediterranea genome database (http://smedgd.neuro.utah.edu) (Robb et al., 2008). Genomic and mRNA sequence alignment was performed using software Dnaman 6.0. Open Reading Frame Finder at the NCBI was used to examine putative coding sequences of SmedHSP90 and the output viewed against the Genbank database.

2.2 Promoter prediction

The Promoter Inspector facility and Chip2Promoter (www.geomatix.de/software_services/software/Promoter-Inspector/) were used to predict a promoter (Scherf, 2000). FirstEF (http://rulai.cshl.org/tools/FirstEF) was used to idenfify a promoter sequence, and the Promoter 2.0 Prediction Server (ww.cbs.dtu.dk/server/Promoter) to generate likelihood scores. The Eukaryotic Promoter Database (http://www.dpd.isb-isb.ch) was search and Neural Net-work Promoter Prediction (NNPP)

(http://www.fruitfly.org/seq_tools/promoter.html) used to generate putative promoter elements.

2.3 Characterization of SmedHSP90

The deduced amino acid sequence was analyzed with the Expert Protein Analysis System (http://www.expasy.org). Motif scan was performed against a number of motif databases (http://myhits.isb-sib.ch/cgi-bin/motif_scan) and by Simple Modular Architecture Research Tool (http://smart.embl-heidelberg.de/). Three-dimensional domain structure of SmedHSP90 was predicted by SWISS-MODEL Server (http://www.expasy.org/swissmod/SWISS-MODEL.html) (Arnold et al., 2006). Additional assessments of domain structures were performed on ProSA-Web (https://prosa.services.came.sbg.ac.at/prosa.php) and Verify3D Structure Evaluation Server (http://nihserver.mbi.ucla.edu/Verify_3D/).

2.4 Multiple sequence alignment and phylogenetic analysis

The similarity analysis of nucleotide and protein sequence was carried out by using blastn and blastp at web servers of the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast). A phylogenic tree was constructed using the programs of Clustal X 1.83 and Mega 3.1 based on the amino acid sequences of DjHS90 and other known HSP90 sequences. Bootstrap analysis was used with 1000 replicates to test the relative support for the branches produced by neighbor-joining analysis.

3. Results

3.1 Identification of the SmedHSP90 nucleotide sequence

Though a 'key word' search at the NCBI, we identified a previously reported planarian *Dugesia japonica* HSP90 (designated DjHSP90) cDNA sequence (Genbank accession no. FJ628362). Based on the nucleotide sequence of DjHSP90, we blasted the *Schmitdea mediterranea* genome dabatase (Robb et al. 2008), and found that three contigs sequences (contig 31.05671; 31.005595; 31.004356) show high homology with that of DjHSP90 (Fig.1).

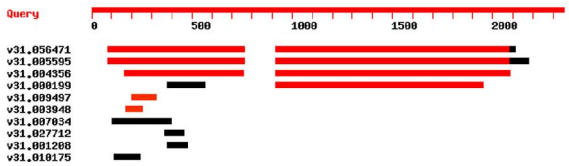

Fig.1 The nucleotide BLAST output for planarian *Schmitdea mediterranea* HSP90 genomic sequence. Three contigs (31.056471; 31.005595; 31.004356) have high alignment scores.

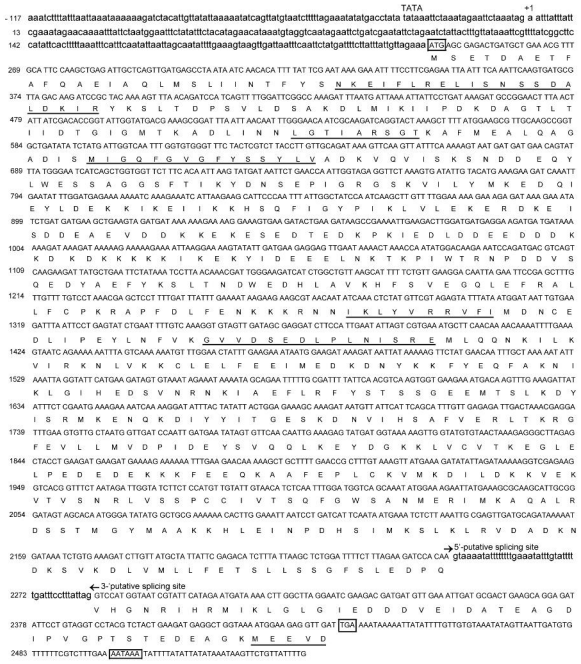

Fig. 2 The nucleotide sequence of the SmedHSP90 gene and the deduced amino acid sequence in the coding region are shown. The transcription start site is indicated by +1. The five highly conserved amino acid segments that characterize all member of the HSP90 family and the C-terminal pentapeptide MEEVD are shown underlined. The start codon, stop codon and the polyadenylation signal (AATAAA) were boxed. Arrow indicated the intron 5'- and 3'- splicing site, which is followed the typical "GT-AG" rule.

3.2 Gene structure and characterization of SmedHSP90

Alignments of DjHSP90 cDNA sequences and SmedHSP90 genomic sequences, we determined the ORF (open reading frame) sequences of SmedHSP90. The ORF of SmedHSP90 is 2148 bp encoding a polypeptide of 715 amino acids with a predicted molecular mass of 82.14 kDa and theoretical isoelectric point of 4.82. The deduced amino acid sequence of SmedHSP90 gene shares five conserved motifs of HSP90 protein family signature (NKEIFLRELISN[S/A]SDALDKIR, LGTIA[K/R]SGT, MIGQFGVGFYSS[Y/F]LV, IKLYVRRVFI, GVVDS[E/D]DLPLN[I/V]SRE) and the consensus sequence MEEVD at the C-terminus (underlined in Fig.2). Between the codon CAA and GTC, there is a 48 bp intron. The intron 5'- and 3'- splicing site follows the typical "GT-AG" rule (Fig.2). Viewscan the nucleotide sequence downstream the stop codon, there exists a canonical polyadenylation signal sequence AATAAA (Boxed in Fig.2). Using the promoter prediction tools, we found the putative TATA-box element upstream the start codon 269 bp.

The transcription start site is indicated by +1 (Fig. 2). Expasy program analysis revealed that the typical histidine kinase-like ATPase domain, which was ubiquitous in all HSP90 family members, was located amino acid residues 29-183 (Fig. 3).

Fig. 3 Prediction of histidine kinase-like ATPase domain.

3.3 Multiple sequence alignment and phylogenetic analysis

The deduced amino acid sequence of SmedHSP90 was close matched to other HSP90s in invertebrates and vertebrates (more than 72% similarity in all matches). It displays high similarity to HSP90s of *Dugesia japonica* (92.31%), *Schistosoma japonicum* (75.2%), *Mytilus galloprovincialis* (78.3%), *Drosophila melanogaster* (75.9%), *Homo sapiens* (76.24%), *Danio rerio* (74.52%) and so on. Multiple sequence alignment of SmedHSP90 with other known HSP90 amino acid sequences reveals that they are highly conserved, especially in the regions of HSP90 family signatures.

In vertebrates, there exists two different cytosolic isoforms of HSP90 gene (HSP90α and HSP90β) which are different in the structure of glutamine-rich sequence (QTQDQ) at the N-terminus (Csermely et al., 1998; Gao et al., 2007). In contrast, most invertebrates possess only one HSP90 gene. Obviously, SmedHSP90 and DjHSP90 lack QTQDQ sequence (Fig.4), which suggested that planarian HSP90 was more closely related to the vertebrate β-isoforms

```
Human α                 MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Rat α                   MPEETQTQDQPMEEEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Chicken α               MPEAVQTQDQPM--EEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Zebrafish α             MPEAHEQQ----MMEDEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Human β                 MPEEVHHG----------EEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Rat β                   MPEEVHHG----------EEEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Chicken β               MPEQVQHG----------EDEVETFAFQAEIAQLMSLIINTFYSNKEIFLR
Zebrafish β             MPEEMRQ-----------EEEAETFAFQAEIAQLMSLIINTFYSNKEIFLR
Dugesia japonica        MS----------------------DTDTETFAFQAEIAQLLSLIINTFYSNKEIFLR
Schmitdea mediterranea  MS----------------------ETDAETFAFQAEIAQLMSLIINTFYSNKEIFLR
```

Fig. 4 Alignment of N-terminal amino acid sequences of planarian HSP90s and other known HSP90s. Planarian HSP90s lack QTQDQ sequence at the N-terminus.

Based on the HSP90 of above and other species, a phylogenetic tree was constructed using programs of CLUSTAL X1.83 and MEGA3.1 (Fig.4). In the root of HSP90 phylogenetic tree is unicellular yeast, next to plant, coelenteratas, platyhelminthes, molluscs, arthropodas, vertebrates. All the vertebrates were clustered together and formed two branches (HSP90α and HSP90β isoform groups). Two species of planrians, *Schmitdea mediterranea* and *Dugesia japonica* were clustered together. The relationships displayed in the phylogenic tree was in good agreement with the concept of traditional taxonomy.

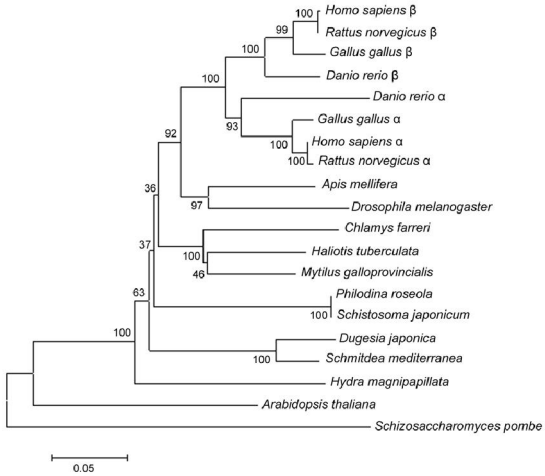

Fig. 5 A phylogenetic tree of HSP90 family members constructed with the neighbor-joining method. Number at each brach indicated the percentage of times a node was supported in 1000 bootstraps pseudoreplication by neighbour joining. The species names and the Genbank accession numbers were as follows: *Homo sapiens*, NP_005339 (HSP90α), NP_031381 (HSP90β); *Rattus norvegicus*, NP_786937 (HSP90α), P34058 (HSP90β); *Gallus gallus*, P11501 (HSP90α), CAA49704 (HSP90β); *Danio rerio*, NP_571403 (HSP90α), O57521 (HSP90β); *Apis mellifera*, XM_623936; *Drosophila melanogaster*, CAA27435; Chlamys farreri, AAR11781; *Haliotis tuberculata*, AM283515; Mytilus galloprovincialis, CAJ85741; *Philodina roseola*, ACC43981; *Schistosoma japonicum*, AAW27659; *Hydra magnipapillata*, XM_002155074; *Arabidopsis thaliana*, BAA00615; *Schizosaccharomyces pombe*, AAC41646.

3.4 The three-dimensional structure analysis of SmedHSP90

The three-dimensional structure analysis of SmedHSP90 revealed that it contains two clearly distinguishable domains: N-terminal domain and C-terminal domain. The highly conserved N-terminal polypeptides fold into a relatively flexible sphere domain (Fig.6A), which contains a pocket structure. Some important structural motifs, nests such as Ala113, Gly114, Ala115, Phe123, Gly124, Val125, are located in this pocket, which form binding sites for binding of ATP/ADP, antibiotic, substrate polypeptides and substrate target proteins, etc. The C-terminal polypeptides fold into a structurally flexible

domain (Fig.6B), which facilitat the forming dimer structure of SmedHSP90. Other reported that the C-terminal dimerization domain also provides the binding site for a set of co-chaperone molecules that function with HSP90 as part of a dynamic series of multi-chaperone complex.

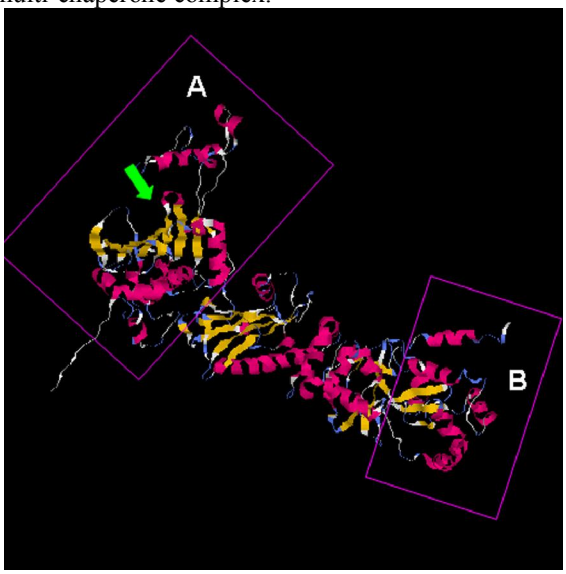

Fig. 6 Three-dimensional structure of SmedHSP90. A: N-terminal dodmain; B: C-terminal domain. Green arrow showing the position of bound ATP/ADP and others substrate polypeptides. α helices are shown in red, β sheets are shown in orange.

4. Discussions

The establishment of planarian *Schmitdea mediterranea* genome database provides an important tool to find genes of interest and their homologs in other species (Robb et al. 2008). Using this database, we have derived the gene structure and protein characterization of SmedHSP90. SmedHSP90 gene contains an ORF of 2148 bp encoding a polypeptide of 715 amino acids with the characteristics as those described in other invertebrate. Especially the highly conserved MEEVD sequences on the C-terminus is a character shared by all of the cytosolic HSP90 protein. In addition, SmedHSP90 gene cnontains one intron located upstream of the stop codon, which is very different from the inducible HSP70 gene (Ma et al. 2009). Three-dimensional structure of SmedHSP90 contains two important domain: N-terminal domain forms binding sites for binding of substrate molecules and C-terminal domain is recognized by co-chaperones which mediate the association of HSP70 and HSP90 into a multi-chaperone complex (Scheufler et al, 2000). Bioinformatic analysis of SmedHSP90 gene provides basic data for the study of stress response in planarians.

Acknowledgements:

This work was supported by the projects for young teachers in colleges and universities of Henan Province (2011GGJS-163

Corresponding Author:
Dr. Xingzi Xi
Department of Education Sciences
Xinxiang University
East Jinsui Road, Xinxiang 453003, China
E-mail: xingzixi2003@sina.com

References

1. Jackson SE. Hsp90: structure and function. Top Curr Chem 2013, 328:155-240.
2. Picard D. Heat shock protein 90, a chaperone for folding and regulation. Cell Mol. Life Sci 2002, 59: 1640-1648.
3. Csermely P, Schnaider T, Soti C, Prohaszka Z, Nardai G. The 90–kDa molecular chaperone family: structure, function, and clinical application: A comprehensive review. Phamacol Ther 1998, 79: 129-168.
4. Gao Q, Song L, Ni D, Wu L, Zhang H, Chang Y. cDNA cloning and mRNA expression of heat shock protein 90 gene in the haemocytes of Zhikong scallop *Chlamys farreri*. Comp. Biochem. Physiol. B 2007, 147: 704-715.
5. Wang SJ, Wu MJ, Chen XJ, Jiang Y, Yan YB. DsHsp90 is involved in the early response of Dunaliella salina to environmental stress. Int J Mol Sci 2012, 13(7): 7963-7979.
6. Saló E, Agata K. Planarian regeneration: a classic topic claiming new attention. Int J Dev Biol 2012, 6(1-3): 3-4.
7. Bowen ED, Ryder TA, Dark C. The effects of starvation on the planarian worm Polycelis tenuis Iijima. Cell Tissue Res 1976: 169: 193-209.
8. Robb SM, Ross E, Sanchez Alvarado A. SmedGD: the Schmidtea mediterranea genome database. Nucleic Acids Res 2008, 36: D599-D606.
9. Scherf M, Klingenhoff A, Werner T. Highly specific localization of promoter regions in large genomic sequences by PromoterInspector: a novel context analysis approach. J Mol Biol 2000, 297: 599-606.
10. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 2006, 22: 195-201.
11. Ma K-X, Chen G-X, Lou H, Fei L-N. Cloning and expression analysis of *hsp70* gene from freshwater planarian *Dugesia japonic*. Biologia 64:1018-1024.
12. Scheufler C, Brinker A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H. Structure of TPR domain-peptide complex: critical elements in the assembly of the hsp70-hsp90 multichaperone machine. Cell 2000, 101: 199-120.