Journal of Genetics, Vol. 96, No. 4, September 2017, pp. 613–623
DOI 10.1007/s12041-017-0816-5

RESEARCH ARTICLE

The first report of *diablo* in *Megalobrama amblycephala*: characterization, phylogenetic analysis, functional annotation and expression

NGOC TUAN TRAN[1,2], IVAN JAKOVLIĆ[1] and WEI-MIN WANG[1,3]*

[1]*Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, College of Fisheries, Huazhong Agricultural University, Wuhan 430070, Hubei, People's Republic of China*
[2]*Key Laboratory of Aquaculture Disease Control, Ministry of Agriculture, and State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, Hubei Province, People's Republic of China*
[3]*Collaborative Innovation Center for Efficient and Health Production of Fisheries, Changde 41500, Hunan Province, People's Republic of China*
*For correspondence. E-mail: wangwm@mail.hzau.edu.cn.

Received 4 October 2016; revised 5 December 2016; accepted 4 January 2017; published online 5 September 2017

Abstract. *Smac/DIABLO* gene is essential for the apoptosis mechanism in mammals. This study is the first report of the *Megalobrama amblycephala* (*ma*) *diablo* gene, and the first report of the tertiary structure of a Diablo polypeptide in fish. *Madiablo* is 1540-bp long with an open reading frame of 792 bp, encoding a putative protein of 263 amino acids with a molecular weight of 29.2 kDa. Phylogenetic analysis indicates that it is closely related to the zebrafish Diablo-a homologue. It also indicates the existence of two *diablo* copies (a and b) in teleosts; apart from the Percomorpha group, where *diablo-b* has been lost, but *diablo-a* had undergone an independent duplication. Madiablo protein contains a long Smac_DIABLO super family domain (Leu32–Asp263) and alpha helices were prevalent in the secondary structure. Homology model of madiablo protein was constructed using the comparative modelling method. Expression of *madiablo* mRNA transcript was investigated using qPCR: (i) in five tissues from a healthy blunt snout bream, indicating the highest constitutive expression level in liver. (ii) During the embryo and juvenile development, indicating a spike in expression during hatching and in later phases of the juvenile development. (iii) In response to *Aeromonas hydrophila* infection, indicating the downregulation in liver, spleen and kidney during the first 12 h postinfection and upregulation in spleen and kidney after 24 h postinfection (hpi). The results imply that *madiablo* is homologous to *Diablo* orthologues in other species, both structurally and functionally, and that, it probably plays a role in the immune system of *M. amblycephala*.

Keywords. Diablo; physicochemical characterization; homology modelling; functional annotation; gene expression; *Megalobrama amblycephala*.

Introduction

Apoptosis, or programmed cell death, has a crucial role in the normal development, homeostasis and immune response of all multicellular organisms. By committing to this type of cell suicide, redundant or abnormal cells that represent a threat to the organism are efficiently removed without affecting the neighbouring normal cells (Steller 1995; Alnemri *et al.* 1996; Takle and Andersen 2007; de Almagro and Vucic 2012). Apoptosis can be activated through the extrinsic or death-receptor-mediated pathway and the intrinsic or mitochondrial pathway. Extrinsic apoptotic pathway is initiated by the activation of death receptors of the tumour necrosis factor receptor super family by their respective ligands, leading to the activation of caspases, while the intrinsic pathway is initiated by the

Electronic supplementary material: The online version of this article (doi:10.1007/s12041-017-0816-5) contains supplementary material, which is available to authorized users.

cellular stress, growth serum withdrawal, DNA damage, radiation, or other stress signals resulting in alterations of the outer mitochondrial membrane potential and permeability. The response promotes release of cytochrome c and second mitochondrial activator of caspase / direct inhibitor of apoptosis protein-binding protein with low pI (Smac/DIABLO) from mitochondria. Smac/DIABLO binds and neutralizes the 'inhibitor of apoptosis' proteins (IAPs), leading to the activation of caspases and apoptosis (Du *et al.* 2000; Verhagen *et al.* 2000; Verhagen and Vaux 2002; Justo *et al.* 2003; Honda *et al.* 2004; Shiozaki and Shi 2004; Montesanti *et al.* 2007; de Almagro and Vucic 2012). When cytochrome c is released into the cytosol it binds to apoptosis protease activating factor (APAF1) and forms the apoptosome, leading to the activation of caspase-9, caspase-3 and caspase-7 directly by binding to them. Meanwhile, Smac binds XIAP, preventing its inhibition of caspase-3, caspase-7 and caspase-9, thus, removing the apoptosis blockade (Vaux and Silke 2003). Previous researches are overwhelmingly associated Smac/DIABLO with apoptosis, including the regulation of apoptotic responses of cancer cells (Mizutani *et al.* 2005), both in humans (Verhagen *et al.* 2000; Liu *et al.* 2013; Zhang *et al.* 2013) and in mice (Okada *et al.* 2002; Justo *et al.* 2003).

Apoptosis attends to the development of virtually all cell lineages and is crucial in response to many physiological and pathophysiological stimuli (Ekert and Vaux 1997; Opferman 2007). Apoptotic pathways are quite conserved among mammals and aquatic animals, including fish, including the existence of two different pathways for caspase activation (Yamashita 2003; dos Santos *et al.* 2008; Romano *et al.* 2013). Apoptosis is observed in response to aquatic diseases and environmental stressors, but also during the ontogeny (Cole and Ross 2001; Takle and Andersen 2007). However, while there are several fish *diablo* sequences deposited in the NCBI GenBank, which are transcriptome studies, whereas only two publications have so far studied this gene in fish, both as a biomarker for pollutant exposure (Zacchino *et al.* 2012; Asker *et al.* 2013). Consequently, expression patterns of *diablo* and its function in fish immune system are presently poorly understood.

Blunt snout bream (*Megalobrama amblycephala* Yih 1955), endemic to China, is becoming increasingly economically important aquaculture species in China (FAO 2013). Along with the increase in production, it has also become increasingly vulnerable to various pathogens, often causing catastrophic economic losses, prompting the need for a better understanding of its immune system, including the genes involved in pathogen recognition and of triggering the immune response (Zhou *et al.* 2008; Ji *et al.* 2014).

Hence, the objectives of this study were to determine the presence (indirectly) and functions of *diablo* gene in *M. amblycephala*. To corroborate this, it is indeed a structural and functional homologue of the mammalian Smac/DIABLO, its physicochemical, structural and functional properties were investigated. Constitutive *madiablo* mRNA expression has been analysed in five different tissues of healthy blunt snout bream specimens. As *diablo* homologues in other species have been associated with both the immune response and development (in association with the apoptosis), also the expression was studied during and after the embryonic development. To study the role in the immune response, the main pathogen causing haemorrhagic septicaemia (a disease causing particularly severe losses in this species) is *Aeromonas hydrophila* (Nielsen *et al.* 2001; He *et al.* 2006; Tran *et al.* 2015a), which was chosen to investigate the expression levels after a pathogen challenge.

Materials and methods

Fish, tissue samples, embryo samples and challenge experiment

Healthy blunt snout bream (260 specimens, average body weight: 27.3 ± 6.8 g) were obtained from the Tuanfeng fish farm (Hubei, China). The fish were kept in 1 m^3 tanks with aeration at about 28°C for two weeks at the College of Fisheries, Huazhong Agricultural University, Wuhan (China) and fed commercial pelleted feed twice a day. For the tissue-specific expression analyses, three healthy specimens were anesthetized with 100 mg/L tricaine methanesulfonate (MS-222, Sigma-Aldrich, St Louis, USA) and dissected to collect five type-tissue samples (gills, liver, spleen, kidney and intestine). The remaining specimens were divided into control ($n = 87$) and experimental ($n = 151$) groups. Fish from both the groups were intraperitoneally injected equal volumes (0.1 mL) of sterile physiological saline (control) and *A. hydrophila* suspension at 1.8×10^6 colony forming unit (cfu)/mL (experimental). Six specimens from each group were sampled at five timepoints (30 specimens from each group in total): 4, 12, 24, 72 and 120 h postinfection (hpi), anesthetized in MS-222 and dissected. Liver, spleen and kidney were sampled from all specimens, immediately flash-frozen in liquid nitrogen and stored at −80°C.

To determine the expression patterns at different embryo and juvenile developmental stages, samples (embryos were obtained through artificial breeding of three female and three male adult from the Tuanfeng fish farm) were taken at 17 developmental stages (Li and Zhang 1993), and confirmed by microscopic observation: fertilized egg, 0-h postfertilization (hpf) (30 eggs); 32-cell stage, 3 hpf (30 embryos); early gastrula, 5 hpf (30 embryos); body segment appearance, 14 hpf (30 embryos); caudal bud appearance, 17 hpf (30 embryos); caudal fin enlargement, 21 hpf (30 embryos); heart appearance, 24 hpf (30 embryos); heart-beat appearance, 27 hpf (30 embryos); hatching, 33 hpf (50 specimens); body pigment appearance, four days postfertilization (dpf) (50 specimens); air bladder

formation, 5 dpf (50 specimens); intestine appearance, 6 dpf (50 specimens); chordal tip lifting, 12 dpf (50 specimens); anal fin formation, 28 dpf (15 specimens); 40 dpf (five specimens); 66 dpf (three specimens); 90 dpf (1 specimen). Samples were immediately frozen and stored at −80°C. All fish were handled and experiments were conducted in accordance with the Experimental Animal Regulation in Hubei Province, Standing Committee of People's Congress of Hubei Province (code no. 50; issue date: 29 July 2005).

Total RNA preparation and cDNA synthesis

Total RNA was extracted from each sample with RNAiso-Plus reagent (Takara Bio, Dalian, China), according to manufacturer's instructions. In a single sample, RNA from all the specimens were pooled. Quality and quantity of the extracted RNA were checked by electrophoresis in 1% agarose gel and Nanodrop 2000 spectrophotometer (Thermo Scientific, Delaware, USA). Equal amounts of the total RNA of specimens sampled at each time-point were pooled. cDNA library was synthesized using PrimeScript RT Reagent kit with gDNA Eraser (Takara Bio, Dalian, China) following the manufacturer's instructions (RT Primer Mix contains Oligo DT primer and random 6-mers), serially diluted 10-fold and used as the template for quantitative real-time PCR (qPCR).

Identification of madiablo gene

Full-length putative blunt snout bream *diablo* (*madiablo*) transcript was obtained from the blunt snout bream transcriptome profile, constructed in a previous study using Solexa/Illumina technology (Tran *et al.* 2015b) and identified by BLAST homology search against other orthologous sequences available in the GenBank database (http://blast.ncbi.nlm.nih.gov/Blast). The open reading frame (ORF) and amino acid sequence of the putative Madiablo polypeptide were inferred using NCBI's Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder/), signal peptide using the SignalP server (http://www.cbs.dtu.dk) and domain structures using CD-Searchserver (http://www.ncbi.nlm.nih.gov). Homologous Diablo amino acid sequences from other species were obtained from the NCBI database. Multiple-sequence alignment was performed using ClustalW2 server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Phylogenetic analyses were performed, their relationships were analysed and trees were constructed using MEGA 5.2 (Tamura *et al.* 2011), MrBayes (Ronquist and Huelsenbeck 2003) and FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Bayesian analysis was performed twice with 1,700,000 iterations. Figures were edited using Gimp software (www.gimp.org).

Protein physicochemical and functional characterization

Expasy's ProtParam prediction server (http://web.expasy.org/protparam/) was employed to determine the following physicochemical properties of Madiablo (263 amino acid, aa) polypeptide chain: molecular weight, aa composition, theoretical isoelectric point (pI), total number of positive and negative residues, extinction coefficient (EC), instability index (II), aliphatic index (AI) and grand average of hydropathicity (GRAVY). SOSUI tool (Hirokawa *et al.* 1998) was used to identify the types of protein (soluble or membrane) and CYS_REC (http://linux1.softberry.com/) to predict the presence of disulphide bonds and their bonding patterns.

Protein structure prediction

Secondary structure of Madiablo was predicted by self-optimized prediction method using Alignment (SOPMA) server (Geourjon and Deleage 1995) with default parameters (window width, 17; similarity threshold, 8; number of states, 4). Three-dimensional (3D) homology models were constructed by using online servers GeneSilico Metaserver (https://www.genesilico.pl/meta2/), SWISS-MODEL (http://swissmodel.expasy.org/) and COFACTOR (http://zhanglab.ccmb.med.umich.edu/). These programs align the input target with pre-existing PDB templates to generate a series of predicted models. The most suitable template to build the 3D model is then selected on the basis of sequence identity (Fiser 2010). Stereochemical quality and accuracy of the predicted models were analysed using PROCHECK's Ramachandran plot analysis, ERRAT, Verify3D (all three are available at SAVES server, http://services.mbi.ucla.edu/SAVES/), ProQ protein quality prediction (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) and ProSA-web protein structure analysis (https://prosa.services.came.sbg.ac.at/prosa.php).

qPCR and statistics

qPCR was performed using SYBR Premix Ex Taq (Takara, Dalian, China) in a Rotor-Gene Q real-time PCR cycler (Qiagen, Dusseldorf, Germany). The reaction volume of 10 mL contained 5 μL of SYBR Premix Ex Taq II (2×), 0.4 μL of each primer (10 μM), 0.8 μL of cDNA template and 3.4 μL of distilled H_2O. Both 18*S* rRNA and *beta*(β)*-actin* were tested, and the latter chosen as the most suitable reference gene for developing embryos; and 18*S* rRNA was used as the reference gene for gene expression under *A. hydrophila* infection (Luo *et al.* 2014). Primers were designed and synthesized using Sangon Biotech (Shanghai, China) (table 1). Thermal conditions were as follows: denaturation at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, annealing temperature (shown in table 1), for 20 s and elongation at 72°C for 15 s. All

Table 1. Primers used in this study.

Primer name	Primer sequence (5′-3′)	T_m (°C)
Madiablo-F	CCAAGGCACTACATACACT	56
Madiablo-R	TCTATCACTAACCTCCACTCT	
β-actin-F	CCAAAGCCAACAGGGAAA	58
β-actin-R	GAGGCATACAGGGACAGCAC	
18S rRNA-F	CGGAGGTTCGAAGACGATCA	60
18S rRNA-R	GGGTCGGCATCGTTTACG	

reactions were performed in triplicate and data was analysed using Rotor-Gene Q series software 1.7 (build 94) (Qiagen, Dusseldorf, Germany). Ct values were obtained using the threshold value of 0.05 for genes. Relative gene expression was quantified by the comparative Ct method, expressed as $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen 2001). Relative expression of the target genes was normalized to an endogenous reference (*18S* rRNA), where ΔCt was calculated as $Ct_{Test} - Ct_{18S\,rRNA}$, and ΔΔCt was calculated as $\Delta Ct_{Test} - \Delta Ct_{Control}$. The qPCR data were analysed statistically using Microsoft Excel, with one-way analysis of variance (ANOVA) performed using SPSS 16.0 software. Differences were considered statistically significant at $P < 0.05$ and $P < 0.01$.

Results

Sequence and phylogenetic analyses

Madiablo transcript, deposited in the NCBI GenBank database under the accession number KT805200 is 1540-bp long, contain an ORF of 792 bp, encoding a putative protein of 263 amino acids. Protein BLAST search revealed that the closest Madiablo polypeptide match, with 90% sequence identity, is the *Danio rerio* Diablo sequence in the EST database (accession number NM_200346.1). Multiple alignment analysis revealed that Madiablo also shares a relatively high amino acid sequence similarity (40–90%) with homologues in other teleost species, such as Mexican tetra (*Astyanax mexicanus*) (67%) and, again, zebrafish (*D. rerio*) (90%) (figure 1). The sequence contained a single typical polyadenylation signal (AATAAA) at the position 1519 and two mRNA instability motifs (ATTTA) at positions 1203 and 1507. Start (ATG) and stop (TAG) codons were determined from the 5′ end of the sequence at the nucleotide positions 274 and 1063 bp, respectively. Most of the deduced Madiablo amino acid sequence (Leu32 to Asp263) corresponds to a typical super family domain Smac_DIABLO. No N-terminal signal peptide was found and only one N-linked glycosylation site ('NLSH' at position 88) was found in the protein sequence. The consensus IAP-binding motif (IBM) 'AVPF' was identified in the sequence at the position 80.

Phylogenetic tree, constructed using homologous amino acid sequences from multiple species (figure 2; figure 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/), also suggests that the closest relative is the zebra fish Diablo-a orthologous sequence. There were no other cyprinid sequences available in the GenBank. Tetrapod homologues, including a mammalian (*M. musculus*), a reptilian (*C. picta*), and an amphibian (*X. laevis*) sequence, were evolutionary distant from teleosts. Two nonteleost fish sequences included in the analysis (*L. chalumnae* and *L. oculatus*) were more closely related to tetrapods than to the (monophyletic) teleost clade. The latter was divided into two subclades: Diablo-a and Diablo-b. Madiablo sequence was placed within the Diablo-a subclade. Diablo-a sequences belonging to the Percomorpha group in this clade were further subdivided into two clades (aa and ab). Apart from the *C. harengus* (62) node, all the remaining nodes were supported with the maximum posterior probability value (100) and Jones was selected ($P = 1.0$) as the amino acid substitution model.

Physicochemical and functional characterization

A putative protein sequence of 263 aa was used as the template for physicochemical characterization. Molecular weight of the putative protein was 29.2 kDa. The pI of the protein was 6.54, implying acidic character (<7). The EC of Madiablo measured at 280 nm was 20440 M^{-1} cm^{-1} (assuming all pairs of cysteine residues form cysteines) and 19940 M^{-1} cm^{-1} (assuming all cysteine residues are reduced). The total number of negative and positive residues were 33 and 32, respectively. The estimated half-life of the protein is 30 h in mammalian reticulocytes *in vitro*, over 20 h in yeast *in vivo* and over 10 h in *Escherichia coli in vivo*. The II value of Madiablo protein (II = 54.45) suggests that it is probably unstable in a test tube (II > 40) (Guruprasad *et al.* 1990). High AI value (AI = 98.37) indicates high thermostability (Ikai 1980), while low GRAVY value (−0.170) implies that Madiablo is a hydrophilic protein. It is rich in alanine (11.8%) and leucine (11.0%) and classified as 'soluble protein'. There are nine cysteine residues in the sequence and the most probable pattern of their pairing is Cys_{79}–Cys_{201} and Cys_{100}–Cys_{171}, indicating the presence of disulphide

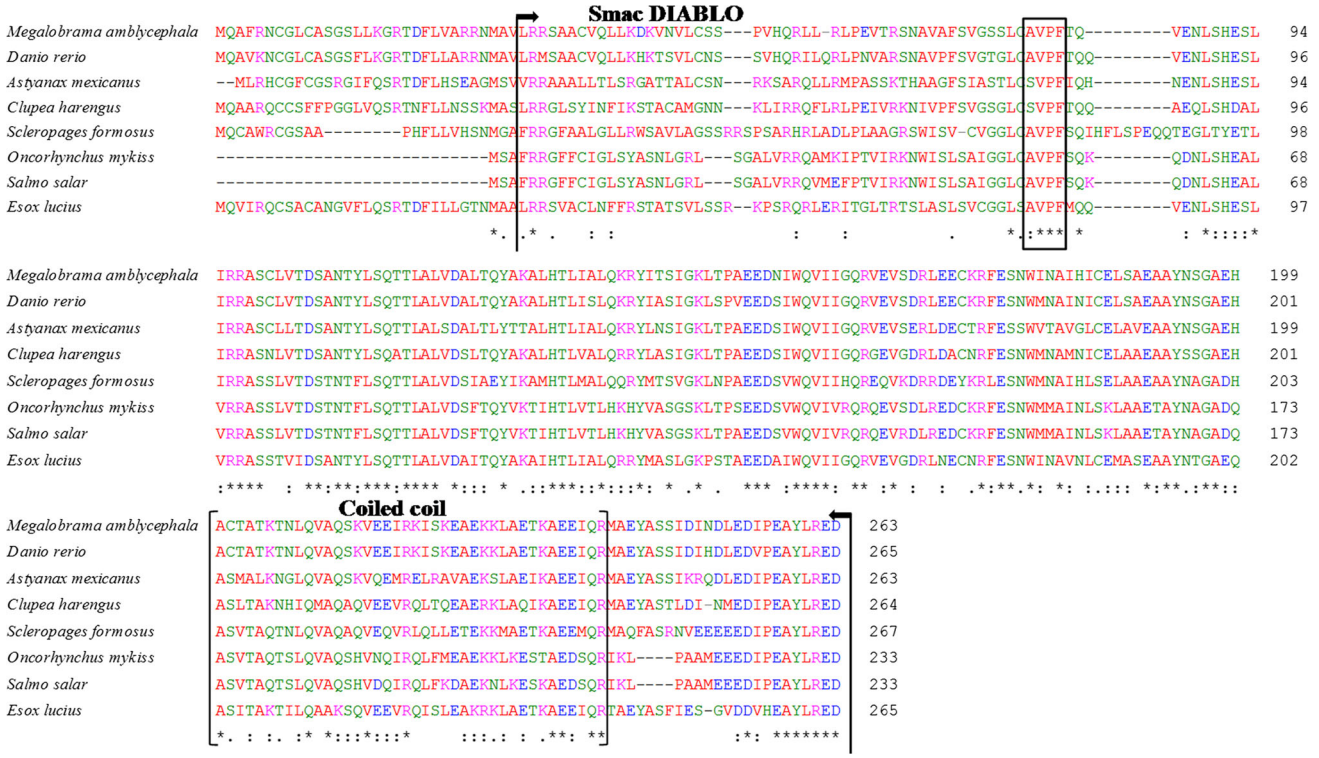

Figure 1. Multiple amino acid sequence alignment of Madiablo with Diablo homologues of six fish species: *Danio rerio* (GenBank accession number: NP_956640.1), *Astyanax mexicanus* (XP_007237773.1), *Clupea harengus* (XP_012694273.1), *Scleropages formosus* (KPP70654.1), *Oncorhynchus mykiss* (ACO07657.1), *Salmo salar* (ACI70050.1) and *Esox lucius* (ACO14405.1). *Identical amino acids. Sequences are numbered on the right; ':' conserved substitutions; '.' semiconserved; '–' deletions. Smac_Diablo of the IAP-binding motif 'AVPF' is boxed.

bonds that are essential in the folding of the protein and responsible for stabilization of protein structure (Hogg 2003).

Protein structure prediction and model validation

Alpha helix (65.02%) was dominant among the secondary structure elements, followed by random coil (19.01%); extended strand (11.03%) and beta turn (4.94%). Other structure elements, including 3_{10} helix, Pi helix, beta bridge, bend region and ambiguous states, were not predicted. SOPMA server was also used to analyse the remaining available fish diablo sequences: *D. rerio* (acc. no.: NP_956640.1), *E. lucius* (acc. no.: ACO14405.1), *S. salar* (acc. no.: ACI70050.1) and *O. mykiss* (acc. no.: ACO07657.1); in all of which alpha helix represented between 66.1% (*O. mykiss*) and 71.3% (*E. lucius*) of the secondary structure.

On the basis of the GeneSilico Metaserver search, human Smac/DIABLO (PDB ID: 1few, Chai *et al.* (2000)) at 2.2 Å resolution was identified as the best template for Madiablo with a high Z-score of 47.21 (a cut-off of Z-score ≥ 6.0 indicates a 99% confidence) (Kurowski and Bujnicki 2003). The same template (PDB ID: 1few) was also confirmed by Swiss-Model analysis, which revealed 43.65% identity and model coverage from Asn88 to Asp263. Three-dimensional structure of the predicted model is shown in figure 3a. Structural similarity between Madiablo and human Smac/DIABLO (1few) was additionally corroborated by COFACTOR analysis, where the calculated TM-score was 0.982. As TM-scores >0.5, generally indicate that two proteins have the same fold (Roy *et al.* 2012), this suggests a very high level of structural identity (figure 3b).

The proposed 3-D model was validated using many different tools, all of which indicate a good quality of the proposed model. Ramachandran plot analysis results reveal that the Madiablo model has 94.7% of residues in the most favoured region, suggesting a good quality (over 90%) of the homology model (Ramachandran *et al.* 1963). The overall average G-factor of dihedral angles and main-chain covalent forces was 0.26, indicating a very good quality (>−0.5) of the proposed model (Xu *et al.* 2000). Results from ERRAT analysis showed 97.619 overall quality, indicating a reliable high resolution and good quality (>95%) of the structured model (Colovos and Yeates 1993). Verify 3D analysis revealed that 45.45% of the residues had an averaged 3D-1D score ≥0.2, and 77.27% of the residues had positive scores (cut-off score is >0), implying that the predicted model is valid (Liithy *et al.*

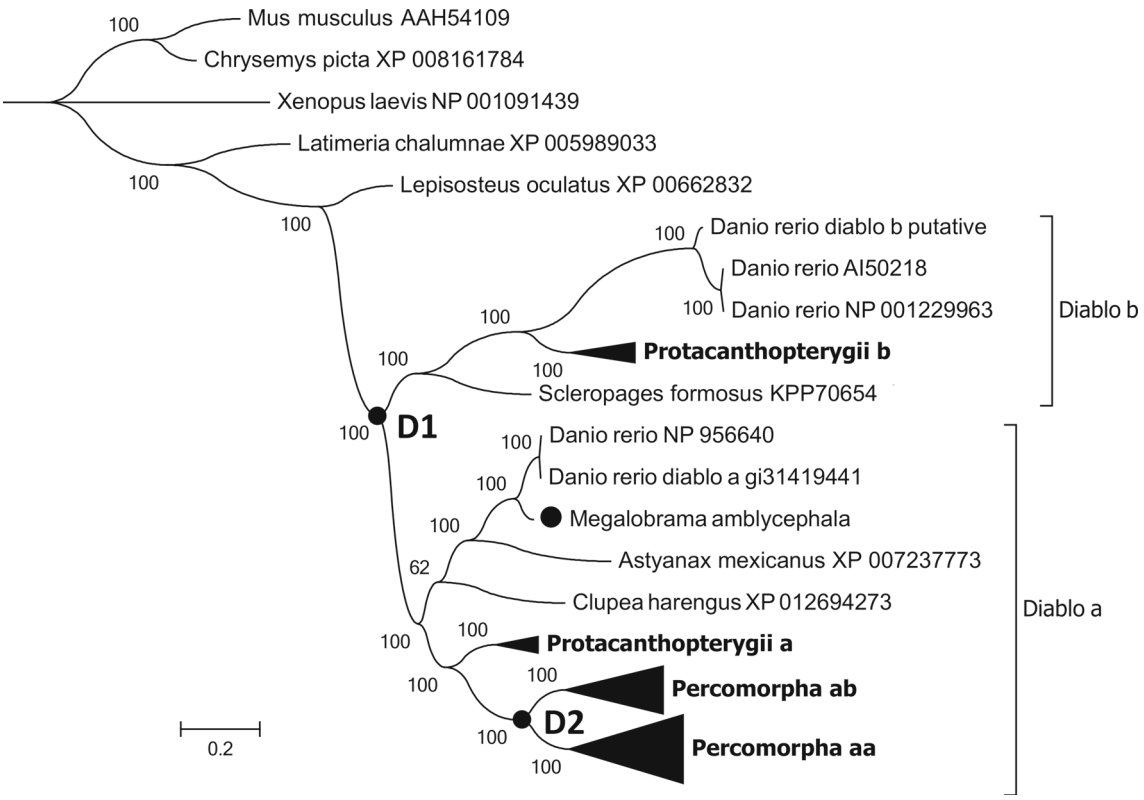

Figure 2. Phylogenetic tree showing relationships between the Madiablo and selected 58 homologues, as inferred by Bayesian analysis. Numbers at nodes indicate posterior probabilities, the bar (0.2) indicates genetic distance, and Madiablo is marked by a black dot. Putative positions of Diablo duplication events are indicated in the figure by black dots: D1, teleost-specific whole genome duplication; D2, percomorpha-specific Diablo duplication event. Full information about sequences, including the ones in the collapsed clades, is available in electronic supplementary material.

1992). The quality of the model was evaluated as 'correct' by both LG score (1.942) and MaxSub (0.198) analyses (Cristobal *et al.* 2001). The Z-score (−6.43) of the predicted model is within the range of scores typically found in the native proteins of similar size. The plot of residue energies, where positive values correspond to problematic or erroneous parts of the input structure, revealed that most of the calculated values are negative, also confirming good quality of the model (Wiederstein and Sippl 2007) (figure 4). The verification results suggest that the proposed model of Madiablo can be accepted with high confidence.

Expression of madiablo in tissues and during the development

Expression of *madiablo* was analysed in five different tissues from healthy blunt snout bream specimens. For expression levels normalized (set as 1) to the (control) tissue with the lowest expression level (intestine), *madiablo* was barely detectable in gills (0.0) and spleen (0.1), seven times higher in kidney, while extremely high expression was detected in liver (2339 times higher).

Madiablo expression pattern was also analysed during the development from the fertilized egg to 90 days-old juveniles (figure 5). Expression was normalized to the 'fertilized egg' stage (set as 1). *Madiablo* expression was detected in all the sampled stages, albeit it was barely detectable ($\leq$0.1) during the 'air bladder' and 'intestine appearance' stages. A spike in the expression (4-fold) was observed during hatching. After a period of low expression during the early larval stages, *madiablo* was highly upregulated in the entire period after the 'chordal tip lifting' (12 days) stage, with the highest expression at 66 days posthatching (10.2 $\pm$ 0.11-fold).

Expression of madiablo transcript after A. hydrophila infection

The expression of *madiablo* transcript after challenge with *A. hydrophila* was studied in the blunt snout bream liver, spleen and kidney, at five different time points (4, 12, 24, 72 and 120 hpi) using qPCR (figure 6). In comparison to the control group (normalized to the internal control gene *18S* rRNA expression levels), *madiablo* gene was significantly ($P < 0.01$) downregulated in liver with four (0.00-fold) and 12 (0.03-fold) hpi, with expression returning to the normal level (1-fold) subsequently. A similar expression pattern was observed in spleen, with downregulation at four (0.05-fold, $P < 0.01$) and 12 (0.22-fold, $P < 0.05$) hpi, but (nonsignificant) upregulation at 24 and 72 hpi.

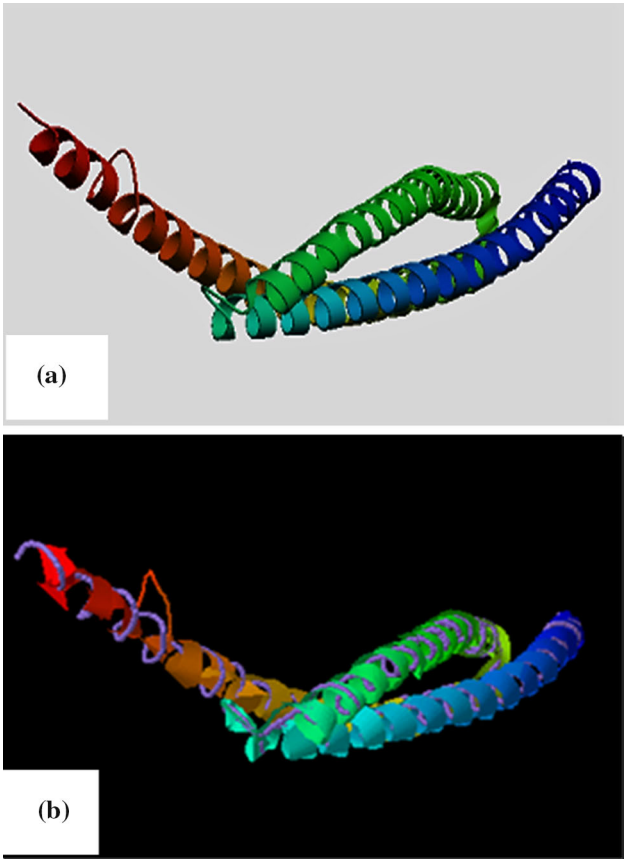

Figure 3. Three-dimensional structure of the Madiablo protein: (a) 3D homology model rendered by the SWISS-MODEL program. (b) Structural alignment of Madiablo (shown in cartoon) and human Smac/DIABLO structural analogue (PDB ID: 1few) (backbone trace) rendered by COFACTOR server.

Expression of the gene in kidney at 4 hpi was close to the control group level (1.08-fold), downregulated at 12 hpi (0.05-fold, $P < 0.05$), upregulated at 24 hpi (4.25-fold, $P < 0.05$), standard at 72 hpi (0.76-fold) and nonsignificantly upregulated at 120 hpi (2.12-fold).

Discussion

This study is the first identification and characterization of a *diablo* homologue in the blunt snout bream. Structural characteristics indicate that Madiablo polypeptide is a homologue of Smac/DIABLO reported in other animal species. The presence of a conserved super family Smac_DIABLO domain and the IAP-binding motif 'AVPF' in the tetrapeptide consensus sequence, both previously identified in other species, evidences a high level of structural and functional conservation among the members of this protein family (Du *et al.* 2000; Montesanti *et al.* 2007). As the secondary structure of a protein is crucial for further folding and higher structural levels, high similarity among Madiablo and other available fish homologues suggests that the tertiary structure of this protein is probably also very similar among these species. High similarity with mammalian homologues was further corroborated by homology modelling. And finally, as the tertiary structure of a protein is in close association with its function (Hoang *et al.* 2003), this further indicates a high level of functional conservation.

Phylogenetic analysis further confirmed that the putative Madiablo sequence is certainly a homologue of zebrafish Diablo, as it was placed in a monophyletic cyprinid cluster with two zebrafish sequences. Two Diablo paralogues (a and b) have been found in zebrafish; and as Madiablo clustered with Diablo-a, this suggests that: (i) the studied sequence could be *madiablo-a* paralogue; (ii) *madiablo-b* paralogue might also be present in the blunt snout bream genome. A previous research (Zacchino *et al.* 2012) of Diablo in European flounder (*Platichthys flesus*) has used 13 polypeptide sequences from eight vertebrate species to construct a phylogenetic tree. Based on the topology, they have concluded that tetrapods have lost a *diablo* gene retained by cyprinids, while in Acanthopterygians, a loss of one of the paralogues was followed by a subsequent intrachromosomal duplication of *diablo*. A much larger number of sequences included, as well as an exhaustive Bayesian approach used in this analysis, provide a much better topology resolution that corroborates these propositions, albeit with some modifications. Since two Diablo copies appear to be present only in teleosts, it is likely that they originate from the whole genome duplication in the common ancestor of teleosts (Hoegg *et al.* 2004). This was most probably followed by subsequent intrachromosomal *diablo-a* duplication (common in fish) (Robinson-Rechavi *et al.* 2001) in the common ancestor of Percomorpha (possibly even Acanthopterygii, but more sequences would be needed to corroborate this). As none of the Percomorph sequences included in the analysis clustered within the Diablo-b subclade, this indicates a possibility of the loss of Diablo-b paralogue in this teleost group. It is possible that the existence of two functional *diablo-a* copies (aa and ab) subsequently led to a loss of functionality in *diablo-b*. It would be necessary to identify and compare Diablo homologues in a number of other species to corroborate this scenario.

High *madiablo* expression levels observed in healthy blunt snout bream kidney and liver are comparable to observations for human Smac (Du *et al.* 2000) and mouse Smac/Diablo (Justo *et al.* 2003). Most importantly, the expression of *pfDiablo1* homologue in the European flounder tissues exhibited a very similar pattern: it was also barely detectable in gills, and by far, the highest in liver. However, *pfDiablo2* homologue expression was high in gills and very low in liver (Zacchino *et al.* 2012), suggesting that another Diablo homologue might exist in the blunt snout bream genome, and that *madiablo* described in this study could actually be *madiablo 1*. Expression of *madiablo* mRNA was consistent throughout the embryonic and

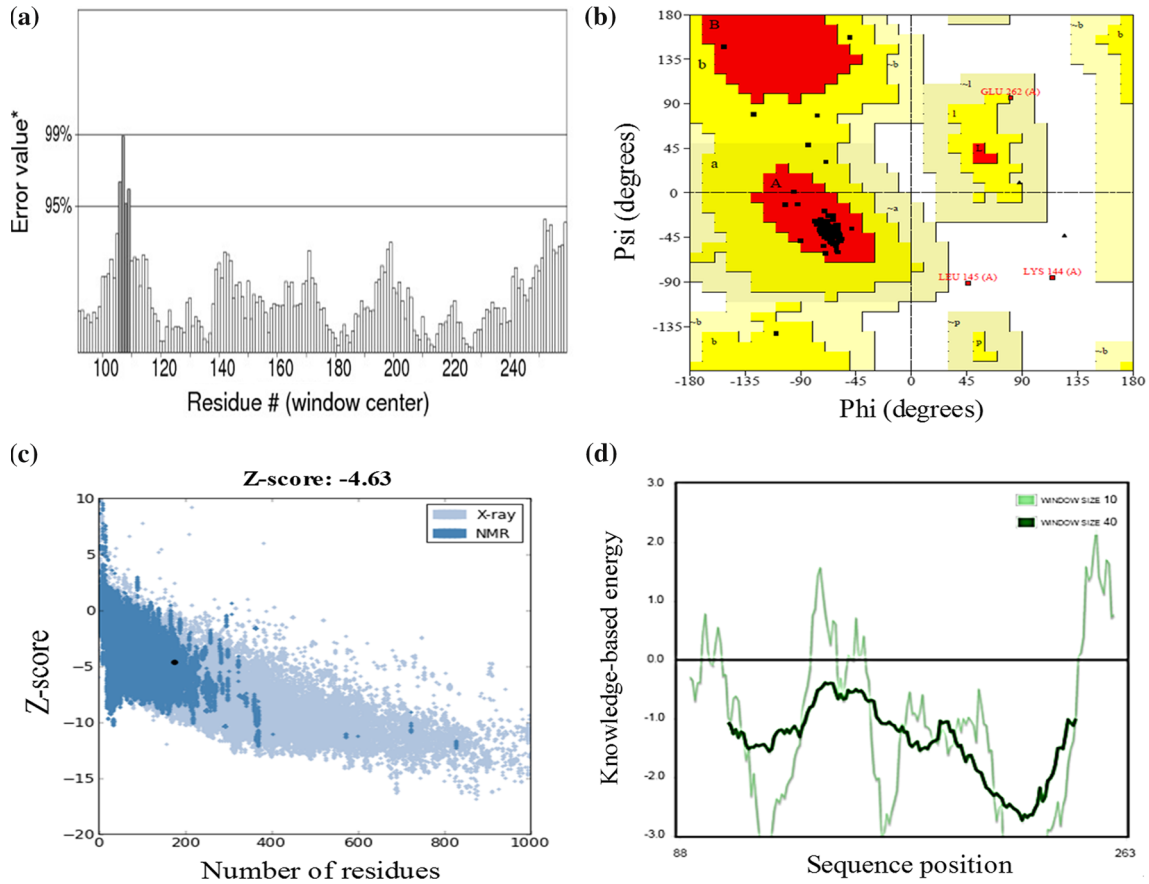

Figure 4. Validation results of the predicted Madiablo protein model: (a) overall quality of the model evaluated by the ERRAT program. On the error axis, two lines (95 and 99%) indicate the confidence with which it is possible to reject regions that exceed that error value. Regions of the structure highlighted in grey and black are rejected at 95 and 99% confidence levels, respectively. (b) Ramachandran plot analysis, indicating residues in the favoured regions (red), allowed regions (yellow), generously allowed regions (light yellow) and disallowed regions (white). (c) Z-score (highlighted as a black dot) is displayed in a plot that contains the Z-scores of all experimentally determined protein chains currently available in the Protein Data Bank. Groups of structures from different sources (X-ray and NMR) are distinguished by different colours (light blue and dark blue, respectively). (d) Plot of single residue energies, where window sizes of 40 and 10 residues are distinguished by dark green and light green lines, respectively. Positive values indicate problematic or erroneous parts of the structure.

larval developmental stages, and while there are currently no comparable published data in fish, this is similar to a previous observation in preimplantation mouse embryos (Honda *et al.* 2004). A spike in expression during the hatching was also observed in zebrafish *p8* gene, which is similar to *diablo*, has been associated with apoptosis, stress, cellular growth, tumour development, embryo development and immunity (Sun *et al.* 2010). It is possible that the increased expression of these two genes during hatching is a consequence of high physiological stress levels during this period. Higher (in comparison to the early development stages) expression of *madiablo* during the later stages of juvenile development might indicate that it plays an important role in cellular processes involved in development of blunt snout bream juveniles. However, as the number of sampled specimens decreased with time, individual variability in expression levels may have played a larger role in the last three stages (40–90 days old juveniles).

Only a few studies of this gene have so far been undertaken in fish, two of which studied *diablo* as a biomarker for pollutant exposure. Asker *et al.* (2013) have reported higher mRNA expression levels in *Zoarces viviparus* from a contaminated harbour in comparison to a reference site. Zacchino *et al.* (2012) found that *PfDiablo1* homolog expression in liver three days after treatments was not significantly different from the control group in all the pollutants investigated. In contrast, *PfDiablo2* mRNA was significantly upregulated after Arochlor 1254 and 3MC treatments. In this study, the upregulation of the expression of *madiablo* observed in spleen (24 and 72 hpi) and kidney (24 and 120 hpi) indicates its importance in response to bacterial infections, especially in the later stages of infection, as observed in spleen and kidney. Moreover, as *madiablo* was downregulated in all tested tissues during the first stages of infection (from 4 to 12 hpi), this indicates that the two (putative) *diablo* homologues might have

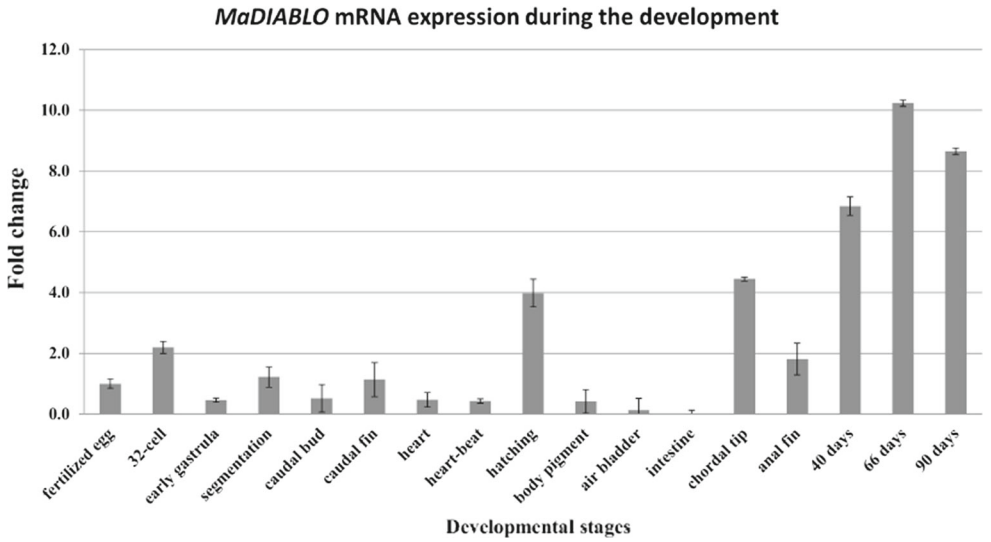

Figure 5. Expression profile of *madiablo* mRNA during the physical development of blunt snout bream: from the fertilized egg to 90 day-old juveniles.

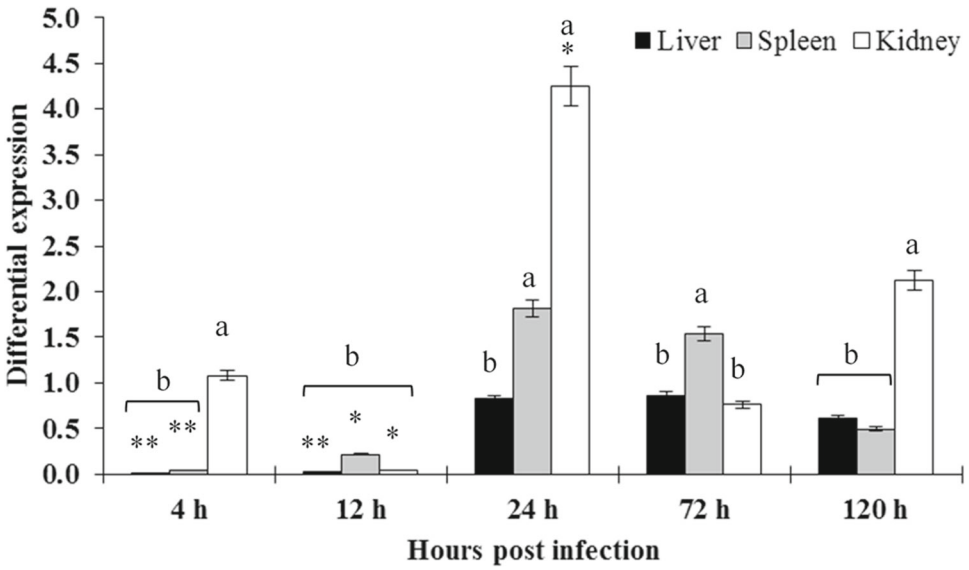

Figure 6. Expression profile of *madiablo* mRNA in liver, spleen and kidney of blunt snout bream at 4, 12, 24, 72 and 120 h postinfection with *A. hydrophila*. Expression in both control and experimental group was normalized to the *18S* rRNA as a reference gene. The control group expression level is designated as 1, so (a) values > 1 indicate upregulation, whereas (b) values < 1 indicate downregulation. Each histogram represents the mean ± SE of three replicates. Statistically significant differences from the control group are marked as *($P < 0.05$) and **($P < 0.01$).

somewhat different functions and that further research is required to elucidate their functions in the blunt snout bream immune system. Although, the association of *diablo* with pathogen infection has not yet reported in fish, a study demonstrated a significant correlation between Smac/DIABLO protein expression levels and stressors in humans (Zhang *et al.* 2013). The study result revealed a similar downregulation pattern of Smac/DIABLO was observed in the mitochondria of umbilical vein endothelial cells 12 and 24 h after fluid shear stress stimulation (Zhang *et al.* 2013). A previous study by Mizutani *et al.* (2005) has reported that the regulation of Smac/DIABLO in human renal cell carcinoma (RCC) kidney could be associated with the long-term disease-specific survival of patients, as longer survival was associated with positive expression, and vice versa. The study also strongly indicated that reduced expression of Smac/DIABLO in RCC is associated with drug/immune resistance. Downregulation of Smac/DIABLO was also proven as significant in hypertrophic scars in human skin tissues, especially, through regulation of fibroblast apoptosis and proliferation (Liu *et al.* 2013). Since Smac/DIABLO is believed to have a

crucial role in the facilitation of apoptosis (Vaux and Silke 2003), a possible explanation for the observed downregulation of *madiablo* in response to *A. hydrophila* infection could be to prevent the cells from entering apoptosis. However, this hypothesis should be verified by further specific studies. As Smac/DIABLO binds XIAP, leading to induced apoptosis (in mammals), further studies should focus on identifying XIAP in the blunt snout genome and investigate the interaction of these two proteins in response to *A. hydrophila* stimulation.

In conclusion, this (first) report of *diablo* gene in blunt snout bream uses *in silico* approach to provide a better understanding of the properties, structure and functions of Madiablo protein, as well the first published homology model of a Diablo tertiary structure in fish. Phylogenetic analysis indicates the existence of two *diablo* copies in teleosts; apart from the Percomorpha group, where *diablo-b* has been lost, but *diablo-a* underwent an independent duplication. *Madiablo* gene was constitutively expressed (albeit at rather low levels) during the embryogenesis and early juvenile development, with a spike in expression during the hatching. Expression was much higher in later juvenile development stages, while in adult fish, it was by far the highest in liver. Expression of this gene in immunity-related organs in blunt snout bream showed regulation in response to *A. hydrophila* infection. These results suggest that *madiablo* gene may play an important role in blunt snout bream development and immunity. However, further studies are necessary to understand completely the apoptotic functions of this gene during different development stages as well as its immunity-related functions in the interaction between the host and the pathogens in the blunt snout bream.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable suggestions. The first author, Ngoc Tuan Tran would like to thank the China Scholarship Council for providing scholarship of doctoral programme in Huazhong Agricultural University, Wuhan, Hubei, China.

References

Alnemri E. S., Livingston D. J., Nicholson D. W., Salvesen G., Thornberry N. A., Wong W. W. and Yuan J. 1996 Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171.

Asker N., Kristiansson E., Albertsson E., Larsson D. J. and Förlin L. 2013 Hepatic transcriptome profiling indicates differential mRNA expression of apoptosis and immune related genes in eelpout (*Zoarces viviparus*) caught at Göteborg harbor, Sweden. *Aquat. Toxicol.* **130**, 58–67.

Chai J., Du C., Wu J.-W., Kyin S., Wang X. and Shi Y. 2000 Structural and biochemical basis of apoptotic activation by Smac/DIABLO. *Nature* **406**, 855–862.

Cole L. and Ross L. 2001 Apoptosis in the developing zebrafish embryo. *Dev. Biol.* **240**, 123–142.

Colovos C. and Yeates T. O. 1993 Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511.

Cristobal S., Zemla A., Fischer D., Rychlewski L. and Elofsson A. 2001 A study of quality measures for protein threading models. *BMC Bioinformatics* **2**, 5.

de Almagro M. and Vucic D. 2012 The inhibitor of apoptosis (IAP) proteins are critical regulators of signaling pathways and targets for anti-cancer therapy. *Exp. Oncol.* **34**, 200–211.

dos Santos N. M. S., do Vale A., Reis M. I. R. and Silva M. T. 2008 Fish and apoptosis: Molecules and pathways. *Curr. Pharm. Des.* **14**, 148–169.

Du C., Fang M., Li Y., Li L. and Wang X. 2000 Smac, a mitochondrial protein that promotes Cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42.

Ekert P. G. and Vaux D. L. 1997 Apoptosis and the immune system. **53**, 591–603.

FAO 2013 *FAO Statistical yearbook 2013: world food and agriculture*. FAO, Rome, Italy.

Fiser A. 2010 Template-based protein structure modeling. In *Computational biology* (ed. D. Fenyö), pp. 73–94. Springer, New York, USA.

Geourjon C. and Deleage G. 1995 SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11**, 681–684.

Guruprasad K., Reddy B. B. and Pandit M. W. 1990 Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng.* **4**, 155–161.

He L.-J., Liao L.-K., Yuan J.-F., Tang H.-Y., Wu Q. and Zhang G.-W. 2006 Pathological observation of bacterial septicemia in *Megalobrania amblycephala*. *J. Sichuan Agric. Univ.* **3**, 35.

Hirokawa T., Boon-Chieng S. and Mitaku S. 1998 SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378–379.

Hoang T. X., Seno F., Banavar J. R., Cieplak M. and Maritan A. 2003 Assembly of protein tertiary structures from secondary structures using optimized potentials. *Proteins* **52**, 155–165.

Hoegg S., Brinkmann H., Taylor J. S. and Meyer A. 2004 Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* **59**, 190–203.

Hogg P. J. 2003 Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* **28**, 210–214.

Honda Y., Tanikawa H., Fukuda J., Kawamura K., Sato N., Sato T. *et al.* 2004 Expression of Smac/DIABLO in mouse preimplantation embryos and its correlation to apoptosis and fragmentation. *Mol. Hum. Reprod.* **11**, 183–188.

Ikai A. 1980 Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898.

Ji W., Zhang G.-R., Ran W., Gardner J. P., Wei K.-J., Wang W.-M. and Zou G.-W. 2014 Genetic diversity of and differentiation among five populations of blunt snout bream (*Megalobrama amblycephala*) revealed by SRAP markers: implications for conservation and management. *PLoS One* **9**, e108967.

Justo P., Sanz A., Lorz C., Gómez-Garre D., Mezzano S., Egido J. and Ortiz A. 2003 Expression of Smac/Diablo in tubular epithelial cells and during acute renal failure. *Kidney Int.* **64**, S52–S56.

Kurowski M. A. and Bujnicki J. M. 2003 GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.* **31**, 3305–3307.

Li J. and Zhang H. 1993 Observation on the embryonic and larval development of blunt snout bream (*Megalobrama arnblycephala* Yih). *Fish. Sci. Technol. Inform.* **4**, 002.

Liithy R., Bowie J. and Eisenberg D. 1992 Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85.

Liu B.-H., Chen L., Li S.-R., Wang Z.-X. and Cheng W.-G. 2013 Smac/DIABLO regulates the apoptosis of hypertrophic scar fibroblasts. *Int. J. Mol. Med.* **32**, 615–622.

Livak K. J. and Schmittgen T. D. 2001 Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.

Luo W., Zhang J., Wen J.-F., Liu H., Wang W.-M. and Gao Z.-X. 2014 Molecular cloning and expression analysis of major histocompatibility complex class I, IIA and IIB genes of blunt snout bream (*Megalobrama amblycephala*). *Dev. Comp. Immunol.* **42**, 169–173.

Mizutani Y., Nakanishi H., Yamamoto K., Li Y. N., Matsubara H., Mikami K. *et al*. 2005 Downregulation of Smac/DIABLO expression in renal cell carcinoma and its prognostic significance. *J. Clin. Oncol.* **23**, 448–454.

Montesanti A., Deignan K. and Hensey C. 2007 Cloning and characterization of *Xenopus laevis* Smac/DIABLO. *Gene* **392**, 187–195.

Nielsen M. E., Hoi L., Schmidt A., Qian D., Shimada T., Shen J. and Larsen J. 2001 Is *Aeromonas hydrophila* the dominant motile *Aeromonas* species that causes disease outbreaks in aquaculture production in the Zhejiang Province of China? *Dis. Aquat. Org.* **46**, 23–29.

Okada H., Suh W.-K., Jin J., Woo M., Du C., Elia A. *et al.* 2002 Generation and characterization of Smac/DIABLO-deficient mice. *Mol. Cell. Biol.* **22**, 3509–3517.

Opferman J. T. 2007 Apoptosis in the development of the immune system. *Cell Death Differ.* **15**, 234–242.

Ramachandran G., Ramakrishnan C. and Sasisekharan V. 1963 Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.

Robinson-Rechavi M., Marchand O., Escriva H. and Laudet V. 2001 An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. *Curr. Biol.* **11**, R458–R459.

Romano N., Ceccarelli G., Caprera C., Caccia E., Baldassini M. R. and Marino G. 2013 Apoptosis in thymus of teleost fish. *Fish Shellfish Immunol.* **35**, 589–594.

Ronquist F. and Huelsenbeck J. P. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Roy A., Yang J. and Zhang Y. 2012 COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.* **40**, 471–477.

Shiozaki E. N. and Shi Y. 2004 Caspases, IAPs and Smac/DIABLO: Mechanisms from structural biology. *Trends Biochem. Sci.* **29**, 486–494.

Steller H. 1995 Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449.

Sun Y., Liu Z. and Zhang S. 2010 Tissue distribution, developmental expression and up-regulation of p8 transcripts on stress in zebrafish. *Fish Shellfish Immunol.* **28**, 549–554.

Takle H. and Andersen Ø. 2007 Caspases and apoptosis in fish. *J. Fish Biol.* **71**, 326–349.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Tran N. T., Gao Z.-X., Milton J., Lin L., Zhou Y. and Wang W.-M. 2015a Pathogenicity of *Aeromonas hydrophila* to blunt snout bream *Megalobrama amblycephala*. *IJSRP* **5**, 1–7.

Tran N. T., Gao Z.-X., Zhao H.-H., Yi S.-K., Chen B.-X., Zhao Y.-H. *et al.* 2015b Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish Shellfish Immunol.* **45**, 72–82.

Vaux D. L. and Silke J. 2003 Mammalian mitochondrial IAP binding proteins. *Biochem. Biophys. Res. Commun.* **304**, 499–504.

Verhagen A. M. and Vaux D. L. 2002 Cell death regulation by the mammalian IAP antagonist Diablo/Smac. *Apoptosis* **7**, 163–166.

Verhagen A. M., Ekert P. G., Pakusch M., Silke J., Connolly L. M., Reid G. E. *et al.* 2000 Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53.

Wiederstein M. and Sippl M. J. 2007 ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–W410.

Xu Y., Tao X., Shen B., Horng T., Medzhitov R., Manley J. L. and Tong L. 2000 Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* **408**, 111–115.

Yamashita M. 2003 Apoptosis in zebrafish development. *Comp. Biochem. Physiol. Biochem. Mol. Biol.* **136**, 731–742.

Zacchino V., Minghetti M., Centoducati G. and Leaver M. J. 2012 Diablo/SMAC: a novel biomarker of pollutant exposure in European flounder (*Platichthys flesus*). *Ecotoxicol. Environ. Saf.* **79**, 176–183.

Zhang F., Zhang L., Sun L.-L., Meng X.-L., Zhao Y. and Jin X. 2013 Effects of fluid shear stress on expression of Smac/DIABLO in human umbilical vein endothelial cells. *Curr. Ther. Res. Clin. Exp.* **74**, 36–40.

Zhou Z., Ren Z., Zeng H. and Yao B. 2008 Apparent digestibility of various feedstuffs for bluntnose black bream *Megalobrama amblycephala* Yih. *Aquac. Nutr.* **14**, 153–165.

Corresponding editor: Indrajit Nanda