Serine/Threonine Protein Phosphatase 1- α (STPP1- α) from Black Tiger Shrimp, *Penaeus monodon*, an Immune-Related Gene

Jacqueline Marjorie R. Pereda^{1,2,*}, Neil Andrew D. Bascos³, and Mudjekeewis D. Santos¹

¹Genetic Fingerprinting Laboratory, National Fisheries Research and Development Institute, 101 Mother Ignacia Ave., South Triangle, Quezon City, 1103, Philippines

²Far Eastern University, Nicanor Reyes Street, Sampaloc, Manila, 1015, Philippines

³National Institute of Molecular Biology and Biotechnology, University of the Philippines Diliman, Quezon City, 1101, Philippines

*Author for correspondence; E-mail: jacquerpereda@gmail.com

Received: June 14, 2021/ Revised: October 02, 2021/ Accepted: November 17, 2021

Reversible protein phosphorylation is a significant regulatory mechanism in many cellular functions, such as the dephosphorylation of Serine/Threonine protein residues catalyzed by protein phosphatase. In this study, the full-length STPP1- α gene from *Penaeus monodon* was cloned, characterized, and analyzed for its constitutive expression in WSSV-negative P. monodon organs. The gene was originally an isotig isolated from the gills of *P. monodon* that survived WSSV infection. *PmSTPP1-* α gene (GenBank: KX385833) has a total of 2,171 bp, with a 990 open reading frame (ORF) that encodes 329 amino acids (aa), sharing a 93% sequence identity with human Serine/Threonine PP1- α catalytic subunit. The protein has a single conserved catalytic domain and shares almost all the conserved sites and functional residues of the human protein phosphatase, which particularly might have putative functions to viral protein synthesis. Using clustering analysis, *PmSTPP1-* α was verified to be the $-\alpha$ isoform while the reported *L. vannamei* STPP1 is the β form. In silico homology modelling predicts similar structures for PmSTPP1- α and STPP1 from H. sapiens. Conserved functional domains for metal binding, target protein interaction, and toxin binding sites were identical in sequence and predicted structure. The observed variations in amino acid sequence were outside these conserved domains but should be further studied to determine their potential effects on function. Current molecular docking predictions for $PmSTPP1-\alpha$ against three proteins from known P. monodon pathogens suggest specific functional interactions with the target protein binding domain and the molecular toxin interaction sites. *PmSTPP1-\alpha* was found to be ubiquitously and highly expressed in organs of WSSVnegative P. monodon, further investigations on the interactions of this protein will help validate its predicted involvement in the P. monodon immune response.

Keywords: molecular cloning, immunity, disease, white spot syndrome virus, Crustaceans

Abbreviations: WSSV – white spot syndrome virus, *Pm STPP1-* α – *Penaeus monodon* Serine/Threonine Protein Phosphatase 1- α , PP1 – Protein Phosphatase 1, STPP1- α – Serine/Threonine Protein Phosphatase 1- α

INTRODUCTION

In the cell, reversible phosphorylation of proteins is an essential process that is triggered by a stimulus from its surface causing various changes in the function and activities of the intracellular proteins. Regulated changes in the state of protein phosphorylation and dephosphorylation are caused by either the protein kinases or phosphatases (Cieśla et al. 2011). Serine/ Threonine Protein Phosphatase 1 (PP1) has been found to be highly involved in a variety of cellular processes such as protein synthesis, apoptosis, meiosis and cell division, cytoskeletal reorganization, regulation of membrane receptors and channels, and metabolism (Bollen et al. 2010). Collectively, PP1 isoforms have been shown to act on a broad substrate, although there are some such as PP1 holoenzyme, catalyzes a specific substrate to affect a definite biological response (Shi 2009). Though, the three subunits of Serine/Threonine Protein Phosphatase 2A (PP2A) have been fully elucidated in black tiger shrimp (Zhao et al. 2016), further studies on shrimp's PP1 mechanism are still lacking.

White spot syndrome virus (WSSV) is one of the most dreaded viruses in the shrimp industry as it causes massive mortality and major damage to many types of shrimp farming operations (Dieu et al. 2004). The limited knowledge regarding the disease's pathogenicity and the shrimp's immunity warrants further understanding in order to find means of combating the disease, and help the shrimp industry. While studies have been conducted to understand WSSV-shrimp pathogen interaction, and identifying and analyzing genes that are related to WSSV (Li and Xiang 2013; Song and Li 2014), these studies have likewise produced unknown, partial gene fragments that are potentially involved in WSSV infection (Maralit et al. 2014). It is important that these unknown, partial fragments be fully elucidated, as these fragments, when completed, could lead to a better understanding of the shrimp immunity in general. The generation of fully annotated novel genes, if confirmed to occur in shrimp tissues, could then be used later for possible functional in vivo experiments.

In the study of Maralit et al. (2014), an unknown partial gene fragment Isotig00463 (i00463) was singled out from a suppression subtractive hybridization (SSH)-next-generation sequencing (NGS) transcriptome database of gene isotigs isolated from the gills of black tiger shrimp, *P. monodon* that survived WSSV infection. Here, this i00463 was fully-cloned and characterized to be the shrimp ortholog of Serine/Threonine Protein Phosphatase 1- α , named here as *PmSTPP1*- α . We provide its complete nucleotide sequence, protein structure, and its constitutive organ expression in WSSV-negative *P. monodon*. This is the first cloning report of the α isoform of STPP1 in shrimp following the report of the β isoform in Pacific white shrimp *Litopenaeus vannamei*.

MATERIALS AND METHODS

Shrimp Samples A live adult P. monodon used for the cloning part of this study, was collected in a local market in Pasay City, Philippines. On the other hand, three (3) live adult P. monodon, procured from a shrimp pond located in barangay Sta. Monica, Hagonoy, Bulacan, Philippines, were used for the organ expression analysis in gills, hepatopancreas, intestine, muscle, lymphoid organ, heart and hemolymph. All of the P. monodon samples tested negative for WSSV by PCR using the primers of Flegel et al. (2006) (136198F: 5' GTACGGCAATACTGGAGGAGGT 3' and 136429R: 5' GGAGATGTGTAAGATGGACAAG 3'). These tissues and organs were dissected out and preserved in screw cap tubes containing RNALater. The collected tissue samples were shipped to the Genetic Fingerprinting Laboratory and stored in a -80°C Ultralow Refrigerator prior to the experiment.

Cloning of Full Length PmSTPP1-a cDNA Gene

An unknown partial fragment Isotig00463 (i00463) was obtained from transcriptome database of Maralit et al. (2014) accessible at NCBI Sequence Read Archive with accession number SRR57708030 (http:// www.ncbi.nlm.nih.gov/sra). RNA from gills of *P. monodon* was extracted using QIAzol® Lysis Reagent (QIAGEN) following the manufacturer's protocol. Total RNA of the samples was purified using RNeasy Mini Cleanup Kit (QIAGEN) following the manufacturer's protocol while its concentration (μ g per μ l) was determined using ImplenTM nanophotometer P-Class.

The total RNA template used in the first strand cDNA synthesis was 0.103 μ g. A total of 10 μ l reaction mix was composed of the following: 0.5X first strand buffer, 2 mM DTT, 0.1 mM dNTP mix, SMARTerOligo IIA (for 5' cDNA synthesis), 5U RNAse inhibitor, 10 U SMARTScribe Reverse Transcriptase, and 1 mM of respective primers for 5' and 3' cDNA synthesis. The reaction mix was placed in a thermal cycler and programmed to 42°C for 90 min and heated to 70°C for 10 min. After which, 20 μ l of tricine-EDTA buffer was added to the samples for dilution.

Rapid Amplification of cDNA Ends (RACE) PCR was employed to complete the full length of i00463. Gene specific primers were designed (Table 1) and primer annealing sites are shown in Fig. S1. Reactions were done utilizing a total of 25 µl reaction mix using Advantage 2 Kit (Clontech) with composition as follows: 1X PCR buffer, 1 mM dNTP mix, 1X polymerase mix, 1X universal primer, and the generated gene specific primers. The reaction mix was amplified in a thermal cycle with conditions as follow: initial denaturation at 94°C for 3 min, 38 cycles of 94°C for 30 sec, 70°C for 30 sec and 72° C for 3 min, then final extension at 72°C for 10 min. The quality of the PCR amplicons were assessed and documented by visualizing on a 1% gel stained with ethidium bromide in SynGene G:BOX. The resulting amplicons were sent to 1st Base (Malaysia) for sequencing.

In Silico Analysis

In silico analysis was performed following Santos et al. (2006 and 2007) with modifications. The sequences, including **Table 1 Primer sequences used to amplify and sequence the**

Table 1. Primer sequences	s used to a	implify and sec	quenc	e the
full-length serine/threonin	ne protein	phosphatase	1–α	gene
from <i>P. monodon</i> using R	ACE PCR.			-
	SEQUENCE	(61 01)		

PRIMER NAME	SEQUENCE (5' \rightarrow 3')
PmSTPP1-α 5RACE-1	CTCTGGACTTGAGGCATAGTCCACG
PmSTPP1-α 3RACE-1	GGTGACATCCACGGACAGTACTACG
PmSTPP1-α 3RACE-2	CCAGACCAGGGCTTATTGTGCGATC
PmSTPP1-α 3RACE-3	GCAAGTCTGAGGTTGGCCTTTGTTC
PmSTPP1-α 3RACE-4	GTCCAGTTGCCGTAGTAGCAGGAGC

the nucleotide and translated amino acids, average molecular weight, isoelectric point, and extinction coefficient were examined and obtained using Geneious 6.1.8 (Biomatters). The identities of nucleotide and amino acid sequence were analyzed through regions of similarity between other biological sequences using BLASTn and BLASTp. The functional and conserved domain of the amino acid sequence was determined using the Conserved Domain Architecture Retrieval Tool (http://www.ncbi.nlm.nih.gov/Structure/ (CDART) lexington/lexington.cgi). Complete multiple alignments were obtained using ClustalW with default parameters. The alignment (.aln) file format produced from the former tool was utilized for the clustering analysis in MEGA 6 (http://www.megasoftware.net/) using the Neighborjoining (NJ) tree based on Poisson model with 500 bootstrap replications and complete deletion of sites (Tamura et al. 2013). The sequences of the members of MPP superfamily (GENBANK: cl13995) were retrieved from NCBI.

Structural Analysis

BLAST Analysis

Structures related to the $PmSTPP1-\alpha$ sequence were determined by submitting the polypeptide sequence for BLAST analysis at the UniProt website (uniprot.org 2021). Related sequences with available 3D-structures were extracted and compared with the target sequence to determine percentage identity, and sites of conservation and variation. Structures of the related sequences were acquired from the Protein Data Bank (www.rcsb.org; Berman et al. 2000) for further analysis. Structures of the *PmSTPP1-\alpha* sequence were predicted using the Magic Fit function of the DeepView Molecular Viewer (Guex et al. 1997). This function fits the submitted sequence unto a reference structure (e.g. *H. sapiens* STPP1-α; PDBID 3e7a (Kelker et al. 2009)). Comparison of the fit and reference sequences was done to observe predicted locations of conservation and variance between the two proteins.

Homology Modelling

Predictions on the potential 3D-structure of the target protein sequence (i.e. $PmSTPP1-\alpha$) were made using the ITASSER Homology Modelling server (https:// zhanggroup.org/I-TASSER/). This system makes predictions on the potential structures and functions of target proteins based on comparisons with related/ homologous proteins with curated structures in the protein data bank. Model structures for the target protein are returned, as well as identities of the top threading templates that served as references for the structure

prediction. Predictions for enzyme class and ligand binding sites are also provided in the results. The returned models for the submitted sequence were acquired and subsequently analyzed.

In Silico Protein Docking Predictions

The predicted involvement of *Pm STPP1-a* with *P. monodon* immune response was assessed based on predicted associations with protein targets from known Shrimp pathogens. Specifically, protein docking experiments were done with three proteins (PirA, PirB, and VP24). Both PirA and PirB are toxins from *V. parahaemolyticus* (Lee et al. 2015). While V224 is a major envelope protein of White Spot Syndrome Virus (Sun et al. 2016). Molecular structures for these target proteins were acquired from the Protein Data Bank (www.rcsb.org: Berman, et al. 2000); PirA and PirB: PDBIDs: 3X0T and 3X0U; Wang et al. 2014; VP24: PDBID 5HLJ; Sun et al. 2016).

Docking experiments were conducted using the ClusPro Protein Docking server (https://cluspro.org/; Kozakov et al. 2017). This service generates predictions of docking interactions between submitted "receptor" and "ligand" protein partners. The predictions are then assessed for relevance based on their prevalence for the multiple replicate predictions. Models are generated based on these prevalent dock clusters and returned to the user. Models of the top 10 docked structures based on electrostatic-favored; (balanced, specific criteria hydrophobic-favored; VDW + electrostatic-favored) are then used to assess the interactions between the submitted protein pairs. As interactions between proteins are often based on combined interaction types (i.e electrostatic, hydrophobic, and VDW interactions) the top ranked models based on "balanced" factors were used as bases for comparing the different partner proteins. However, additional information on the potential bases that promote specific docking conformations were derived on observations of similar structures in models that were ranked highly when considering the other factors.

To assess the relevance of the predicted docking models for *Pm STPP1-a* and the target proteins, the placement of the docked structures was compared relative to documented functional domains for STPP1- α (e.g. toxin-binding-sites; target-protein binding site; etc.).

Constitutive Expression Analysis

Expression in tissues and organs was determined using the Reverse Transcription–Polymerase Chain Reaction (RT–PCR) method following Santos et al. (2006 and 2007) with some modifications. Total RNA from various organs and tissues of three (3) WSSV-negative P. monodon samples were obtained using QIAzol® Lysis Reagent (QIAGEN) following the manufacturer's protocol. The reverse transcription (RT) reaction mix of each sample was composed of the following: 1X Buffer RT, 0.5 mM dNTPs, 1 µM Oligo-dT primer, 10 units RNase inhibitor, 4 units Omniscript Reverse Transcriptase (QIAGEN), with 0.254 µg of total RNA template and amplified using Gene (GSP) (5'-Specific Primers forward CATCCGAGTCTGTGGTGTAG-3') and reverse (5'-GATCGCACAATAAGCCCTGG-3'). The Elongation Factor 1- α (EF1- α) was amplified using the following primers: EF1- α forward primer (5'-ATGGTTGTCAACTTTGCCCC-3') and EF1- α reverse primer (5'-TTGACCTCCTTGATCACACC-3') (Dang et al. 2010). The cycling parameters used for the amplification were: initial denaturation at 95°C for 5 min, 30 cycles of denaturation at 95°C for 30 sec, annealing at 60°C for 30 sec, 72°C for 30 sec, and final extension at 72°C for 10 min. The negative control which contained no template was also included. PCR amplicons were assessed and visualized on a 1% gel stained with ethidium bromide and documented with SynGene G:BOX.

RESULTS AND DISCUSSION

PmSTPP1-α cDNA and Protein

BLAST and in silico analysis showed that the unknown partial gene fragment as Serine/Threonine Protein Phosphatase 1 Catalytic Subunit – α isoform, named here as PmSTPP1-a (GenBank: KX385833) (Fig. 1). PmSTPP1-a cDNA sequence has a total of 2,171 base pairs (bp) with an open reading frame (ORF) of 990 bp encoding for a 329 aa protein with a molecular weight of 37.6 kDA, an isoelectric point of 6.75, and an extinction coefficient of 36, 620. The upstream- and downstream untranslated region (UTR/DTR) consisted of 139 base pairs and 1,042 base pairs, respectively. The promoter sequence, such as the downstream promoter element (AGACA), was in the location of +27 to +31 nucleotides downstream from the initial codon (ATG). The Kozak's sequence (ATCATGG) is also present in the generated full-length gene sequence. Whereas, two (2) of the poly-adenylation signals (AATAAA) have occurred near the C-terminus noncoding region of the gene, which ended with poly-A tail.

PmSTPP1-a has 93% identity with Serine/Threonine Protein Phosphatase 1 Catalytic Subunit – α of humans (*Homo sapiens*) (Accession no.: NP_002699.1) and 92% identity with that of African clawed frog (*Xenopus laevis*) (Accession no.: NP_001080222.1) (Fig. S2). A recent study identified a novel protein phosphatase (PPs) derived from the Pacific white shrimp (Litopenaeus vannamei) cell (Lu and Kwang 2004). It was reported to only consist of 199 aa, contains almost all the functional catalytic domains of human protein phosphatase except the C-terminal non catalytic sequence. Protein Phosphatase 1 gene was also found in other invertebrates such as in parasitic protozoan (Trypanosoma brucei brucei) (Accession no.: AAZ10947.1), and fruit fly (Drosophila melanogaster) (Accession no.: NP_001262919.1); however, these two sequences have low identity match with the *PmSTPP1-* α . Interestingly, BLAST search showed that *PmSTPP1-a* is most closely identical to a hypothetical protein from water flea (Daphnia pulex) (Accession no.: EFX67868.1) with 98% identity. Moreover, PmSTPP1-a has 99.77% identity with 100% query cover matched with a recent NCBI sequence record of a predicted sequence of Penaeus monodon Serine/Threonine-Protein Phosphatase alpha-2 isoform (Accession no.: XM_037948540.1), which was derived from a genomic sequence of Penaeus monodon isolate on its chromosome 41 with a whole genome shotgun sequence (Accession no.: NC_051426.1) (The NCBI sequences of the latter two (2) mentioned accessions are not included in the clustering tree).

NJ analysis showed $Pm STPP1-\alpha$ clustered with STPP1 clade (Fig. 2). Moreover, a 100% bootstrap value separated the 2 subclades, the STPP1- α /- γ (where the $PmSTPP1-\alpha$ clustered) and the STPP1- β (where the previously reported *L. vannamei* PPs grouped). Within the STPP1- α /- γ subclade, $PmSTPP1-\alpha$ clustered with STPP1- α , hence we concluded that $PmSTPP1-\alpha$ was indeed of the $-\alpha$ isoform while the *L. vannamei* is the STPP1- β termed here as $LvSTPP1-\beta$ (Lu and Kwang 2004).

CDART shown *PmSTPP1-\alpha* to be under superfamily Metallophosphatase (MPP) (GenBank: ofcl13995). The said gene possesses a single catalytic domain homologous with other STPP subfamily that starts with Leucine (L7) and ended with Alanine (A299) (Fig. 1 and Fig. S2). And, based on the result of the tertiary structure (3D) prediction (Phyre² software), PmSTPP1- α was predicted as not conserved from Lysine (K301) up until Lysine (K319). This structure is very similar to that of mammalian PP1s. In mammals, all PP1 catalytic domains (PP1Cs) are affected by alternative splicing, which causes its non-conserved N and C termini to be cleaved. The Nterminus of PP1C α and PP1C γ are nearly similar but the C-terminus of all PP1C isoforms is divergent (Korrodi-Gregório et al. 2014). On the C-terminal end of PP1C α structures, about 25 - 30 aa shown a simple coil (Terrak et al. 2004). This structure is thought to form a complex with some regulatory subunits to mediate isoform specificity (Ma et al. 2015); and also functions as an epitope for

$PmSTPP1-\alpha$

A CAT GOGOGO GO C CAT CIT CA AT AGT C TA T GOGO T T G G AGG T G T G G C AGT G T A T C G G C T A T C AGG G T C T C T A T T A T A GAC CAT C C G AG T C AT C G T G G C AG T T A T T A G A A T C T T C T A T A A T A G A C T C T C T A T A A T A G A C T C T C C AG A G C A G C A G A T A T T A G A A T G C G C AG A T A T T A G A A T G C C G A A T A T A G A A T C G T C A A A A T C G T G C A G A A T C G C C A G A A T C C T C C C A A A A T C G T C C C C A A A T C C C C C A A A T C C C C	_																								_	
CTTATCTGCCCTCATTGAACCATCGGAGTCTGTGGTGTATGGAAAGGATCAGGATAACCTTCTCATCC 13 ATO GCGGAGAGCG GATAAGCTTATATATATAGACAGC ATC GTC GC AGATTATTGAGAGTTCAGGGATAACCTTCTCCCACGA (21) ATA A E T D K L N I D S I I A R L L E V R G S R P G AGG AAT GG CA GG ATTGAAATTGGT GGACTC TGC CTCAG TTAGAGAGATCTCTCTCCACACCA (24) K N V Q L T E N E I R G L C L K S R E I F L S Q P ATC CTTCGGAATTGGAAGGT CT CTCAAAATT TGTGGT GGAC TCCAC GGACAG TACTAC GAT TGCTCCATCA (24) K N V Q L I E N E I R G L C L K S R E I F L S Q P ATC CTTCGGAATTGGAAGGC CT CTCAAAATT TGTGGT GGAC ATCCAC GGACAG TACTAC GAT TGCTCCATCA (24) F E Y G G F P F E S N Y L F L G D Y V D R G K Q S TTG GAAACCATATGTCCCC CCAGAGTCCAAC TAC TG TG TTTTAGGA GATTAC CTCGAT CGT GGTAAACAC (24) F E Y G G F P F E S N Y L F L G D Y V D R G K Q S TTG GAAACCATATGTCCCCC TGGCC TACAAGATCAT TC CAGAGAAGTTC TCCTCT CT AGAGGC AACCAT 51 L E T C L L A Y K I K Y P E N F F L L R G N H S I N R I Y G F Y D E C K R Y N I K L W K ACT TTCAAAGACTGT TG TCTAC GGAC GGG GG GG GG GG TATAAC ATC AAACTG TG GAAG F F N C L F V A A I V D E K I F C C H G G TTG GACCCGAC CTG CAGAGAATCTAT GGGAG CACTAT GG GGAGAAATGTCTC TGCT GCCAGAGC GG G TTG TGACAGCCTG TG TTCCAAT GGAACAGATC CGT CGC GC GG GG GG GG GG AGAAAATGT CCAGGG CTAC (36) F T T D C F N C L F V A A I V D E K I F C C H G C T F T D C F N C L F V A A I V D E K I F C C C A G TTG GACCCCGAC CTC CAGAGCATGGAACGAACGGAC CTT GG GG GGA GAAATGGAAGG GC CACGACCAG GG CTT A G A E V V A K F L H K H D F D L I C R A H Q V 250 CTT GAAGATGTTC CAGATGT TCTTGCC AGCGA CACTA GG GG CTG TGCGCCT CTCGTG CGAGGCTGAT CAG GTT TTG GAC AATGTAGT TCCAGATGT TTTGCC AGGGAC ACTA GG GG CGC GC GGAGAAATGGA AGGAC ACCAC GC AGT TG TG GAAAGTGAAGT TAAAACTAGGGGAT GGACCTT TAGGT GG GCCT TCTGTC CAGGT CAC GCC GAC GC TT G GAAGAGGGAATTAAAAACAGGGAT GGAACCCT AGG GG GACCC CTT CAGGT GG GCCT TCAGG GG GAA A C AAGAAGAAATGAA TAA AAACTAGT GT TGGAC TAGAGGGA CACCT TGGT GG GGAGAATGAAATGA				ACA	TGG	GGGG	GCGC	CCAT	CTT	GAA	TAG	TCT	ATG	GGG	TGG	GCI	GTT	GTG	AGG	TTG	TGG	CAGI	GTA	TCG	GCT	69
ATO GOG GAGAGO GATA ARGOTT ANA CONCENT CATC GOC GAGATTATTAGAA GATT COG GOG TO COC COG AGOL 21 M A B T D N I J A R L E V R G S R P G 21 M A R V Q L T E N R C C L R R P G S R P G C L R S R P G C C L L R R C G D I C R D X L R C G D Y V L R C G D Y V L R C G D Y V L R Y L L R X I Y D R I R G R Y N L L R X I R C R I <td< td=""><td></td><td></td><td>C</td><td>TTA</td><td>TCT</td><td>GCCI</td><td>TCTO</td><td>CATI</td><td>GAA</td><td>CCA</td><td>TCC</td><td>GAG</td><td>TCT</td><td>GTG</td><td>GTG</td><td>TAGI</td><td>GCA</td><td>AAG</td><td>TGA</td><td>TCA</td><td>GGA</td><td>TAAC</td><td>CTI</td><td>CTC</td><td>ATC</td><td>139</td></td<>			C	TTA	TCT	GCCI	TCTO	CATI	GAA	CCA	TCC	GAG	TCT	GTG	GTG	TAGI	GCA	AAG	TGA	TCA	GGA	TAAC	CTI	CTC	ATC	139
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	ATG	GCG	GAG	ACG	GAT	AAG	CTT	AAT	ATA	GAC	AGC	ATC	ATC	GCC	AGA	TTA	TTA	GAA	GTT	CGA	GGG	TCT	CGC	CCA	GGA	214
ARE ARE FEECABETRACCGRARTGRARTEGG GENCETEC CECCEGAG GEAGATC CEC CECTCE CECCEGAG GEAGATC CECTCE CECTC CECCEGAGATEG GACATC CECEGAG GEAGATC CECTA CAGATTEG CECTC CECEGAGATEG GACATCCAC GACTEG CECTATEG CECTEG GEACATCCAC GACTEG CECTATEG GEACATCCAC GACTEG CECTATEG GEACATCCAC GACTEG CECTATEG GEACATCCAC GACTEG CECTAGAG GEAC GEACATCG GEAC GEAC GEAC GEAC GEAC GEAC GEAC GE	м	A	E	т	D	к	L	N	I	D	s	I	I	A	R	L	L	E	v	R	G	s	R	P	G	25
K N V Q L T R G L C L K S R P L S Q P J ACC CTC C C C L L R L L R L K I C G D I H G Q Y D L L R L K I C J T T T G G F P P S N Y L F L G D Y V D R G N H 1 Z J T T T C L X Y K Y Y L X Y L Y L Y L X Y L X K Y L L L X Y L X X Y L L L X X L X Y </td <td>AAG</td> <td>AAT</td> <td>GTG</td> <td>CAG</td> <td>TTA</td> <td>ACC</td> <td>GAG</td> <td>AAT</td> <td>GAA</td> <td>ATT</td> <td>CGT</td> <td>GGA</td> <td>CTC</td> <td>TGC</td> <td>CTC</td> <td>AAG</td> <td>TCC</td> <td>AGA</td> <td>GAA</td> <td>ATC</td> <td>TTC</td> <td>CTC</td> <td>TCA</td> <td>CAA</td> <td>CCA</td> <td>289</td>	AAG	AAT	GTG	CAG	TTA	ACC	GAG	AAT	GAA	ATT	CGT	GGA	CTC	TGC	CTC	AAG	TCC	AGA	GAA	ATC	TTC	CTC	TCA	CAA	CCA	289
ATC CITC TG GAA TT GAA GT CT CT CAAAATT TG GG TG AC AT CAC GG ACA TA CCATA CGATTAC GAT TT GT TG ACAT CG GA GT CT CC CCAGA GT CCA CT AC TT TT TAGA GAGATTAC GT GG AT TG GT GAAACA TT CC CC CAGAG TC CC CCAGA GT CCA CT AC TT TT TAGA GAGATTAC GT GG AT GG GG TAAACAG TC CC 43 F E Y G G F P P E S N Y L F L G D Y V D R G K Q S 100 TT GGAAA TC ACTAT TG TC TC TT GG GC TC CAAGATAC AT TAT CCAGAA AATT TT CT CT TC AGAGG CAA C CAT 1 L E T I C L L L Y Y K I F L G D Y V D R G K Q S 100 TT GGAAA CCATTAT TG TC TC TT GG GT TC TAC GAC GGAA GAT TAC CT TC TC TC AGAGG CAA C CAT 1 L E T I C L L L Y Y K Y P E N F F L L R G N H 12 CAO TG TG CATCC ATC AACAAACT TAT GG TT TC TAC GAC GGAA GAAAAGAT CT TC TC TC AGAGG CAA C CAT 1 C C A S I N R I Y G F Y D E C K R R Y N I K L W K 150 ACT TT C ACAGAC TGT TC CAAT TG CT TC TC TC GC GG CC ATT AT GAGG CCT AC GGA GGT GAAAAGAT CT TC TG C GC CC CAG C CG GG GGA AAAGAAT CT TC TG TG C C CAGAC AG C CAGA C C GAA GC C AA T C AG AG C T T T G AG C C C GAG C C AG C C AG AC C AG C C A	ĸ	N	v	Q	L	T	E	N	E	I	R	G	L	С	L	ĸ	S	R	E	I	F	L	S	Q	P	50
TT GALAGGE GGETTECCCC CAGAGET CACALCTACTEST TT TAGGEGAT TAGGET TO TEGT GALAGEGET GGET ALACGE TCC TT GALACGE GGETTECCCC CAGAGET CACAGET ACTEST TT TAGGEGAT TAGT GGET ALACGET GGET ALACGE TCC TT GALACCCATA TGT CT CCT CT GGCC TACALGAT CALT TT TT TAGGEGAT TAGT GGET CT TC CAGAGEC ALC CAT L E T I C L L L A Y K I K Y P E N F F L L R G N H 122 GAS TGT GCATCCATC ALCAGAAT CT AL GT TT TT CAC GAC GALAGEGEGEG TATALAC AT CALCAGEGE CAL CALT TT CACAGE TT TT CALT TGCTTACCT GT CGC GGC GC ATT GT GGAT GALAGEGEG TATALAC AT CALCAGE GGA GG T F T D C F N C L P V A A I V D E K I F C C H G G 175 TTG AGCCCCGAC CT CAGAGEAT GGAT GALAGEGEC CAC GGA GGA GGA GALAGET TT TG GGC CT CC AGGAGE AT GGAT GALAGEGEG CT AC GALAGEGEC TA ACT TTT GACAGEC CT CAGAGEAT GGAT CAGAT CGAC GALAGEGE CT AC GALAGEGEC TA 1 S P D L Q S M E Q I R R I M R P T D V P D Q G L 201 TTG TGC GALCTT TTG TGGT CAGAT CCAGACAAGE GGC TGG GGAGAALAGAT GALAGEGT GT AT CT TC 1 C D L W S D P D K D T M G W G E N D R G V S F 221 ACT TTT GGGC CAGAA GT AGT TGCALAGEGA CALAGE GGC TGC GGAGAALAGAT AGA GGGT GT AC TT 1 G AG GATGGT TAC GAAT CCA GALAGEGA CALAGE GGC TGC CT CACT CT GT CGA GCT CAC GG T 1 F G A E V V A K F L H K H D F D L I C R A H Q V 2 GT GALAGEGE GGA GT AGT TG TT GGAL GGAT TG AGT GGA GGA GGA TAC TT TG GG GAAL GALAT GALAT TA TT TG GALAGT TT TT GGA GAGT TT TT AGGC CALCAGEGA CT AL TA TT GALAGE CCC AGC CALT CAT GG GG GALAGE CCC AGC CALT CAT GGA GALACT ACT GGG GGALAGEGA CCALT ACT TT TT CCC AGACE ACT TA TG AGCCA CALCAGEG CGA CCC CAGACE ACT TA TG AGA AGT AC ACAGEGE CAAT CCA AGC GGA CCA CT AT GG AGG GGA TT GALAGEAGEGA TT GALACT AGT GGALACT AGT GGALACCAAGEG GGA CCC CT ALCG CAAGE CCA CAC ACT ACT TG GG GALA 1 T T GALAAAGAAT AA ALACTAGT GT TG AGCA AGGG GGA CCC AT TA TG GALACT TA AGCCAAGE AGA CCAACGA GGA TT GALAGEAGET TT TT TG GALAGAAGAAT AA ALAACT AGT GT TG TG TT TT TT TT TT TT TT TT TT T	ATC	CTT	CTG	GAA	TTG	GAA	GCT	CT	CTC	AAA	ATT	TGT	GGT	GAC	ATC	CAC	GGA	CAG	TAC	TAC	GAT	TTG	CTT	CGA	TCC	364
TT GAALACGAC GG F P P E S N Y L P L G D Y V D R G K Q S TG GAALCATA TG CCCCT CTGGCCTACAAGATC AAA TAT CAGAAATTT CTGCTT CT AGAGGCAAC CAT I E T I C L L L A Y K I K Y P E N F F L L R G N H 121 GAG TGT GCATCA TCA ACAGAATTAT GG TT CTAC GAC GAA TG CAAG CGCG GG TATAACATC AAACTG TG GAAG E C A S I N R I Y G F Y D E C K R R Y N I K L W K ACT TTCACAGAC TGT TCAAT GG TT GT AC GAC GG GG GC AT TG GAAG GG GG GG TATAACATC AAACTG TG GAAG E C A S I N R I Y G F Y D E C K R R Y N I K L W K ACT TTCACAGAC TGT TCAAT TG CTTA CTG GG CG GG GG CCAT GT GG GAG GAAAGAG TCT CTG TG CCAC GG GG GG T F T D C F N C L P V A A I V D E K I F C C H G T TTG AGC CCCGAC CTC CAGAGC AT GG GAACAGATC CG T CG GG CCA TT AGAGG CCTAC CG ACGG GG CCA ACT TTT GG GG CCCAGAG AT GG AA CAGATC CG C CCAT TA TG AGG CCTAC CG ACGT GG GG GAA AATGAT AGA GG TG TA TC I S P D L Q S M E Q I R R I M R P T D V P D Q G L CT TT GG GG CAGAG GT AGT CCAGAC AGAG CAC AATG GG CTG GG GG AA AATGAT AGA GG TG TA TC I C D L L W S D F D K D T M G W G E N D R G V S F 221 ACT TTT GG GG CAGAG GT ACTT GT GG CCAAGG GA CCA AGG GG CAA CAT TG TG CCAA CAC TG TG GG GG GA ACT ATT GG GG C GAG TA GT GCAAAG TC CT CCAC AG CA CAT GT GAACTT TA TCCAC GAC CCA AC TA CTG GG GG GA ACT ATT GG GG C GAG TA GT TG CAAGC CAAG GG CAA CTA GT AACTT ATT TCCA GAC CCCA AC TA CTG GG GG GA F D N A G A W M S V D E T L M C S F Q I L K F A D AAA AAG AAG TAC TA GG A T M M S V D E T L M C S F Q I L K F A D AAA AAG AAGTAC TCCT TT GG AG GAT GT GAACAC GG GG CAC CCG TAACC CG GG GG GG GG GG CCA CCCA AC TA TG GG T TT TA TT TA TT GAT TA TT TT	-				-		•	-			-				-			*	-	-				~	-	400
TTG GAA CCATA TGT CTC TT GGC TACAAGATC AAA TAT CCAGAA AATTTC TT CCT CT GAGAGC AAC CAT 1 E T I C L L L A Y K I K Y P E N F F L L R G N H 121 GAG TGT GCATCC ATC AACAGAATCTAT GGTTTC TAC GAC GAA TGCAAG CGGCGG TATAAC ATC AAACTG TGG AAG E C A S I N R I Y G F Y D E C K R R Y N I K L W K 151 ACT TTC ACAGAC TGT TC AAT TGCTTA CTGT GC GGC GAT GT GGAT GAAAAGATCTT CGC TGCCAGAC GAA GGT G G A C C T C CAGAGC AT GGAA CGAAT CG C ATT ATGAGG CT ACC GA CGT G C C A G G 177 TG ACCCCCCAC CT C CAGAGC ATGGAA CGAATC CGT CG C ATT ATGAGG CT ACC G G G G C TA 1 S P D L Q S M E Q I R R I M R P T D V P D Q G L 200 TTG TGC GATCTT TTG TGGT CAGAC CGAAT GGAA CGAAT G GC TGG GGGAGAAAATGAT AGAGG TAT C T TC 1 C D L L W S D P D K D T M G W G E N D R G V S F 221 ACT TTT GGG CC GAAG GT AGT CCAGAC CAGAG GAC ACAATG GGC TG ACTT TG CGACTCAT CGT CGAGCT CT C G G TT TG GGG CG GAA GT AGT GCAAAG T CCT CAG CACAG GGC CACATG GC CG GGG GAA AATGAT AGAGG TG T T C T C C D L L W S D P D K D T M G W G E N D R G V S F 225 ACT TTT GGG CT GGAG TAGT TG CAAAG T GCAAAG GAC ACAATG GGC TA ACT T T G GG GG CT AC C T T G GG GG T T A C A R Y V T T G G A E V V A K F L H K H D F D L I C R A H Q V C G T GAAGATGT TC CAAGG GG CAACTA GT AGAACTT ATTCT CG CG ACCCC AC T ACT GT G GG GA F D N A G A M M S V D E T L M C S F Q I L K F A D 300 AAAA AAG AAG TAC C T ATGGA GGATT GAACACA GGG CGA CCC GT AACG CCG CG GGAGG AG CCCAAT C AG AAA 111. K K K F P Y G G L N T G R P V T F P R G A A N Q K 322 TATTT AGACATATTTT TG TG TATCCAGG GGAT TAACAG GGG GGA CCC GT AACG CGG GGA GCC CAAT C AG AAA 114. K K K F P Y G G L N T G R P V T F P R G A A N Q K 322 TATTTAT GAATTATCAGT GT TG GT GT GT GT GT CT T T T T T T T	F	GAA	Y	GGC	GGC	F	P	P	E	S S	AAC	Y	L	F	L	GGA	D	TAC	V	D	R	GGT	K	O	S	439
THE GRANT CARATERISTIC FOR THE GOLD THE ANGAE CARATER CARATERIC TREATER AND THE TECHT OF THE THE CHILD THE CHILD THE THE CHILD THE THE CHILD THE C	-	-	-		-	-	-	-	-	-		-		-	-		-	-		-				~	-	514
CASTGET CATC ALCAGA ATCTAT GETTECTAC GAC GAA TGCAAG CGGCGG TATAAC ATC AAACTGE GA AG E C A S I N R I Y G F Y D E C K R R Y N I K L W K ACT TTC ACAGAC TGT TTCAAT TGCTTACCTGTC GCG GCCAT GTGGAT GAAAAG ATCTC TGC TGCCAC GGA GGC T F T D C F N C L P V A A I V D E K I F C C H G G TTG AGC CCCGAC CTC CAGAGC ATGGAA CAGATC GGT CGC ATTATGAGG CCTACC GACGTG CCA GACCAG GGC TTA I S P D L Q S M E Q I R R I M R P T D V F D Q G L TTG TGC GATCTT TTG TGGTCA GAT CAGACAGG GAC ACAATG GGC TGG GAGAAAATGAT AGA GGT GTAT CC TTC ACT TTT GGGGCA GAAGTAGT GCAAAG TCCCC CACAAG GAC ACAATG GGC TGG GAGAAAATGAT AGA GGT GTAT CC TTC ACT TTT GGGGCA GAAGTAGT GCAAAG TCCCC CACAAG GAC AGATG GGC TGG GAGAAAATGAT AGA GGT GTAT CC TTC GTT GAGGCA GAAGTAGT GCAAAG TTCCTC CACAAG CAT GACTTA TGCCCA TCTGT CGA GCTCAT CAG GTT 881 T F G A E V V A K F L H K H D F D L I C R A H Q V 500 GTT GAAGATGGT TAC GAGTC TTTGGCC AAGCGA CAACTA GTA ACATTATTCTCAGCACCC AAC TACTGT GG GAA 966 V E D G Y E F F A K R Q L V T L F S A P N Y C G E 77 TTT GACAATGCAGGT CCTATGGA GAATGAT GAG ACACTTATGTGT TCCTTC CAGATAC TTAAGCCAGCA GAT 1031 F D N A G A M M S V D E T L M C S F Q I L K F A D AAA AAG AAGTTC CCC TATGGTGTGTGCCACCAGGA GGC GCC CTTGTGGCCATT CTGTTATACCT 1190 N K K K F P Y G G L N T G R P V T P P R G A A N Q K 322 AAC AAG AAGAATAA AAACTAGTGTTGTGTGGCCACTGTCTTGGTGTGGCACTTGAGGTTATTTA 126 TTATTTAACATAT TTAACCTGGATTGCTATCCAGGAGATATGAATTATTGCAAATTATTGGACATAGAAATTATTTAAGGGCAAAGT 137 TATTTATCTATTTGAACTTCCTGATTTTATCCTTGAGATATAGCAGGACAATAAATCTGGTTATTTA 136 AAACTATTTAACCTGGATTGCTATCCAGGAGATGTGACTCAAGGTATTATATTATATTAATTA	TTG	GAA	T	TA	C	L	L.	TTG	A	Y Y	AAG K	ATC	K	TAT	P	E	AAT	F	F	L	L.	R	GGC	AAC	H	125
$ \begin{bmatrix} c & c & s & i & n & r & i & y & c & r & y & d & c & k & r & c & k & r & r & n & i & k & l & w & k \\ c & c & t & c & t & n & r & t & k & l & w & k & l & s \\ c & t & t & t & c & k & r & r & v & n & t & k & l & w & k & l & s \\ c & t & t & t & d & c & r & n & c & l & p & v & a & a & i & v & d & e & k & i & p & c & c & r & g & g \\ t & s & p & d & l & Q & s & m & e & Q & i & r & r & i & m & r & p & t & d & v & p & d & q & d & l \\ c & s & p & d & l & Q & s & m & e & Q & i & r & r & i & m & r & p & t & d & v & p & d & q & d & l \\ c & s & p & d & l & Q & s & m & e & Q & i & r & r & i & m & r & p & t & d & v & p & d & q & d & l \\ c & c & d & l & u & w & s & d & p & d & k & d & t & m & g & w & g & e & n & d & r & g & v & s & p & 2 \\ c & t & t & t & t & g & g & g & g & c & 1 & r & r & s & d & r & g & v & s & p & 2 & 2 \\ c & t & t & t & f & g & a & e & v & v & a & k & p & l & h & k & h & d & p & d & l & i & c & r & a & h & q & v & 2 & 5 \\ c & t & t & t & f & g & a & g & v & v & a & k & p & l & h & k & h & d & p & d & l & i & c & r & a & h & q & v & 2 & 5 & 2 & 2 \\ c & c & t & t & s & g & g & t & f & r & h & k & r & l & h & k & h & d & p & d & l & l & c & r & a & h & q & v & 2 & 5 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2$	GRG	- TGT		-	ATC	220	- D G D	ATC	T A T	- 667	ттс	-	GAC	- GDD	TGO	- 200	caa	- GGG	- 	-	-		CTG	TGG	220	580
ACT TICACAGAC TGT TICANT TGCTTA CCTGTC GCG GC ATT GT GGAT GANANG ATCTTC TGC TGCCAC GGA GGA T F T D C F N C L P V A A I V D E K I F C C H G G TTG AGC CCCGAC CTC CAGAGC ATGGAA CAGAT CGT CG CAT ATGAGG CCTACC GACGTG CCA GAC CAG GGC TT A S P D L Q S M E Q I R R I M R P T D V P D Q G L TTG TGC GATCTT TTG TGGTCA GATCCA GACAAG GAC ACAATG GGC TGG GGAGAAAATGAT AGA GGT GTATCCT TC L C D L L W S D P D K D T M G W G E N D R G V S F ACT TTT GGGGCAGAAGTAGTT GCAAAG TTCCTC CAC AAG GAT TA GACGT CAT CAG GTT GT GAAGATGGT TAC GAGTTC TTTGCC CACGAC CAA TG GAC TT GACCTC AT CTGT CGAGCT CAT CAG GTT A F G A E V V A K F L H K H D F D L I C R A H Q V GTT GAAGATGGT TAC GAGTTC TTTGCC AAGCGA CAA CTA GTA ACATTA TTCTC AGCAC CAACTACTGT GGG GAA V E D G Y E F F A K R Q L V T L F S A P N Y C G E TTT GACAATGCA GGT GCAATG AT GTC GTAGAT GAG ACCT AT GTG TTCCT TC CAGATA CTT AAGCCAG GAT F D N A G A M M S V D E T L M C S F Q I L K P A D ANA AAG AAGTTC CCC TAT GGA GGAT TG GAC ACCT GTG GGC GAAG CCC GAAG GAC GCC CAAC CAA	E	C	A	s	I	N	R	I	Y	G	F	Y	D	E	C	K	R	R	Y	N	I	K	L	W	K	150
T F T D C F N C L F V A A I V D E K I F C C H G G 17 TIG AGC CCCGAC CTC CAGAGC ATGGAA CAGAT C GT C G	ACT	TTC		GAC	TGT	TTC	лат	TGC	ττδ	сст	GTC	GCG	GCC	סדד	GTG	GAT	GAA	AAG	ATC	TTC	TGC	TGC	CAC	GGA	GGC	664
TTG A GC CC GAC CT C CAGAGC AT GGA C AGAT C GT CGT CGT AT AT GAGG CCT AC C GAC GTG CCA GAC CAG GGC TTA I S P D L Q S M E Q I R R I M R P T D V P D Q G L 200 TTG TGC GAT CTT T GTG TGC CAG AC CAG GAC CAA GG GG CAG AA AT GAT AGA GG GT AT CC TTC C D L L W S D P D K D T M G W G E N D R G V S F 222 ACT TTT GG GG CA GA GT AGT T G CAAAG TT C CAC AAG CAA T GAC TT T G GG CCA ACT ACT GG GT AT CC TTC T P G A E V V A K F L H K H D F D L I C R A H Q V 250 GTT GAA GAT GGT T C CAATG CT T GAC AT GTA ACT T AT G T G C A C A C T A C T A C G C C A C T A C T A C G C C A C T A C T A C G C C A C T A C T A C G C C A C T A C T A C C A C A C C C A C T A C T A C C C A C A	T	F	T	D	C	F	N	C	L	P	v	A	A	I	v	D	E	K	I	F	C	C	н	G	G	175
L S P D L Q S M E Q I R R I M R P T D V F D Q G L TIG TGG GATCTT TIG TGGTCAGATCCA GACAAGGAC ACAATGGTCGGGGAGAAAATGATAGAGGGTATCC TTC L C D L L W S D P D K D T M G W G E N D R G V S F 222 ACT TIT GGGGCA GAA GTAGTT GCAAAGTCC CACAAGCAC GATTAGACCT ACTGGT GGCACATCAG GTT T F G A E V V A K F L H K H D F D L I C R A H Q V GTT GAA GATGGT TAC GAGTTC TITGCC AAGCGAC CAA CTAGTAACATTATTCTCA GCACCC AAC TACTGT GGG GAR V E D G Y E F F A K R Q L V T L F S A P N Y C G E TIT GACAATGCA GGT GCAATGATGATGAGT GAGACAC TAGTGAT GGT GTTCCTTC CAGATACT TAAGCCAG CAA CTA F D N A G A M M S V D E T L M C S F Q I L K F A D AAA AAGAAGTTC CCCTATGGA GGATTGAACACAGGG CGACCC GTAACGCCGCCGCGGAGGA GCAGCCATACAG AAA K K K F P Y G G L N T G R P V T P P R G A A N Q K AAC AAGAAGAATAA AAACTAGTGTTGATGATGACATGGTGTGGCAAGTTGGAGGTTGGCCTTGTTCATCCT III N K K K * TGCCTTAAGTTTTG GGATTGCTATCAGGGGGATGGACTCAAGAGAAATAAATATGAATTTAACTGGGTTATTTA 12 60 TTTTAACATATTTTAACCTGGATGATGACCAGGGGTAGTCACTTCAAAAAGAAATAAAT	TTG	A GC		GAC	CTC	CAG	a GC	ATC.	699	CAG	a TC	CGT	cac	ስጥጥ	370	100	CCT	200	GAC	GTG		GAC	CAG	GGC	773	730
TTG TGC GATCTT TG TGGTCA GATCCA GACAAG GAC ACA ATG GAC TG GG GGA GAA AATGAT AGA GG GT A TCC TTC L C D L L W S D P D K D T M G W G E N D R G V S F ACT TTT GGGGCA GAA GTAGTT GCAAAG TT CCTC CAC AAG CAT GACTT GACCTC ATCTGT CGA GCTCAT CAG GTT T F G A E V V A K F L H K H D F D L I C R A H Q V GTT GAA GATGGT TAC GAGTTC TTTGCC AAGCGA CAA CTA GTAACATTATTCTCA GCACCC AAC TACTGT GGG GAA V E D G Y E F F A K R Q L V T L F S A P N Y C G E TTT GAC AATGCA GGT GCAATG ATGCTT GTAGAT GAG ACACTT ATGTGT TCCTC CAGATACTT AAGCCA GCA GAA F D N A G A M M S V D E T L M C S F Q I L K P A D AAA AAG AAGTTC CCT ATGGAG GAACACA GGG CGA CCC GTAACG CCGCCG CGAGGA GCA GCCAAT CAG AAA K K F P Y G G L N T G R P V T P P R G A A N Q K 322 TGCCTTAAGTTT GTGGATT GCTAACGTGTGTGGTAG GAACTT ATGTGT TGTGAGTAGTT TAACT GGGTTATTTA 126 TTATAACATATTTATAACATATTATATCAGTGTGTGGAACAAGGGGAACTT CAAAAAGAAGTTGGACCTAGGGTTGGCCTTGTTCATCCT 119 AAA AAG AAGAAA TAA AAACTAGTGTGTGGTGTGCACTGTCTGGTGGCAAGTCTGAGGTTGGCCATGT GAAGTTAATTA ATTTTATGAAATAA AAACTAGTGTTGGTGGTGCACTGTCTTGGTGTTGGCAAGTCTGAGGTTGGCCATGTGAAATT 142 TTATTATGAATTTAACATATTTATATCGTGGTGGAACAAGGGGAACTTAATTGAACTTGGAACATAAATTTAACTGGGTTATTTA 142 TTATTATGAATTTCAGTTGCTAAGGTGTTTGTTGTGTAGCAAGGGTAATTAAGACTTGGAACTAAATTTAACTGGGTTATTTA 142 TTATTTATGAATTTAACATATTTATTTGGTATATCCATGGAAGTAAAGGGTAATTAAGACTTGGAACTTAGGACATAAATCTGTGAAATTTCT 142 TTATTATGAATTTAATATATATTATTTTTTTTTTTTTT	L	S	P	D	L	Q	S	M	E	Q	I	R	R	I	M	R	P	T	D	V	P	D	Q	G	L	200
L C D L L W S D P D K D T M G W G E N D R G V S F ACT TIT GGGGCA GAA GTAGT GCAAAG GT CCC CAC AAG CAT GACT GAC								~ ~ ~		~~~~			1.01		~~~			~ • • •								014
ACT TIT GGGGCA GAA GTAGTT GCAAAG TIT CCT CAC AAG CAT GACTTT GACCTC ATCTGT CGA GCT CAT CAG GTT T F G A E V V A K F L H K H D F D L I C R A H Q V GTT GAA GATGGT TAC GAGTTC TITGCC AAGCGA CAA CTA GTA ACATTATTCT CA GCACCC AAC TACTGT GGG GAA V E D G Y E F F A K R Q L V T L F S A P N Y C G E TTT GAC AATGCA GGT GCAATG ATGCTT GTAGAT GAG ACACTT ATGTGT TCCTC CAGATA CTT AAGCCA GCA GA F D N A G A M M S V D E T L M C S F Q I L K F A D AAA AAG AAGTTC CCC TATGGA GGACACTG GTA GGG CGA CCC GTAACG CCGCCG CGAGGA GCA GCCAAT CAG AAA K K F F Y G G L N T G R P V T P P R G A A N Q K ACC TAAGTATGTTGTGGATTGCTGGTGTGGTGTGGCAAGTCTGGAGGTGGGCCTTGTCATCCT N K K K * TGCCTTAAGTTTGTGGATTGCTAGGTGGACAAGGGGTAATTAGACATGGACACTGGAGGTCGACGTGGCCTTGTCATCTT TAATTAACATATTTATATCTGGGGTGACACAGGG CGA CCC GTAACG CCGCG CGAGGA GCA GCCAAT CAG AAA 111- K K K * T P Y G G L N T G R P V T P P R G A A N Q K 322 TGCCTTAAGTTTGGGATTGCTAGGTGTGGGTGGACTTCAAAAAGGAAATTAAA TATTTATGGAATTACAGTGGTGGATGAACAAGGGGTAATTAGACTTGGCAAGTCTGAGGTTGGCCATGTTATTAT TAATTAACATATTTATATCTGGGATGAACAAGGGTAATTAGACTTGGCAAGTCTGGAGCAAAATTATATTATGGATTTAACTGGGTTATTAT 126 TTATTATGGAATTCCGTGGATGAACAAGGGGTAATTAGACATTGGTAATGGACATTGGACATAAATCTGTTAATGAAATA 134 TTTTTTTTGTGTTTTTTATCAGTTGTGAAAAAGGGTAATTAGACATTGCTAACTGGACTAAAATCTGTTAATGAAATA 134 TTTTTTTTTTGAATTTCAGTTGCTAGATTATCCATGATAATTGTTAATTATTATTATGGACTTTGGTACTAGGCAAAGTC 142 TAATTAACATATTATATATATTCCTGGATTATCCATGATAATTGTTAATTATTATTATTATGGACTTTGGTACAATTGTTCA 142 150 CAAAAGAATTGGCTTGCAAAGAGCTGGAATTAGGATTTATATTATTATTATTATGGAATTAGGAATTAGGCAATTTCT 151 CAAAAAGAATTGGCTGCAAAGGACTCGAGGTGAATTAGGCATTTAACGCATTAAGCTCTGGAAATTTCT 152 153 154 154 154 154 155 155 155 155	L	C	D	L	L	W	S	D	P	D D	K	D	T	M	GGC	W	GGA	E	N	D	R	GGT	V	S	F	225
ACT TIT GOGGC GAA GTAGTI GCAAAGT TCCTC CACAAG CAT GACTTI GACCICATCTG CGAGCTATCAG GTT T F G A E V V A K F L H K H D F D L I C R A H Q V GTI GAA GATGGT TAC GAGTTC TITGCC AAGCGA CAA CTA GTAACATTATTCTC A GCACCC AAC TACTGT GG GAA V E D G Y E F F A K R Q L V T L F S A P N Y C G E TIT GAC AATGCA GGT GCAATG ATGCTC GTAGAT GAG ACACTT ATGTG TCCTC CAGATA CTT AAGCCA GCA GAT AAA AAG AAGGTC CCC TATGGA GGATGT CT GTAGAT GAG ACACTT ATGTG TCCTC C CAGATA CTT AAGCCA GCA GAT K K K F P Y G G L N T G R P V T P P R G A A N Q K AAC AAG AAGAAA TAA AAACTAGTG TGGTG TGGTG TGGC CAACCG CCAGGA GCAGCCAAT CAG AAA N K K K * T Y G G L N T G R P V T P P R G A A N Q K ACC TTAAGT TTGGGATT GCTACAGGG GGA CTT CATAAAAGAAATTAAATT	_	-	-	-	_		_	-	-	-		-			-			_		-		-		-	-	
GTT GAA GATGGT TAC GAGTTC TT GCC AAGCGA CAA CTA GTA ACATTA TT CT CA GCACCC AAC TACTGT GGG GAA V E D G Y E F F A K R Q L V T L F S A P N Y C G E TTT GAC AATGCA GGT GCAATG ATGCTC TG GAG GAG CCA CTT ATGTGT TC CT CC CAGATA CTT AAGCCA GCA GAT F D N A G A M M S V D E T L M C S F Q I L K P A D AAA AAG AAGTTC CCCC TATGGA GGATGA CACA GGG CGA CCC GTAACG CCGCCG CGAGGA GCA GCCAAT CAG AAA K K F P Y G G L N T G R P V T P P R G A A N Q K AAC AAG AAGAAA AA AAACTAGTGTGTGGTGTGGCTCCTTGTGTGGCAAGTCTGAGGTGGCCTTTGTCATCCT N K K K * TGCCTTAAGTTT TGATGGATGAT GCACAGGG GGA CCT GTAAGACCAGGG CGAGCCAAT CAG AAA ATTTAT TGAATATCAGTGTGTGGTGGGTGCACTGCTTGGTGTGGCAAGTCTGAGGTTGGCCCAAT CAG AAA AAC AAG AAGAAA AA AAACTAGTGTGGGTGGGTCCCCTGTTGGTGTGGCAAGTCTGAGGTTGGCCCAAT CAG AAA A K K K * TGCCTTAAGTTT TGTGGATTGCTAAGGGTGTGGTTGGGACTTCAAAAAGAAATTAAATATTATTAAGAATTTAACTGGGTTATTTA 12 6 TTATTATAGAATATTTAATCTGGGATGAACAAGGGGTAATTAGACTTGGCAAGTCGAGGACATAAATCTGTTAATGAAATA ATTTTTT TGAATTTCAGTTGCTAGATTATCCATGATATTATATT	ACT	TTT	GGG	GCA	GAA	GTA	GTT	GCA	AAG	TTC	CTC	CAC	AAG	CAT	GAC	TTT	GAC	CTC	ATC	TGT	CGA	GCT	CAT	CAG	GTT	250
GTT GAA GATGGT TAC GAGTTC TTTGCC AAGCGA CAA CTA GTA ACATTA TT CT CA GCACCC AAC TACTGT GG GAA 96 V E D G Y E F F A K R Q L V T L F S A P N Y C G E 273 TTT GAC AATGCA GGT GCAATG ATGCTC TG GAG AG GAG AC CTT ATGTGT TC CT CC CAGATA CTT AAGCCA GCA GAT F D N A G A M M S V D E T L M C S F Q I L K P A D 300 AAA AAG AAG TT CCC CTATGGA GGATTG AACACA GGG CGA CCC GTAACG CCG CG CG AGGA GCA GCCAAT CAG AAA 1114 K K K F P Y G G L N T G R P V T P P R G A A N Q K 323 TGC CTTAAGTTT GTGGATTG CTACAGGG GGA CTT GTGTGTGGCAAGTCTGAGGTTGGC CTTGTCTCTCCT 119 TAA TAACATATTTAACTAGTGTGTGGTGTGGTGTGGA CTTCAAAAAGAAATAAA ATTTTAT GGATGC CTAGGATGAACAAGGG GTAGTTAGACATTGCAACCCAGGACATAAATCTGTGTCATCCT 119 TAATTAACATATTTATACCTGGATGAACAAGGG TG GA CTTCAAAAAGAAATAAATTTAACTGGG TTATTAT 126 TTATTAACATATTTATACCTGGATGAACAAGGG TAATTAGACTTGCAACCCAGGACATAAATCTGTTATGAAATA 134 ATTTTTTTCTATTTGAGTTGCTAGATTACCATGATAAGGGTAATTAGACATTGCAACCCAGGACATAAATCTGTTAATGAAATA 134 ATTTTTTTTTTTTTATGAATTTCAGTTGCTAAGTATATCCATGATATTTATT	•	-	•		-				~	-					2	-	2	-	-		~		~	×		200
TTT GACAATGCA GGT GCAATG ATGCT GTAAGAG AGG ACACTT AT GTGT TCCTC CAGATA CTT AAGCCA GCG CAGCA GCA GCA GCA GCA GCA GC	GTT	GAA	GAT	GGT	TAC	GAG	TTC	TTT	GCC	AAG	CGA	CAA	CTA	GTA	ACA	TTA	TTC	TCA	GCA	CCC P	AAC	TAC	TGT	GGG	GAA	964
TTT GACAATGCA GGT GCAATG ATGTCT GTAGAT GAGACA CTT AT GTGT TCCTTC CAGATA CTT AAGCCA GCA GAT 103 P D N A G A M M S V D E T L M C S F Q I L K P A D AAA AAG AAGTTC CCC TATGGA GGATTG AACACA GGG CGA CCC GTAACG CCGCCG CGAGGA GCA GCCAAT CAG AAA K K K F P Y G G L N T G R P V T P P R G A A N Q K AAC AAG AAGAAAAAA AAACTAGTGTGGGTGTGGGTGTGGCTTGGCAAGTCGAGGTGGGCTTGTCATCCT 119 N K K K * TGC CTTAAGTTTGGTGATTGCTACAGGGGATGTGGACTTCAAAAAGA <mark>AATAAA</mark> TATTGAATTTAACTGGGTTATTAA 126 TAATTAACATATTTAATCCTGGATGACAAGGGGTAATTAGACATTGCAACCCAGGACATAAATCTGTAATGAAATA 134 ATTTTATGAATTCCAGTGCTAGATGACAAGGGGTAATTAGACATTGCAACCCAGGACATAAATCTGTTAATGAAATA 134 ATTTTATGAATTTCAGTTGCTAGATTATCCATGATAGGTTATTATTATTGGTACATTGGCCCAGGTTCCCCCAGTAAAGT 142 TTAATTATCAGTTGCTGAAATAGGTTTTATCAGCAGGAGATTATATATTATTATTATGGTCCAAATTTCT 158 AGAAAAGAATTGGCTTGCAAAGGGTAATTAGGCATGGTATTAAACGCATTAGCCATGGTCCAAATTTCT 177 TATAAACTTAGGGCCCAAAGAGACTCGAGTGAAACGGGAGGAGGATAGTTAGACATTGGCTGAAATTTGT 173 TATAAACTTAGGTCCGGAGAGCCCGAGGGAGGAGGAGGAGGAGGTAGGT		P	2	0	-	5	-	-	~	R	R	×		•			-	3	~	F		-	C	0	-	275
AAA AAG AAGTTC CCC TATGGA GGATTGAACACA GGG CGA CCC GTAACG CCG CG AGGA GCA GCCAAT CAG AAA 111 K K K F P Y G G N T G R P Y T P R G A N Q K 323 AAA AAGAAGAAATAA AAACTAGTGTTGGTGTGGGTGTGGCACTGTCTTGGTGTGGCAAGTCTGAAGGTGGGCCTTGTCCATCCT 119 323 323 TGC CTTAAGTTTGTGGATTGCATTCAGGGGATGTGACTGCATGTGGAAGTGGACTTCAAAAAGAAATTAAACTGGGTTATTAA 126 324 TGC CTTAAGTTTGTGGATGGCATGCAAGGGGATGTGACTTCAAAAAGAAATTGGAACTAAATTTAACTGGGTTAATTAA	TTT	GAC	AAT	GCA	GGT	GCA.	ATG	ATG	TCT	GTA	GAT	GAG	ACA	CTT	ATG	TGT	TCC	TTC	CAG	ATA	CTT	AAG	CCA	GCA	GAT	1039
AAA AAG AAGTTC CCC TATGGA GGATTGAACACA GGG CGA CCC GTAACG CCG CCG AGGA GCAGCCAATCAG AAA 111' K K K F P Y G G L N T G R P V T P P R G A A N Q K 323 AAC AAG AAGAAATAA AAACTAGTGTGGGTGGGTGGGCTCTTGGTGTGGCAAGTCGAGGTGGGCTTGTCATCCT 119 N K K K * TGCCTTAAGTTTGTGGATTGCTATCAGGGGGATGTGACTTCAAAAAGAAATAAA TTATATAACATATTTAATCTGGGATGACCAGGGGATGTGACTTCAAAAAGAAATAAAT	-	D	IN	A	0	A	м	м	3	•	0	Б			м	C	3	F	×	-	5	R	F	A		300
K K K F F F G K	AAA	AAG	AAG	TTC	CCC	TAT	GGA	GGA	TTG	AAC.	ACA	GGG	CGA	CCC	GTA	ACG	CCG	CCG	CGA	GGA	GCA	GCC	AAT	CAG	AAA	1114
AAC AAG AAGAAA TAA AAACTAGTGTTGGTTGGTCACTGTCTTGGTGTTGGCAAGTCTGAGGTTGGCCTTTGTTCATCCT 119: N K K * 32: TGCCTTAAGTTTGTGGATTGCTATCAGGGGATGTGACTTCAAAAAGAAATAAAT	ĸ	ĸ	ĸ	r	P	ĭ	G	G	г	м	т	G	ĸ	P	v	т	P	P	ĸ	G	A	A	м	8	r	320
NKKKKK 32 TGCCTTARGTTTGTGGATTGCTATCAGGGGATGTGACTTCARARAGAATAAATATTGAATTTAACTGGGTTATTTA 126 TAATTAACATATTTTAATCCTGGATGACAAGGGGATGTGACTTCAAAAAGAAATAGAATTGCAACCCAGGACATAAATCTGTTAATGAAATA 134' ATTTTATGAATTTCAGTGCTAGATATCCATGATAGTTATATTTATT	AAC	AAG	AAG	AAA	TAA	AAA	CTA	GTG	TTG	GTGI	FCA	CTG	TTT	GGT	GTI	GGC	AAG	TCTO	GAGG	TTO	GCC	TTT	GTT	CAT	CCT	1191
TECCTTARGTTTEGTEGRTTGCTATCAGEGGATGTGCACTTCARARAGAATAAATATAGAATTAACTGGGTTATTAACT TAATTAACATATTTTAATCCTEGATGACAAGGGTAATTAGACATTGCACCCGGGACATAAATCGTTAATGAAATA ATTTTATGAATTTCAGTTGCTAGATTATCCATGATAGTTATATTAATTGGTACATTGGCCCAGGTCCCCTCAGTAAAG 1421 TAATTTTATCAGTTGCTTCAGATTATCCATGATATTGTTAATTTAATTGGTACATTGGCCCAGGTCCCCTCAGTAAAG 1421	м	ĸ	ĸ	ĸ	·																					329
TAATTAACATATTTTAATCCTGGATGAACAAGGGTAATTAGACATTGCAACCCAGGACATAAATCTGTTAATGAAATA 134' ATTTTATGAATTTCAGTGCCAGATTATCCATGATAGTTATTTAT	TGC	CTT	AAG:	TTTI	IGTG	GAT	TGC	TAT	CAG	GGGA	ATG:	TGAC	TTC	AAA	AAG	AAA	TAA	ATA1	TGA	ATI	TAA	CTG	GGT	TATI	TA	1269
ATTTTATGAATTTCAGTTGCTAGATTATCCATGATAGTTATATTTATT	TAA	TTA	ACA	FATI	TTA	ATC	CTG	GAT	GAAG	CAAG	GG	FAAT	TAG	ACA	TTG	CAA	CCCI	AGGI	CAT	AAA	TCT	GT T.	AAT	GAA	ATA	1347
TTATTTTTCTATTTTGAACCTTCATTTTGTTTTTTTTATTCATTTATTAATTA	ATT	TTA	TGA	ATTI	CAG	TTG	CTA	GAT	TATO	CAI	GA	TAG	TAT	ATT	TAT	TGG	TAC	ATTO	GCC	CAG	GTT	ccc	TCA	GTA	AAG	1425
AAACTTTTGTTTTATCCGAAATAGGTTTTATCAGCAGTATTAATGCTAAAAGGTTTGTCTATTGGTTCCAAATTTTCT 158: AGAAAAGAATTGGCTTGCAACTGGAATTACGGAAAATTTGTTCAACTTCAGTTTATATCTTGGAAGATAAGTTTGAC 165: TTGTATTAATATATATATTTCAGTGTGTAAATAGCCATGGTATTTAACCGCATATAAGCATTACCATGTTCTG 173: TATAAACTTTAGGGCCCAAAGAAGACCGAGGGGAGGGAGG	TTA	TTT	TTC	TATI	TTTG	AAC	CTT	CAT	TTT	IGTI	TT	ATTO	CATT	TAT	AAT	TAT	TAA	TAC	TGT	ATI	TTT	TAG	GGC	AAA	GTC	1503
AGAAAAGAATTGGCTTGCAACTGTGATTATCGGAAAATTTGTTCAACTTCAGTTTATATCTTGGAAGATAAGTTTGAC 1650 TTGTATTAATATATATATTTCAGTGTGTAAATAGCCATGGTATTTAACCGCATATAAGCATTTACCATGTTCTTG 1730 TATAAACTTTAGGGCCCAAAGAAGACCGAGTGAAGCAGGGGAGGGA	AAA	CTT	TTG	TTT	TATC	CGA	AAT	AGG	TTT	TATO	CAG	CAGI	TTAT	AAT	GCT	AAA	AGG	TTT	TCT	ATI	GGT	TCC	AAA	TTTT	TCT	1581
TTGTATTATATATATATATTTCAGTGTGTAAATAGCCATGGTATTTAACCGCATATAAGCATTTACCATTGTTCTTG 173' TATAAACTTTAGGGCCCAAAGAAGACCGAGTGAAGCAGCGAGGGAGG	AGA	AAA	GAA	TTGO	GCTI	GCA	ACT	GTG	ATT	ATCO	GA	AAA	TTG	TTC	AAC	TTC	AGT	TAT	ATC	TTO	GAA	GAT	AAG	TTTO	GAC	1659
TATARACTTTAGGECCCAARGAGACTCGAGTGAAGCAGCGAGGGGATGATAGTTAGAGAAGCTGCTGGTTAGGTTAG 181 TGTACARATGTCCAGTTGCCGTAGTAGCAGGAGCATCTGGCTTGGCTTGCTATCGCTTGTATAGTCAGAACAAGAGCGCAC 189 TGGTGGTGGTATGTCGAGAGCCCAACCCCTACCCCAAAACCTTTCGAGAGAAGAAAGA	TTG	TAT	TAAT	TATA	TATA	ATA	TTT	CAG	TGTO	GTAA	AT	AGCO	CATG	GTA	TTT	AAC	CGCI	TAT	AAG	CAT	TTA	CCA	TTG	TTCI	TTG	1737
TGTACAAATGTCCAGTTGCCGTAGTAGCAGGAGCATCTGGCTTTGCTATCGCTTGTATAGTCAGAACAAGAGCGCCC 1890 TGGTGGGGGTATGTCGAGAGCCCAACCCCTACCCAAAACCTTTCGAGAGAGA	TAT	AAA	CTT	TAGO	GCC	CAA	AGA	AGA	CTCC	GAGI	GA	AGCI	AGCG	AGG	GGA	TGA	TAG	TAG	AGA	AGO	TGC	TGG	TTA	GGTI	FCA	1815
TGGTGGTGGTATGTCGAGAGCCCAACCCCTACCCAAAACCTTTCGAGAGAGA	TGT	ACA	AAT	GTCC	CAGI	TGC	CGT	AGT	AGCI	AGGA	GCI	ATCI	GGC	TTT	TGC	TAT	CGCI	TGT	ATA	GTC	AGA	ACA	AGA	GCG	CAC	1893
CGAGGACTTGCATCATACATCCACCAGGTAAACCAGGTATTGGCTAATATACCCATGCTAGATTTGTGGAGTTTGGGC 2049 ATGGTGTGAAAAGTCCCTTAATGTGCAAAAGTAAAATACATGTACCAAAGCAGATGTAAATTGTGGTGTTAATTCAA 2127 TAAAAGCGTTCATCTCGGAAAAAAAAAAAAAAAAAAAAA	TGG	TGG	TGG	TATO	STCG	AGA	GCC	CAA	ccco	CTAC	cca	AAAA	ACCT	TTC	GAG	AGA	GAAJ	GAA	ATG	ACI	CTC	TTT	ccc	CACI	ACC	1971
ATGGTGTGAAAAGTCCCTTAATGTGCAAAAGTAAAATACATGTACAAAGCAGATGTAAATTGTGGTTGTTAATTCAA 212 TAAAAGCGTTCATCTCGGAAAAAAAAAAAAAAAAAAAAA	CGA	GGA	CTT	GCAI	TAD	ACA	TCC	ACC	AGGI	FAAA	ICC1	AGGI	TTAT	GGC	TAA	TAT	ACCO	CATO	CTA	GAT	TTG	TGG.	AGT	TTGO	GC	2049
TAAAAGCGTTCATCTCGGAAAAAAAAAAAAAAAAAAAAA	ATG	GTG	TGA	AAAG	STCC	CTT	AAT	GTG	CAAA	AAGI	AA	AATA	CAT	GTA	CCA	AAG	CAG	ATGI	AAA	TTO	TGG	TTG	TTA	ATTO		2127
	TAA	AAG	CGT	r CA 1	TCTC	GGA	AAA	AAA	AAA/	AAAA	LAAJ	AAA/	AAA	AAA												2171

Fig. 1. Full-length cDNA and *PmSTPP1-\alpha* amino acid sequence showing single conserved domain (highlighted in gray) – CDART result in multi-domain was PP2Ac which is a homologue of PP1 catalytic subunit. The Kozak's sequence and down promoter element are colored red. The two poly-adenylation sites (AATAAA) and poly-A tail were underlined.

developing isoform-specific antibodies (Korrodi-Gregório et al. 2014). The PP1C C-terminus has been documented to interact with myosin phosphatase (MYPT1) to promote dephosphorylation in the smooth muscle tissue. This interaction has been shown to promote structural changes in the PP1C C-terminus, from a simple coil to a more ordered state (Terrak et al. 2014).

PmSTPP1- α possesses two (2) aspartic acids (D64 and D92), three (3) histidines (H66, H173, and H248), and one (1) asparagine (N124) that are conserved in human PPPs.

Jacqueline Marjorie R. Pereda et al.

These conserved amino acids are thought to catalyze reactions when two metal ions bind and activate a water molecule, inducing a nucleophilic attack on the phosphorus atom. The two metal ions that contributed to the catalysis for *PmSTPP1-* α were assumed to be either Zn2+, Cu2+, Ni2+, Fe2+, and Mn²⁺ (Dancheck et al. 2011; Kolupaeva 2019), while the bimetallic Iron (Fe2+) and Zinc (Zn²⁺) were specific for PPP family (Heroes et al. 2015; Verbinnen et al. 2017). Structures of the PP1 complexes showed that the α -isoform was determined to contain two Mn2+ ions that interacts with spinophilin, which is a neuronal regulator protein bound (Ragusa et al. 2010). The bimetallic center is surrounded by the said metal ligands and is known to be the active site of the catalytic sub-unit.

Within the catalytic sub-unit of *PmSTPP1-\alpha*, a total of three (3) characteristic conserved sequence motifs and three (3) conserved sequence residues were present; GDIHG, GDYVDRG, GNHE, HGG, RG, and H (G, glycine; D, aspartic acid; I, isoleucine; H, histidine; V, valine; R, arginine; N, asparagine; E, glutamic acid; Y, tyrosine). *PmSTPP1-\alpha* showed a conserved RVxF motif suggesting that it functions similarly with the RVxF motif of PP1 in humans and the shrimp could possess that same PP1 regulatory proteins that would be interesting to study further. The said motifs and sequences were inclusive in PPP family (Brautigan 2013) and functions in metal coordination and phosphate binding (Shi 2009). Since the said motifs and sequences were conserved in STPP1 as shown in the sequence alignment (Fig. S2), it is likely that *PmSTPP1-* α also functions the same way.

PP1 regulatory proteins act as either substrate-independent activity regulator, targeting subunits, substrate specifier or substrate that interacts to the catalytic subunit (Heroes et al. 2013). They mostly interact with PP1C via RVxF binding motif (Korrodi-Gregório et al. 2014). The RVxF motif is composed of the consensus sequence K/R][K/R][V/I][x][F/W], denoted in standard single letter amino acid code: Lysine (K) or Arginine (R), Valine (V) or Isoleucine (I), Phenylalanine (F) or Tryptophan (W); where x is any residue other than Phenylalanine (F), Isoleucine (I), Methionine (M), Tyrosine (T), Aspartic Acid (D), or Proline (P). The RVxF



Fig. 2. Neighbor-Joining Tree of *PmSTPP1-a* protein sequence using Poisson Model. *PmSTPP1-a* was denoted with yellow-filled box while *LvSTPP1-β* was marked with broken-line unfilled box. Voucher sequences of Serine/Threonine Protein Phosphatase (PPP, PPM and FCP/SCP families) amino acid of the species indicated above were included in the analysis.

interaction functions in the protein binding regulation and doesn't influence the catalytic activity of the active site of PP1C. The glycogen-targeting subunit – muscle (GM) and inhibitor-2 (I-2) regulatory protein binds to the PP1C with the RVxF binding pockets (Peti et al. 2013) as specified in the sequence alignment (Fig 4).

PmSTPP1-\alpha Structural Analysis

BLAST Analysis

The UniProt BLAST search results identified 2 entries in the database with available 3D structures. These entries were for STPP1- α proteins from *H. sapiens* and *O. cuniculus*. The sequences of these related proteins were determined and aligned pairwise with *PmSTPP1-\alpha* using Emboss Needle (https://www.ebi.ac.uk/Tools/psa/ emboss_needle/). Initial alignment of the related sequence from the *H. sapiens* STPP1a structures (PDBID 3e7a Chain A; 293 aa) and the *PmSTPP1-\alpha* sequence (329aa) revealed 86.9% identity. The variance mainly occurred in the N and C terminal sections which did not align for the two sequences. Pairwise alignment between length matched sequences (293aa) of *Pm STPP1-* α and STPP1- α from *H. sapiens* was found to have 97.6% identity.

Model Fitting

The polypeptide sequence of *PmSTPP1-\alpha* was fit unto the related sequence of *H. sapiens* STPP1 α , HSSTPP1a, (PDBID 3e7a, Kelker et al. 2009) using the Magic Fit function of the DeepView Molecular Viewer (Guex et al. 1997). The generated structures showed good alignment of the conserved regions of $PP1\alpha$ structure as listed in Table 2. Locations of the variant residues between the two structures are shown in Figure 3 B-D. The sites of variation are observed to be outside the conserved areas in STPP1-a, proteins. A comparison of the line models for the H. sapiens STPP1- α residues at the sites of variation in the proposed P. monodon STPP1- α shows similar orientations for most of the residue sidechains (Figure 3 B-D). The most variance is observed for the position of Met208 and Phe235 in the fit *PmSTPP1-* α structure (Fig 3 C).

Homology Model Analysis

Structural models were also predicted using the ITASSER Homology Modelling Server (https://zhanggroup.org/I-TASSER/). The

polypeptides sequence of $PmSTPP1-\alpha$ was submitted for modelling and several models were generated. Structure model predictions are made by threading the submitted sequence through related/homologous reference structures. The top threading reference for $PmSTPP1-\alpha$ was a structure for STPP1-β from *H. sapiens* (PDBID 1s70; Terrak et al. 2004). This top model was given a confidence score (C-score) of 0.16. C-scores range from -5 to 2, with scores nearing 2 representing higher confidence for the predicted structure. As noted in the sequence alignment, most of the variance between $PmSTPP1-\alpha$ and its references occur in the unmatched N and C-terminal sections. In fact, a resubmission of a trimmed *PmSTPP1-\alpha* sequence without these sections returns only one predicted model with a near perfect C-score of 1.82.

To compare the structures generated by model fitting and by homology modeling, the generated models were analyzed based on a common reference structure, *H*.

Table 2. Conserved functional domains in STPP1. Identical residues were found for the models of $HsSTPP1-\alpha$ and $PmSTPP1-\alpha$ for these conserved functional domains. The color scheme used for the functional domains is like that used for the structures shown in Figures 3 and 4.

Conserved Functional Domain	Residues in HsSTPP1-α	Residues in <i>PmSTPP1-α</i>
Conserved active site/	D64; H66;D92; N124; H173; H248	D64; H66;D92; N124; H173; H248
Metal coordination site	*Identical Residues	*Identical Residues
Target protein hinding domain	D71; R74	D71; R74
rarget protein binding domain	*Identical Residues	*Identical Residues
	K168; I169; F170; D242; L243; F257; R261; L289; M290;	K168; I169; F170; D242; L243; F257; R261; L289; M290;
RVXF binding pockey	C291; F293	C291; F293
Molecular toxin interaction sites	R96; E126; S129; 1130; Y134; W206; D220; R221; G222; V223; Y272	R96; E126; S129; 1130; Y134; W206; D220; R221; G222; V223; Y272
	*Identical Residues	*Identical Residues
Destaudies Lass	N271; D277	N271; D277
Protruding Loop	*Identical Residues	*Identical Residues

Table 3. Neighboring residues for the original and variant residues between $HsSTPP1-\alpha$ and $PmSTPP1-\alpha$. Most of the observed residue contacts are similar. But potential residue interactions are observed to be altered due to some variations. These include changes in potential residue contacts as seen with the change from Leu241 to Phe241. The modelling of the C-terminal domain also predicts interactions with other residues in the $PmSTPP1-\alpha$ protein. Investigations on the relevance of these contacts for affecting the protein function will be included in future studies.

HsSTPP1-α (3e/7a/ PmSTPP1-α (model)	H. sapiens STPP1 Residue Contacts	<i>P. monodon</i> STPP1 Residue Contacts
Leu9/ Ile9	7, 8, 9, 10, 11, 12, 13, 14, 38, 109, 112, 113 *Identical contacts	7, 8, 9, 10, 11, 12, 13, 14, 38, 109, 112, 113 *Identical contacts
Gly14/ Ala14	9-18 *Identical Contacts	9-18 *Identical Contacts
Glm20/Arg20	17-22, 73, 77, 81	16, 17-22, 70, 73, 77, 81 *Additional Contacts: Leu16; Tyr70
Ile161/ Val161	51, 121, 122, 159-163, 171-173, 186, 201, 204, 205 *Identical contacts	51, 121, 122, 159-163, 171-173, 186, 201, 204, 205 *Identical contacts but the following residues were predicted to form sheets in the model. Sheets: 51, 121, 122;160-162
Val213/ Thr213	210-215; 216 ; 217; 226-229 *Additional contact: Trp216	210-215; 217; 226-229
Glm214/ Met214	212-216; 228; 229 ; 230; 231 *Additional contact: Ala229, Val231	212-216; 228; 230
Leu241/ Phe241	164, 167, 168, 169 ; 170; 183; 235; 236; 239-243 *Additional contact: Ile169	164, 167-168; 170; 183; 235; 236; 237 ; 239-243 *Additional contact: His237
C-terminal domain (Asp300– Lys329)	N/A	66, 68, 71, 96-98, 125, 133, 134, 138, 142, 250, 267, 270- 277, 298, 299

sapiens STPP1- α , (PDBID 3e7a). Upon fitting the top ITASSER predicted model for the full-length *Pm STPP1-\alpha* sequence unto 3e7a, similar structures were observed between Leu7-Ala299 (Fig 4A). Identical residues were observed for both proteins in the predicted functional domains (i.e. metal coordination sites, target protein binding site, RVxF domain, Toxin binding domains, and the protruding loop; Table 2; Fig. 4 B). The variant residue identities in *Pm STPP1-\alpha* compared to *HsSTPP1-\alpha*, occurred outside these conserved functional domains (Fig 4 C).

While the variant residues in *Pm STPP1-* α were found outside the conserved functional domains, it was still important to note potential changes in residue interaction that may be brought about by the changes. Table 3 lists the residues within a 5 distance from the original and variant residues in *HsSTPP1-* α and *Pm STPP1-* α ,

respectively. This proximity allows the formation of contacts between residues. Most of the variance was conservative (e.g. L9I; I161V, etc.) but several changes in residue contacts were observed (Table 3). One change, from Leu241 to Phe241 results in the loss of an interaction with Ile169, and an additional contact with His 237. While both the Leu241-Ile169 and Phe241-His237 involve non-polar interactions, the potential creation of Pi bonds through the aromatic rings in the latter pair may provide a stronger bond. It is interesting to note that the Leu241 to Phe241 variation has not been observed in the other related STPP1 forms (Table 2).

It must also be noted that the modeled structure included predictions for the conformation of the flanking N and C-terminal sections. The N-terminal sequences (Met1-Lys6); and the C-terminal section from Asp300-Lys329 did not have corresponding structures in the *H. sapiens* STPP1- α structure (Fig 4 D). The C-terminal

Jacqueline Marjorie R. Pereda et al.



Fig. 3. Variant locations between STPP1- α in *H. sapiens* and *P. monodon* in Fit Model structures. The residues observed to vary between the two source species do not occur in the conserved functional domains for STPP1- α . (A) Conserved functional domains in STPP1- α , colored as follows in the ribbon structure: Metal coordination site (Red); Target protein binding site (Orange); RVxF domain (Yellow) and Toxin Binding Sites (Green). (B) Residues in *H. sapiens* STPP1- α at the variant locations rendered as line models and colored in CPK format. (C) Residues in the predicted *P. monodon* STPP1 α at the variant locations rendered as line models and colored in CPK format. (D) Variant residue line models in STPP1- α from *H. sapiens* (green) and *P. monodon* (red).

section is in fact predicted to interact with other sections of the protein, including areas that are associated with toxin binding, the metal coordination site, and the expected target protein binding domain (Fig 4 E, F). How these predicted interactions translate to effects on the expected function of *Pm STPP1-a* would be the subject of further studies.

Functional Domain Predictions

In addition to generating structure models for the submitted sequences, ITASSER also provides predictions on potential function based on the presence of characteristic domains. The submitted sequence for *PmSTPP1-a* was predicted to have a conserved metal coordination site involving D64 and H248. These corresponds with the predicted metal coordination sites predicted based on the polypeptide sequence (Table 2).



Fig. 4. Homology Modeled Structure of PmSTPP1-a. (A) Homology modelling predicts a similar structure for PmSTPP1- α (blue) as that of HsSTPP1- α (white). (B) Identical residues and locations were observed for the functional domains (i.e. Metal coordination site (Red); Target protein binding site (Orange); RVxF domain (Yellow); Toxin binding sites (Green); Protruding Loop (Blue)). (C) The variant residues for PmSTPP1-α occur outside these conserved functional domains. Original residues in the HsSTPP1- α are shown in green; and variant residues in PmSTPP1-a are shown in red. (D) The generated homology models included N and C-terminal domains (violet) which were not observed for the reference HsSTPP1-a structure. (E, F) The modelled Cterminal domain is predicted to interact with other regions of the protein, including the Metal Coordination Site, the Target Protein Binding Domain, and the Toxin binding sites.

The sequence was also associated with structures of enzymes with class designations, EC 3.1.3.16, which corresponds to Protein Phosphatases, and S/T specific Protein Phosphatases. Both results suggest that the function of the predicted $PmSTPP1-\alpha$ protein matches the expected S/T phosphatase function of its related proteins.

Immune-Related Gene from Black Tiger Shrimp, Penaeus monodon



Fig. 5. Predicted Dock positions for PirA and PirB. The functional domains for PmSTPP1-a are colored similarly as in Figure 3 (i.e. Green: Molecular Interaction Sites; Orange: Target Protein Binding domain; Yellow: PVxF region, etc.). (A) Eight of ten docks based on balanced factors were predicted to bind the molecular toxin binding region for PirA. (B) Six of ten docks based on balanced factors were predicted to bind the molecular toxin binding region for PirB. (C) Two docks (colored blue and magenta) were predicted to bind with the Target Protein Binding domain for PirA. (D) Two docks (colored blue and magenta) were predicted to bind with the Target Protein Binding domain for PirB. (C) Two docks (colored cyan and yellow) were predicted to bind with the Target Protein Binding domain for PirB. (E) The predicted position for PirA binding to the molecular toxin interaction site allows PirB binding with the target binding region and PirB is bound to the molecular toxin interaction site. (A-D) Structures are shown in side-view perspective; (E,F) Structures are shown from the top-view perspective. (C) The position of the modelled C-terminal domain is rendered as a blue ribbon with visible VDW spheres. This is seen to be placed between the docking positions for the molecular toxin interaction sites, and the target protein binding domain. The molecular structures used in this figure were rendered using the DeepView molecular viewer (Guex et al. 1997).

In Silico Protein Docking Predictions

The predicted involvement of $Pm STPP1-\alpha$ with *P. monodon* immune response was assessed based on predicted associations with protein targets from known Shrimp pathogens. Specifically, protein docking experiments were done with three proteins (PirA, PirB, and VP24). Both PirA and PirB are toxins from *V. parahaemolyticus* (Lee et al. 2015), while VP24 is a major envelope protein of White Spot Syndrome Virus (Sun et al. 2016).

Predicted Dock Models: PirA

Eight out of the top 10 models generated based on "balanced" factor contribution placed the PirA structure in close association with the expected molecular toxin interaction sites of *PmSTPP1-a* (Table 2; Fig. 5 A). The remaining two models in the top 10 showed predictions for association with the target protein binding domain of *PmSTPP1-a* (Table 2). These observed dock positions

occur at opposite flanking sides of the *PmSTPP1-* α structure (Fig. 5 C).

The position of the toxin interaction site docked models was observed to be similar to the top-docked structures in two other groups, those ranked based on electrostatics and hydrophobics. These results suggest the relevant involvement of these types of interactions in the predicted association of *Pm STPP1-\alpha* and PirA.

Predicted Dock Models: PirB

Six out of the top 10 models generated based on "balanced" factor contribution placed the PirA structure in close association with the expected molecular toxin interaction sites of *PmSTPP1-a* (Table 2; Fig. 5 B). Two of the remaining models in the top 10 showed predictions for association with the target protein binding domain of *PmSTPP1-a* (Table 2; Fig. 5 D). The remaining two models were predicted to bind different positions in *PmSTPP1-a*.



Fig. 6. Predicted Dock positions for VP24. The functional domains for *PmSTPP1-α* are colored similarly as in Figure 3 (i.e. Green: Molecular Interaction Sites; Orange: Target Protein Binding domain; Yellow: PVxF region, etc.). (A) Two of ten docks based on balanced factors (red and magenta) were predicted to bind the Target Protein Binding domain of PmSTPP1-a. The red ribbon represents the position of the top ranked dock. (B) Two of ten docks based on balanced factors were predicted to bind the RVxF domain. (C) Six of ten docks based on balanced factors were predicted to bind near the toxin binding site of PmSTPP1-a. (D) It must be noted, however, that the binding positions for these six docks were not observed to coincide with the predicted toxin interaction site bound by PirA (foreground; red ribbon). The molecular structures used in this figure were rendered using the DeepView molecular viewer (Guex et al. 1997).

The position of the target protein binding domain docked models was observed to be similar to the topdocked structure for the models ranked based on hydrophobics. While not matching the top dock, the position predicted for the 6 models bound to the toxin interaction site was similar as the 2nd ranked model based on hydrophobics. These results suggest the relevant involvement of hydrophobic interactions in the predicted association of *PmSTPP1-α* and PirB through either the toxin interaction sites or the target protein binding surface.

Predicted Dock Models: VP24

Two out of the top 10 models generated based on "balanced" factor contribution placed the VP24 structure in close association with the expected target protein binding domain of *Pm STPP1-a* (Table 2; Fig. 6 A). Two of the remaining models were also near this site but were shifted higher towards the PVxF domain (Table 2; Fig. 6 B). The remaining six models were predicted to bind

different positions in *PmSTPP1-* α near the toxin interaction site, but at distinctly different positions as that predicted for PirA binding. (Fig. 6 C, D)

The position of the docked models near the PVxF domain was observed to be similar to the top-docked structure for the models ranked based on electrostatics. The position predicted for the 6 models near the toxinbinding site was similar as the top ranked model based on hydrophobics. These results suggest the relevant involvement of both electrostatic and hydrophobic interactions in guiding the predicted associations of *PmSTPP1-a* and VP24 through contacts near the toxin interaction sites or the target protein binding surface.

Predicted Dock Models: PirAB vs VP24

The predicted involvement of *PmSTPP1-* α in the immune response for *P. monodon* was investigated based on its potential interactions with two types of viral pathogens. These were a membrane protein from WSSV (VP24) and components of a toxin protein (PirAB) from V. parahaemolyticus. The observed interactions with the two target types had similarities and differences. Both target types were observed to have predicted interactions with the documented target protein binding surface of *PmSTPP1-a*. Associations were also observed for VP24 and both PirAB components (PirA and PirB) near the toxin interaction sites of $PmSTPP1-\alpha$. However, the predicted docking regions near this site for VP24 were distinctly separate from those favored by PirA (Fig. 6 D). This suggests specificity for binding target types bound by the toxin interaction sites. Our current results show selective binding for toxin protein components over other targets (VP24 membrane protein) in the predicted toxinbinding site for $PmSTPP1-\alpha$. The functional relevance of this selective function must be further analyzed to understand the role of *PmSTPP1-* α in the immune response.

Predicted Dock Models: PirA and PirB

It is interesting to note that some of the predicted docks for PirA binding to the $PmSTPP1-\alpha$ toxin binding site allows the binding of PirB to the target protein binding region (Fig. 5 E, F). As both PirA and PirB are components of the PirAB toxin from *V. parahaemolyticus* these observed binding combinations predict the possibility for $PmSTPP1-\alpha$ to associate with the PirAB toxin through multiple surfaces, both as single-surface interactions (i.e. toxin binding site only; target protein binding site only); combined action on two different surfaces; or as a linker protein binding two PirAB toxin molecules. The presence of the modelled C-terminal domain between the toxin interaction site and the protein



Fig. 7. Basal expression at organ level of *PmSTPP1-a* gene relative with the EF1- α , as the internal control. The seven (7) organs from three (3) WSSV-negative *P. monodon* were analyzed (1st lane: low mass DNA ladder; 2nd to 8th lanes were as follows: *G*, gills; *HP*, hepatopancreas; *IN*, intestine; *M*, muscle; *LO*, lymphoid organ; *HE*, heart; *H*, hemolymph; -, negative control).

binding domain (Fig. 5 C) may influence the relative positioning of the PirA and PirB components. The significance of these different interaction possibilities on the potential role of $PmSTPP1-\alpha$ for the immune response will be the subject of future studies.

PmSTPP1-α Gene Constitutive Expression

PmSTPP1-\alpha gene was expressed ubiquitously in tissues and organs of WSSV-negative P. monodon (Fig. 7), suggesting a probable involvement of the gene in the overall physiology of the species. In general, STPP1- α is persistently distributed in tissues, and enriched in the brain and heart (cardiomyocytes) (Aoyama et al. 2011; Peters et al. 2009; Vafiadaki et al. 2013). Specifically, in humans, PP1 was found to be highly involved in a variety of cellular processes such as protein synthesis, apoptosis, meiosis and cell division, cytoskeletal reorganization, regulation of membrane receptors and channels, and metabolism (Bollen et al. 2010). Korrodi-Gregorio et al. (2014) described the distribution of PP1- α in tissues such as the brain (ubiquitous but enriched in striatum and hippocampus) and testis (cytoplasm of Leydig and peritubular cells, spermatogonia, and preleptotene spermatocytes); in polarized cells such as spermatozoa and neurons (dendritic spines, perikaryal cytoplasm and nucleus); in cellular processes such as interphase (cytoplasm especially in centrosome, and nucleus particularly the nuclear matrix and nucleoplasm) and mitosis (also in centrosome which is throughout the mitosis stage, another is in kinetochores which is during the metaphase, and the midbody which is in telophase). Such varied PP1- α gene function could also be found in *P. monodon*, since *PmSTPP1-* α is constitutively expressed in gills, hepatopancreas, intestine, muscle, lymphoid organ, heart and hemolymph.

CONCLUSION

To date, this study is the first cloning report of alphasubunit in STPP1 present in *P. monodon* encoding the complete C-terminal end which is found to be very

similar to human PP1C α (Kim et al. 2015). In silico predictions of *PmSTPP1-\alpha* structure were in observed to share a similar form as human STPP1a, especially in the conserved functional domains. Some residue variations were still observed outside these areas, and their significance for *PmSTPP1-* α function must be further studied. *PmSTPP1-\alpha* was found to be ubiquitously and highly expressed in organs of WSSV-negative P. monodon and appears to be very critical for the maintenance of its general physiology. With this, it may also specifically be involved in shrimp-virus interaction and should be studied further on its probable role in shrimp immunity. Current molecular docking predictions for $PmSTPP1-\alpha$ against three proteins from known P. monodon pathogens suggest specific functional interactions with the target protein binding domain and the molecular toxin interaction sites. Further investigations on the interactions of this protein will help validate its predicted involvement in the *P. monodon* immune response.

REFERENCES CITED

- AOYAMA H, IKEDA Y, MIYAZAKI Y. et al. 2011. Isoform-specific roles of protein phosphatase 1 catalytic subunits in sarcoplasmic reticulum-mediated Ca (21) cycling. Cardiovascular Research, 89:79–88. DOI:10.1093/cvr/cvq252
- BERMAN H, WESTBROOK J, FENG Z, GILLILAND G, BHAT T, WEISSIG H, SHINDYALOV I, BOURNE P. 2000. The protein data bank. Nucleic Acids Research, 28:235-242.
- BOLLEN M, PETI W, RAGUSA MJ, BEULLENS M. 2010. The extended PP1 toolkit: designed to create specificity. Trends in Biochemical Science, 35:450–458. DOI:10.1016/j.tibs.2010.03.002
- BRAUTIGAN D. 2013. Protein Ser/Thr phosphatases the ugly ducklings of cell signaling. The Federation of European Biochemical Societies Journal, 280:324-345. DOI: 10.1111/j.1742-4658.2012.08609.x

- CIEŚLA J, FRACZYK T, RODE W. 2011. Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta Biochimica Polonica, 58:137-148. PMID: 21623415.
- DANCHECK B, RAGUSA M, ALLAIRE M, NAIRN A, PAGE R, PETI W. 2011. Molecular investigations of the structure and function of the protein phosphatase 1-spinophilin-inhibitor 2 heterotrimeric complex. Biochemistry, 50:1238–1246. DOI:10.1021/bi101774g
- DANG LT, KOYAMA T, SHITARA A, KONDO H., AOKI, T., HIRONO, I., 2010. Involvement of WSSV shrimp homologs in WSSV infectivity in kuruma shrimp: *Marsupenaeus japonicus*. Antiviral Research, 88:217-226. DOI: 10.1016/j.antiviral.2010.08.017
- DIEU B, MARKS H, SIEBENGA J, GOLDBACK R, ZUIDEMA D, DUONG T, VLAK J. 2004. Molecular epidemiology of white spot syndrome virus within Vietnam. Journal of General Virology, 85:3607-3618. PMID: 15557233.
- FLEGEL T. 2006. Detection of major penaeid shrimp viruses in Asia, a historical perspective with emphasis on Thailand. Aquaculture, 258:1-33. DOI:10.1016/j.aquaculture.2006.05.013
- GEUX N, PEITSCH M. 1997. Swiss-model and the Swiss-Pdb viewer: An environment for comparative protein modeling. Electrophoresis, 18:2714 - 2723.
- HEROES E, LESAGE B, GORNEMANN J, BEULENNS M, MEERVELT L, BOLLEN M. 2013. The PP1 binding code: a molecular-lego strategy that governs specificity. The Federation of European Biochemical Societies Journal, 280:584-595. DOI:10.1111/j.1742-4658.2012.08547.x
- HEROES E, RIP J, BEULLENS M, VAN MEERVELT L, DE GENDT S, BOLLEN, M. 2015. Metals in the active site of native protein phosphatase-1. Journal of Inorganic Biochememistry, 149:1–5. doi:10.1016/ j.jinorgbio.2015.03.012
- KELKER M, PAGE R, PETI W. 2009. Crystal structures of protein phosphatase-1 bound to Nodularin-R and Tautomycin: a novel scaffold for structure-based drug design of serine/threonine phosphatase inhibitors. Journal of Molecular Biology, Volume 385, Issue 1. http://doi.org/10.2210/pdb3E7A/pdb
- KOLUPAEVA V. 2019. Serine-threonine protein phosphatases: Lost in translation. Biochimica et Biophysica Acta (BBA) – Molecular Cell Research, 1866(1):83-89. DOI: 10.1016/j.bbamcr.2018.08.006

- KORRODI-GREGORIO L, ESTEVES SLC, FARDILHA M. 2014. Protein phosphatase 1 catalytic isoforms: specificity toward interacting proteins. Translational Research, 164(5):366-391. doi:10.1016/j.trsl.2014.07.001
- KOZAKOV D, HALL D, XIA B, PORTER K, PADHORNY D, YUEH C, BEGLOV D, VAJDA S. 2017. The ClusPro web server for protein-protein docking. Nature Protocols, 12(2):255-278.
- LEE C, CHEN I, YANG Y, KO T, HUANG Y, HUANG J, HUANG M, et al. 2015. The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by acquiring a plasmid that expresses a deadly toxin. Proceedings of the National Academy of Sciences, 112 (34):10798-10803; DOI: 10.1073/pnas.1503129112
- LI F, XIANG J. 2013. Signaling pathways regulating immune responses in shrimp. Fish & Shellfish Immunology, 34:973-980. DOI:10.1016/j.fsi.2012.08.023
- LU L, KWANG J. 2004. Identification of a novel shrimp protein phosphatase and its association with latency related ORF427 of white spot syndrome virus. FEBS Letters, 577:141–146. DOI:10.1016/j.febslet.2004.08.087
- MA G, ZHOU R, SONG H, ZHU H, ZHOU Z, ZENG Y. 2015. Molecular mechanisms of serine/threonine protein phosphatase 1 (PP1ca-PP1r7) in spermatogenesis of *Toxocara canis*. Acta Tropical, 149:148-154. doi:10.1016/j.actatropica. 2015.05.026
- MARALIT B, VENTOLERO M, MANINGAS M, AMAR E, SANTOS M. 2014. Substracted transcriptome profile of black tiger shrimp (*Penaeus monodon*) that survived WSSV challenge. Dataset Papers in Science, 2014:1–11. DOI: 10.1155/2014/807806
- NEEDLEMAN S, WUNSCH C. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology, 48(3):443-53. DOI: 10.1016/0022-2836(70) 90057-4. https://www.ebi.ac.uk/Tools/psa/ emboss_needle/
- PETERS M, BLETSCH M, CATAPANO R, ZHANG X, TULLY T, BOURTCHOULADZE R. 2009. RNA interference in hippocampus demonstrates opposing roles for CREB and PP1 alpha in contextual and temporal long term memory. Genes, Brain and Behavior, 8:320-329. DOI:10.1111/j.1601-183X.2009.00474.x
- PETI W, NAIRN A, PAGE R. 2013. Structural basis for protein phosphatase 1 regulation and specificity. The Federation of European Biochemical Societies Journal, 280:596-611. DOI: 10.1111/j.1742-4658.2012.08509.x

- RAGUSA MJ, DANCHECK B, CRITTON DA, NAIRN AC, PAGE R, PETI W. 2010. Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. Nature Structural and Molecular Biology, 17:459–464. DOI: 10.1038/ nsmb.1786
- ROY A, KUCUKURAL A, ZHANG Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5:725-738. https://zhanggroup.org/I-TASSER/
- SANTOS M, YASUIKE M, HIRONO I, AOKI T. 2006. The granulocyte colony-stimulating factors (CSF3s) of fish and chicken. Immunogenetics, 58(5-6):422-432. DOI: 10.1007/s00251-006-0106-5
- SANTOS M, YASUIKE M, HIRONO I, AOKI T. 2007. A novel type-1 cytokine receptor from fish involved in the Janus kinase/Signal transducers and activators of transcription (Jak/STAT) signal pathway. Molecular Immunology, 44:3355-3363. DOI: 10.1016/ j.molimm.2007.02.018
- SHI Y. 2009. Serine/Threonine Phosphatases: mechanism through structure. Cell, 139. DOI:10.1016/ j.cell.2009.10.006
- SONG Y, LI C. 2014. Shrimp immune system special focus on Penaeidin. Journal of Marine Science and Technology, 22(1):1-8. DOI: 10.6119/JMST-013-0813-1
- SUN L, SU Y, ZHAO Y. et al. 2016. Crystal structure of major envelope protein VP24 from white spot syndrome virus. Scientific Reports 6, 32309. https:// doi.org/10.1038/srep32309
- TERRAK M, KERFF F, LANGSETMO K, TAO T, DOMINGUEZ R. 2004. Structural basis of protein phosphatase 1 regulation. Nature, 429:780-784. DOI:10.1038/nature02582

- THE UNIPROT CONSORTIUM. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research, 49:D1.
- TAMURA K, STECHER G, PETERSON D, FILIPSKI A, KUMAR S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution, 30:2725-2729. DOI:10.1093/molbev/mst197
- VAFIADAKI E, ARVANITIS DA, SANOUDOU D, KRANIAS EG. 2013. Identification of a protein phosphatase-1/phospholamban complex that is regulated by cAMP-dependent phosphorylation. PLoS ONE, 8(11).e80867. DOI:10.1371/ journal.pone.0080867
- VERBINNEN I, FERREIRA M, BOLLEN M. 2017. Biogenesis and activity regulation of protein phosphatase 1. Biochemical Society Transactions, 45:89-99. DOI: 10.1042/BST-2016-0154C_COR
- WANG H, KO T, WANG A, LO C. 2014. 3X0T:Crystal structure of PirA. Deposited into the PDB 10/22/2014. DOI: 10.2210/pdb3X0T/pdb
- ZHAO C, WANG Y, FU M, YANG K, QIU L. 2016. Molecular cloning, expression and functional analysis of three subunits of protein phosphatase 2A (PP2A) from black tiger shrimps (*Penaeus monodon*). Comparative Biochemistry and Physiology, Part B, 204:77-89. DOI:10.1016/j.cbpb.2016.11.011

Supplemental Figure 1. Illustration of the derived full-length $PmSTPP1-\alpha$ gene sequence. The position of the original cDNA fragment and the gene product were indicated in the generated full-length gene.



Supplemental Figure 2. Complete Multiple Alignment of PmSTPP1-α with reference PP1 amino acid sequences. Secondary structure elements were represented as cylinders for the α -helices and as arrows for the β -strands as indicated above the alignment. The unordered N and C-termini were marked with question mark (?). Amino acids similar with PmSTPP1-a sequence were represented by dots (...); gaps on the aa positions were designated by dashes (---); conserved aa sequence motifs and other conserved sequences (unique for PPP family) were denoted with filled box; conserved active site or metalcoordinating residues were marked with inverted unfilled triangle; residues responsible in binding targeting protein or substrate that bind in PP1C (a regulatory protein bound, spinophilin, also binds to these residues) were denoted with unfilled box, residues that form the RVxF binding pocket were denoted with broken-line unfilled box, residues critical for associating molecular toxins with STPP1-α [within 5 A° of the bound toxin in each of the molecular toxin structures determined as described in the paper of Peti et al. (2013)] were denoted with broken-line filled box, aa residues that are distinct feature to STPP1 together with PP2A and PP2B subfamilies were marked with asterisk (*); position of the twenty three (23) aa residues in between of the first and second conserved motifs (which is the same position and number of residues with the human STPP1- α) were marked with a line on top; STPP1- α with similar aa sequences STPP1- γ and different with STPP1- β were marked with filled circle; STPP1- α with similar aa sequences STPP1- β and different with STPP1- γ were marked with filled diamond; STPP1- α with unique aa sequences from STPP1- γ and STPP1- β were marked with unfilled diamond; aa residues that separated the acidic and C-terminal groove [noted as a protruding loop which is from N271 – D277 in the paper of Terrak et al. (2004)] were marked with unfilled circle.

	??????	A1	0	A2	0
PmSTPP1-a	1: MAETDKLN	-IDSIIAR	LLEVRGSRPG	KNVOLTENE IRGLCLKSR	EIFLS 48
Daphnia pulex hypothetical protein	1:				48
Xenopus laevis STPP1-alpha	1: .GDGE				48
Homo sapiens STPP1-alpha	3:DSE	- L G			48
Rattus norvegicus STPP1-alpha	3:DSE	- L G	Q		48
Taeniopygia guttata STPP1-alpha	1: D . E	- L S			48
Bos mutus STPP1-alpha	80: P. PATE. HGHLQQG	GHAALWG.AGSGRR	G T G T T . A . Q		144
Myotis davidii STPP1-alpha	1:		P.Q		31
Rattus norvegicus STPP1-gamma	1: DI		K	Q	48
Mus musculus STPP1-gamma	1: DI				48
Gallus gallus STPP1-gamma	1: DI		K		65
Homo sapiens STPP1-gamma	1: DL				65
Litopenaeus vannamei STPP	1: - M A D T E . D	NL.S		. s m a. v	Q 65
Carassius auratus STPP1-beta	1: - M A E G D . D	- V L. S		. I M A . V V	47
Mus musculus STPP1-beta	1: - M A D G E	-VL.T		. I M A . V I	47
Rattus norvegicus STPP1-beta	1: - M A D G E	-VL.T		. I M A . V I	47
Homo sapiens STPP1-beta	1: - M A D G E	-VL.T		. I M A . V I	47
Xenopus tropicalis STPP1-beta	1: - M A D G E	- V L. S		. I M A . V I	47
Zootermopsis nevadensis STPP1-beta	1: -MADID	- V L . Q		. S M A . V	Q 47
Danio rerio STPP1-beta	1: - M A E G E	-VL.S		. I M A . V I	47
		A2			
	<u>B1</u>	A3	B 3	A4	0
		A3	B3		0
PmSTPP1-a	49: QPILLELEAPLKIC	A3 ∇ ∇ * G D I H G Q Y Y DL L RL F	B3	✓ SDYVD[k]GKQSLETICLLLAX) Y K I K 113
PmSTPP1-a Daphnia pulex hypothetical protein	49: QPILLELEAPLKIC	✓ ▽ G D I H G Q Y Y D L L R L F 	B3) Y K I K 113
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha	49: QPILLELEAPLKIC 49:	✓ ✓ ✓ * G D I H G Q Y Y D L L R L F	B3		() YKIK 113 113 113
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha	B1 B2 49: QPILLELEAPLKIC 49:	✓ ✓ * Ø ✓ ✓ * Ø ✓ ✓ *	B3 ** * * * YG G F P P E S N Y L F L	✓ SDYVD ^R GKQSLETICLLLA) YKIK 113 113 113 113
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha	B1 B2 49: QPILLELEAPLKIC 49: 49: 49: 49: 49: 49: 49:	✓ ✓ * G D I H G Q Y Y D L L R L F * · · · · · · · · · · · · · · · · · · ·	B3 ** * * G G F P P S N Y L F L	✓ SDYVDRGKQSLETICLLLA) 7 K I K 113 113 113 113 113
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha	B1 B2 49: QPILLELEAPLKIC 49: 49: 49: 49: 49: 49: 49: 49: 49:	A3 ▼ ▼ G D I H G Q Y Y D L L R L F 	B3	✓ SDYVD ^K GKQSLETICLLLA) Y K I K 113 113 113 113 113 113
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha	B1 B2 49: QPILLELEAPLKIC 49:		B3	✓ SD Y V D ⁱ E ⁱ G K Q S L E T I C L L L A S	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha	B1 B2 49: QPILLELEAPLKIC 49:		B3	✓ SDYVD ^E G KQSLETICLLLA	() Y K I K 113 113 113 113 113 113 209 96
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma	B1 B2 49: QPILLELEAPLKIC 49:		B3	✓ SDYVD ^E G KQSLETICLLLA:	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma	B1 B2 49: Q PILLELEAPLKIC 49:	A3 ∇ ∇ G D I H G Q Y Y D L L R L F 	B3	✓ A4 > > > > > > > > > > > > > > > > > > >	() () () () () () () () () ()
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Rattus norvegicus STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Gallus gallus STPP1-gamma	B1 B2 49: Q PILLELEAPLKIC 49:	A3 ∇ ∇ G D I H G Q Y Y D L L R L F 	B3 ** * YGGFPPESNYLFL	✓ A4 ✓ OF A CONTRACT OF A CONTRAC	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Gallus gallus STPP1-gamma Homo sapiens STPP1-gamma	B1 B2 49: QPILLELEAPLKIC 49:		B3 ** * YGGFPPESNYLFL	✓ SD Y V D ^[R] G K Q S L E T I C L L L A S	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Gallus gallus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vannamei STPP Caraccine auratue STPP1-beta	B1 B2 49: QPILLELEAPLKIC 49:		B3 ** * YGGFPPESNYLFL	✓ Ø Y V D ^[R] G K Q S LETICLLLA S	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Gallus gallus STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta	B1 B2 49: QPILLELEAPLKIC 49:		B3 *** * G G F P P S N Y L F L 	✓ ✓ > D Y V D ¹ K ¹ ₀ G K Q SLETICLLLA S	() () () () () () () () () ()
PmSTFP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Dos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta	B1 B2 49: QPILLELEAPLKIC 49:		B3 *** * G G F P P S N Y L F L *** *** *** *** *** *** *** *	✓ ✓ A4 ✓ SD Y V D ^K G K Q SLETICLLLAS	() (K I K 113 113 113 113 113 113 113 113 113 113 113 113 113 113 112 112 112
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Litopenaeus vanamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Rattus norvegicus STPP1-beta	B1 B2 49: Q PILLELEAPLKIC 49:		B3 ** * * YGGFPPESNYLFL YGGFPPESNYLFL A A A	✓	() () () () () () () () () ()
PmSTPP1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma dallus gallus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vanamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Homo sapiens STPP1-beta Katous stPP1-beta Kenopus tropicalis STPP1-beta	B1 B2 49: Q PILLELEAPLKIC 49:		B3 ** * * * G G F P P E S N Y L F L 	✓ SD YV D ¹² ,6 K Q SLETICLLLA 	(K I K 113 113 113 113 113 113 113 113 113 113 113 113 113 113 112 112 112 112 112
PmSTFF1-a Daphnia pulex hypothetical protein Xenopus laevis STFP1-alpha Homo sapiens STFP1-alpha Rattus norvegicus STFP1-alpha Taeniopygia guttata STFP1-alpha Bos mutus STFP1-alpha Rattus norvegicus STFP1-gamma Mus musculus STFP1-gamma Gallus gallus STFP1-gamma Litopenaeus vannamei STFP Carassius auratus STFP1-beta Mus musculus STFP1-beta Rattus norvegicus STFP1-beta Homo sapiens STFP1-beta Kattus norvegicus STFP1-beta Homo sapiens STFP1-beta Zootermopsis nevadensis STFP1-beta	B1 B2 49: Q PILLELEAPLKIC 49:		B3 ** * YGGFPPESNYLFL YGGFPPESNYLFL A A A A A A A A A A	✓ SD YV D ^I ² ³ ₁ G K Q SLETICLLLA S	() () () () () () () () () ()
PmSTFF1-a Daphnia pulex hypothetical protein Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Des mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Homo sapiens STPP1-beta Kattus norvegicus STPP1-beta Kattus norvegicus STPP1-beta Zootermopsis nevadensis STPP1-beta Danio rerio STPP1-beta	B1 B2 49: QPILLELEAPLKIC 49:		B3 * * * * * YG G FPPESNYLFL 	✓ SD Y V D ^[2] G K Q SLETICLLLA S	() (K I K 113 113 113 113 113 113 113 113 113 113 113 113 113 113 113 112 112 112 112 112 112

Supplemental Figure 2. Continuation.

			B4			5 0-				A6				<u>_</u>		В	5			
			/		<u> </u>		\square										~ ⁄			
					a	a									(-		∇			_
PmS1PP1-a	114:	YPEN	FFLL	RGNH	CASI	INRIY	GFYDI	CKR	RYNI	KLWK	TFTD	CFNC	LPVA	AIV	DEP	(IF)	ссн	GGL	SP 17	8
Daphnia pulex hypothetical protein	n114:					i · · · li	· · · ·	• • •				• • • •	• • • •	• • •	· · ·	···	••••	· · ·	17	8
Xenopus laevis SIPPI-alpha	114:			· · · · j	l i i i	1 ji		• • •					· · · · ·		· · · į·	· · ·	••••	•••	17	8
Homo sapiens SIPPI-alpha	114:		• • • •		1	fi i i i	(· · · · ·	• • •				• • • •		• • •	· · ¦·	· · ·¦	••••		17	8
Rattus norvegicus STPP1-alpha	114:			· · · · ·	l de la composition de la comp	l ji							· · I ·		· · · į·	· · ·	••••	•••	17	8
Taeniopygia guttata STPP1-alpha	114:			· · · · ·	(* * * *	i · · · i	[· · · · ·	• • •							·	···	••••	· · ·	17	8
Bos mutus STPP1-alpha	210:				1.1.	1							· · I ·		· · · ·	· · ·	••••	· · ·	27	4
Myotis davidii STPP1-alpha	97:			· · · · ·	i de la c	(\cdots)	[· · · · ·	• • •				• • • •	I.		· · · ¦·	· · ·i	••••	• • •	16	1
Rattus norvegicus STPP1-gamma	114:				1	1						• • • •	I.		· · !·	· · ·	••••	•••	17	8
Mus musculus STPP1-gamma	114:			\cdot	i de la c	(mi	[· · · · ·	···i	••••	•••	17	8
Gallus gallus STPP1-gamma	114:					lb							I.		· · !·	• • •	••••	•••	17	8
Homo sapiens STPP1-gamma	114:		• • • •	\cdot \cdot \cdot \cdot	i de la c	i · · ·i·	[· · · · ·	• • •				• • • •	I .		· · i·	···	••••		17	8
Litopenaeus vannamei STPP	112:		• • • •	· · · · ·		i · · · ŀ	F.	. R .	G T				I .	I	· · [·	···	••••	•••	17	7
Carassius auratus STPP1-beta	112:			· · · · i	i de la c	i · · ·i·	[••••	• • •	. F				I.	· · ·	···	···	••••	•••	17	7
Mus musculus STPP1-beta	112:			· · · · ·		l	· · · ·	•••	. F				I.	• • •	· · !·	···	••••	•••	17	7
Rattus norvegicus STPP1-beta	112:			$\cdot \cdot \cdot \cdot \cdot$	i - i - i	(· · · þ	[••••	• • •	. F				I .		···	···	•••		17	7
Homo sapiens STPP1-beta	112:			· · · · ŀ	· · · · ·	{· · · ŀ	· · · ·		. F				I .		·	· · ·¦·	••••		17	7
Xenopus tropicalis STPP1-beta	112:			$\cdot \cdot \cdot \cdot $	i - i - i	i · · ·i·	[. F				I .		·	· · ·i	••••		17	7
Zootermopsis nevadensis STPP1-beta	a112:			· · · · ŀ	· • • • •	4	· · · ·						I .	I	·	···	•••		17	7
Danio rerio STPP1-beta	112:			· · · · [.	j j. g	jj.	j		. F				I .	I	· [.	<u></u> i			17	7
													N	_						
			A	7 (-			A8						- <u>B6</u>	\succ	A	9	0	[
			(A	7 ()-			- A8						- <u>B6</u>		A	9	0-	[
Dm C 1700 1 _ a	170.			7 ()-	וחדם		- <u>A8</u>		שחפח	DTMG	WGEN	DRGV			A	9	_)-			
PmSTPP1-a	179:	DLQS	(A Meqi	7)-	RPTD	VPDQG	A8	L L M S	DPDK	DTMG	WGEN	DRGV	- BG		A	9 A K F :)-		D L 24	3 9
PmSTPP1-a Daphnia pulex hypothetical protei	179: n179: 179:	DLQS	MEQI	7 ()- RRIMI	R P T D T	♥ ₽ D Q 0	A8		D P D K	DTMG 	WGEN	DRGV	- B6		A	9 A K F :)- L H K	H D F M T	D L 24	3339
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha	179: n179: 179:	D L Q S	MEQI V	7 ()- RRIMI L	R P T D T	V P D Q G	A8	L L M S	D P D K	DTMG .VL.	W G E N	D R G V	B6			9 A K F :)- LHK	H D F M L	D L 24 24 24	3 3 3 3 3
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Pattus provesjour STPP1-alpha	179: n179: 179: 179:	D L Q S	MEQI V	7 ()- RRIMI L	R P T D 1	V P D Q G	A8	L L W S	D P D K	 .VL. .VQ.	W G E N	D R G V	B6			9 AKF:	••• LHK	H D F I MI LI LI	D L 24 24 24 24	3 3 3 3 4
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Teapiopuja guttata STPP1-alpha	179: n179: 179: 179: 179:	D L Q S	MEQI V	7 ()- RRIMI L	R P T D T	V P D Q G	A8		D P D K	. VL. . VQ. . VQ.	W G E N	D R G V	B6			9 AKF:	••• LHK	H D F . M . L . L . L	D L 2 4 2 4 2 4 2 4 2 4 2 4	3 3 3 3 3 3 3
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Poc putre STPP1-alpha	179: n179: 179: 179: 179: 179: 275:	DLQS	MEQI	7 ()- RRIMI L L	R P T D 1	V P D Q C			 D P D K	. DIMG . VL . . VQ . . VQ . . VQ .	W G E N	D R G V	B6			9 AKF:			D L 2 4 24 24 24 24 24 24	3 3 3 3 3 3 0
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Muntia dunidii STP1 alpha	179: n179: 179: 179: 179: 179: 275:	DLQS	MEQI	7 () RRIMI L	R P T D 1	VPDQC	A8		D P D K	 . VL . . VQ . . VQ . . VQ . . VQ . . VQ .	W G E N	DRGV	B6			9 AKF:			D L 2 4 24 24 24 24 24 24 24 24	3 3 3 3 3 9 4
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Datus energia us STPP1-alpha	179: n179: 179: 179: 179: 179: 275: 162: 179:	DLQS	MEQI	7)- RRIMI L L	R P T D 1	VPDQO	A8		D P D K	:DIMG .VL. .VQ. .VQ. .VQ. .VQ. .VQ. .VQ.	W G E N	DRGV	B6			9 AKF:			D I 2 4 24 24 24 24 24 24 24 24 23	3 3 3 3 3 9 6 9
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Myotus convente STP11 gamma	179: n179: 179: 179: 179: 179: 275: 162: 179:	D L Q S	MEQI	7)- RRIMI	R P T D 1	VPDQC	A8		D P D K	:DIMG .VL. .VQ. .VQ. .VQ. .VQ. .VQ. .VQ.	W G E N	D R G V	B6			9 AKF:			D L 2 4 2 4	3 3 3 3 3 9 6 3 9
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Caller alber STPP1-gamma	179: n179: 179: 179: 179: 275: 162: 179: 179:	D L Q S	MEQI	7 ()- R R I M I L 	R P T D 1	VPDQC	A8	L L W S	D P D K	 . VL . . VQ . . VQ . . VQ . . VQ . . VQ . . VL . . VL .	W G E N	DRGV	B6			9 AKF:			D L 2 4 2 4	3 3 3 3 3 9 6 3 3 9
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma	179: n179: 179: 179: 179: 275: 162: 179: 179: 179:	D L Q S	MEQI	7 ()- R R I M I L 	RPTD	VPDQC			D P D K	 . VL . . VQ . . VQ . . VQ . . VQ . . VQ . . VL . . VL .	W G E N	DRGV	B6			9 AKF:			D I 24 24 24 24 24 24 24 24 24 24 24 24	333333963330
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Homo sapiens STPP1-gamma	179: n179: 179: 179: 179: 275: 162: 179: 179: 179: 179: 179:	D L Q S	MEQI	7 ()- R R I M I L 	RPTD				D P D K	<pre>CDIMG</pre>	W G E N	D R G V	B6			9 A K F :			D I 24 24 24 24 24 24 24 24 24 24 24	3 3 3 3 3 9 6 3 3 3 9
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-ganma Mus musculus STPP1-ganma Gallus gallus STPP1-ganma Homo sapiens STPP1-ganma Litopenaeus vannamei STPP	179: n179: 179: 179: 179: 275: 162: 179: 179: 179: 179: 179: 178:	D L Q S	MEQI	7 ()- RRIMI L	R P T D 1				D P D K	<pre>CDIMG</pre>	W G E N	D R G V				9 A K F :			D L 2 4 2 4	33333396333390
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta	179: n179: 179: 179: 179: 179: 179: 179: 179:	D L Q S	MEQI	7 ()- R R I M I L 	R P T D 1				D P D K	DIMG .VL. .VQ. .VQ. .VQ. .VQ. .VL. .VL.	W G E N	DRGV	B6			9 A K F :	()-		D L 2 4 2 4 	3 3 3 3 3 9 6 3 3 3 9 2 0
PmSTFP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta	179: n179: 179: 179: 179: 275: 162: 179: 179: 179: 179: 179: 178: 178:	DLQS	MEQI	7 ()- R R I M I 	RPTDT					 . VL . . VQ . . VQ . . VQ . . VQ . . VQ . . VL . . VL . . VL . . VL . . VL . . VQ .	W G E N	DRGV	B6		A 1 2 2 2 4 2 2 4 2 2 4 2 4 2 4 2 4 2 4 2	9 A K F : 			D L 2 4 . 2	3333396333922
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Gallus gallus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta	179: n179: 179: 179: 275: 162: 179: 179: 179: 179: 179: 178: 178: 178:	D L Q S	MEQI	7 ()- R R I M I L L 	R P T D T				D P D R	CDIMG .VQ. .VQ. .VQ. .VQ. .VQ. .VQ. .V	W G E N	D R G V	B6		A 1 V V 1 1 V V 2 1 V V V 2 1 V V V V V V V V V V V V V V V V V V V	9 A K F : 	() LHK 	H D F M L L L L L L L L L L L L L L L L L L	$\begin{array}{c} \hline D \ L \\ 2 \ 2 \ 2 \ 2 \ 4 \ 2 \ 2 \ 4 \ 2 \ 2 \$	3 3 3 3 3 9 6 3 3 3 9 2 2 2
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Gallus gallus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Homo sapiens STPP1-beta	179: n179: 179: 179: 179: 275: 162: 179: 179: 179: 179: 178: 178: 178: 178:	D L Q S	MEQI	7 ()- R R I M I L 	R P T D 1			L L/W 5		DIMG 	W G E N	D R G V	Ben Stranger (1997)		A 1 V V : 1 V V : 2 V 2 : 2 2 : (2 2 : (9 A K F : 	() ()	H D F M L L L L L L L L L L L L L L L L L L	D L 24	333339633392222
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Taeniopygia guttata STPP1-alpha Bos mutus STPP1-alpha Myotis davidii STPP1-ganma Mus musculus STPP1-ganma Gallus gallus STPP1-ganma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Homo sapiens STPP1-beta Rattus norvegicus STPP1-beta Kano sapiens STP1-beta Kano sapiens STP1-beta	179: n179: 179: 179: 179: 275: 162: 179: 179: 179: 179: 178: 178: 178: 178: 178: 178:	D L Q S	MEQI	7 ()- R R I M I L L 						DIMG 	W G E N	DRGV	B6 SFTI		A 1 V V : 1 V V : 2 V V V : 2 V V V : 2 V V V · 2 V ·	9 A K F : 	() ()		D L 24	3 3 3 3 3 9 6 3 3 3 9 2 2 2 2 2 2
PmSTPP1-a Daphnia pulex hypothetical protei Xenopus laevis STPP1-alpha Homo sapiens STPP1-alpha Rattus norvegicus STPP1-alpha Dos mutus STPP1-alpha Myotis davidii STPP1-alpha Rattus norvegicus STPP1-gamma Mus musculus STPP1-gamma Homo sapiens STPP1-gamma Litopenaeus vannamei STPP Carassius auratus STPP1-beta Mus musculus STPP1-beta Rattus norvegicus STPP1-beta Rattus norvegicus STPP1-beta Homo sapiens STPP1-beta Rattus norvegicus STPP1-beta Xenopus tropicalis STPP1-beta Zootermopsis nevadensis STPP1-beta	179: n179: 179: 179: 275: 162: 179: 179: 179: 179: 178: 178: 178: 178: 178: 178: 178:	DLQS:	MEQI	7 ()- R R I M I L L 	RPTD	V P D Q C				D T M G . V Q . . V Q .	W G E X	DRGV	B6 SFTI 			9 AKF:			D L 24	3 3 3 3 3 9 6 3 3 3 9 2 2 2 2 2 2 2

Zootermopsis nevadensis STPP1-beta 308: G---M.S...S.Q.-----NPT... 325 87 6.06 Danio rerio STPP1-beta 308: G---V.S......T----A.AP..R 327 87 6.06

Supplemental Figure 2. Continuation.

	BZ	B8	B9	<u> </u>		<u>B10</u>		B11		??	, , , , , , , , , , , , , , , , , , , ,
	5⁄			~	~		r		\neg	•	
PmS TPP 1-a	* 244: ICRA	₩ovve dgye!Fi	FAKIROLVTI	FSAPNITICG	EFD	NAGAMMSV	DET	LMCISIFIC	DILKP	o O	KFPYGG 308
Daphnia pulex hypothetical protei	n244:										
Xenopus laevis STPP1-alpha	244:		i.i			s	s	. .			L.A 308
Homo sapiens STPP1-alpha	244:									N	.GKQ308
Rattus norvegicus STPP1-alpha	244:		!.!				!	.!.		N	.GKQ 308
Taeniopygia guttata STPP1-alpha	244:										.GKQ308
Bos mutus STPP1-alpha	340:		!.!							N	.GKQ404
Mvotis davidii STPP1-alpha	227:		.					!.!.!		N	.GKQ291
Rattus norvegicus STPP1-gamma	244:		¦.i					. .		. E	. P 304
Mus musculus STPP1-gamma	244:		.				j	. .		. E	. P 304
Gallus gallus STPP1-gamma	244:		.!					. .		.е	. P 304
Homo sapiens STPP1-gamma	244:						j	. .		. E	. P 304
Litopenaeus vannamei STPP								i-¦-!-			
Carassius auratus STPP1-beta	243:					G	!	. .		SE	AKYQY. 307
Mus musculus STPP1-beta	243:					G				S E	AKYQY. 307
Rattus norvegicus STPP1-beta	243:		!.!			G		. .		SE	AKYQY. 307
Homo sapiens STPP1-beta	243:					G				SE	AKYQY. 307
Xenopus tropicalis STPP1-beta	243:		!.!			G		. .		SE	AKYQY. 307
Zootermopsis nevadensis STPP1-bet	a243:		.			G				S E	AKYQYS 307
Danio rerio STPP1-beta	243:		.!			G	s	. . .		S E	AKYQYS 307
	_							()			
	?	3333333333	??????????????????????????????????????	?????							
					(%)	Ia					
PmSTPP1-a	309: L	- NTGRPVTPPR	GAANO	KNKKK 329	(-)	6.75					
Daphnia pulex hypothetical protei	n 309:	M	. GPQ.KQN	.G.N. 332	98	6.75					
Xenopus laevis STPP1-alpha	309: V	os	N K N	. Q S 326	92	6.25					
Homo sapiens STPP1-alpha	309: FSGL		N S A	.A 330	93	6.25					
Rattus norvegicus STPP1-alpha	309: FSGL	N P G I	N S A	.A 330	93	6.25					
Taeniopygia guttata STPP1-alpha	309: FSGL	N P A	N S A	. A . A S 331	94	5.99					
Bos mutus STPP1-alpha	405: FSGL	N P G I	N S A	.A 426	94	8.57					
Myotis davidii STPP1-alpha	292: FSGL	N P G I	N S A	.A 313	94	6.75					
Rattus norvegicus STPP1-gamma	305:	AT	M I T	.QA 323	93	6.5					
Mus musculus STPP1-gamma	305:	AT	M I T	.QA 323	93	6.5					
Gallus gallus STPP1-gamma	305:	AS	S M I T	.QA 323	92	6.5					
Homo sapiens STPP1-gamma	305:	A T	M I T	.QA 323	93	6.5					
Litopenaeus vannamei STPP					89	5.37					
Carassius auratus STPP1-beta	308: G	м.s	T A T	PPR 327	88	6.06					
Mus musculus STPP1-beta	308: G	L.S	T A N	PP R 327	89	6.06					
Rattus norvegicus STPP1-beta	308: G	L.S	t A N	PPR 327	89	6.06					
Homo sapiens STPP1-beta	308: G	L . S	t A N	PPR 327	89	6.06					
Xenopus tropicalis STPP1-beta	308: G	L.S	t A N	PPR 327	89	6.06					
Contraction of the second s	- 200. 0			D.T. 005	0.7	6 06					