The evolutionary strata of DARPP-32 tail implicates hierarchical functional expansion in higher vertebrates

CHOONG YONG UNG and TEOW CHONG TEOH*

Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

**Corresponding author (Fax, +603-79674178; Email, ttchong@um.edu.my)*

DARPP-32 (dopamine and adenosine 3′,5′-monophosphate-regulated phosphoprotein of 32 kDa), which belongs to *PPP1R1* gene family, is known to act as an important integrator in dopamine-mediated neurotransmission via the inhibition of protein phosphatase-1 (PP1). Besides its neuronal roles, this protein also behaves as a key player in pathological and pharmacological aspects. Use of bioinformatics and phylogenetics approaches to further characterize the molecular features of DARPP-32 can guide future works. Predicted phosphorylation sites on DARPP-32 show conservation across vertebrates. Phylogenetics analysis indicates evolutionary strata of phosphorylation site acquisition at the C-terminus, suggesting functional expansion of DARPP-32, where more diverse signalling cues may involve in regulating DARPP-32 in inhibiting PP1 activity. Moreover, both phylogenetics and synteny analyses suggest *de novo* origination of *PPP1R1* gene family via chromosomal rearrangement and exonization.

[Ung CY and Teoh TC 2014 The evolutionary strata of DARPP-32 tail implicates hierarchical functional expansion in higher vertebrates. *J. Biosci.* **39** 493–504] **DOI** 10.1007/s12038-014-9438-8

1. Introduction

Protein phosphorylation is one of the most important post-translational events that regulate myriad of biological processes such as cell division, cell differentiation, metabolism and modulation of signal transduction pathways (Ubersax and Ferrell 2007). It is common that one protein usually harbours many phosphorylation sites for a repertoire of serine/threonine or tyrosine kinases within a cell where the phosphorylation states for a given protein is combinatorial in nature. Thus, maintenance and regulation of phosphorylation states of proteome within a cell is crucial for its proper biological functions (Olsen *et al.* 2010). The maintenance of phosphorylation states is mainly mediated via the interplay between kinases and phosphatases. However, a mammalian genome encodes much fewer phosphatase genes than kinase genes (Moorhead *et al.* 2007). This implicates that a phosphatase is much versatile than a kinase. A ubiquitously expressed protein Ser/Thr phosphatase-1 (PP1) is one of such important phosphatases that coupled myriad of biological processes involved phosphorylation. Regulating the versatility of PP1 functions by a number of specific inhibitors at right time and right place is thus crucial (Bollen *et al.* 2010). One of such important PP1 inhibitor that mediates wide dopaminergic signalling pathways in the brain is DARPP-32.

DARPP-32 (dopamine and adenosine 3′,5′-monophosphate-regulated phosphoprotein of 32 kDa), which is also called PPP1R1B (protein phosphatase 1 regulatory subunit 1B), is an important mediator of dopamine signalling in the brain (Walaas *et al.* 1983). DARPP-32 is highly expressed in dopaminoceptive brain regions (Ouimet *et al.* 1984) but is also detected at other peripheral organs such as adrenal medulla, parathyroid cells, ciliary epithelium of the

Keywords. Conserved phosphorylation sites; DARPP-32; phylogenetics and synteny analyses; PP1 inhibitor

Supplementary materials pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/jun2014/supp/Ung.pdf*

Published online: 1 May 2014

eyes, and kidney, albeit at lower levels (Hemmings and Greengard 1986; Stone *et al.* 1986; Meister *et al.* 1989). The expression of DARPP-32 in many of these organs was shown to associate with dopamine signalling (Hemmings and Greengard 1986).

The phosphorylation states of DARPP-32 are affected by a number of neurotransmitters such as dopamine, serotonin, glutamate, GABA, as well as antipsychotic drugs and drugs of abuse, indicating its central role as an integrator to diverse neurotransmission inputs (Svenningsson *et al.* 2004, 2005). Also, reduced expression of DARPP-32 in the brain is associated to schizophrenic pathology (Albert *et al.* 2002). Upon activation of dopamine D1 receptor (D1R), DARPP-32 is phosphorylated by cAMP-dependent protein kinase (PKA) at threonine-34 (T34). Phosphorylation at T34 turns DARPP-32 into a potent inhibitor for PP1 (Hemmings *et al.* 1984). A short motif KKIQF at the N-terminal is a conserved PP1 docking motif. X-ray crystallography studies on other PP1 inhibitors had shown that this motif interacts with the surface of PP1 on the back in relative to the metal-containing active site at front (Goldberg *et al.* 1995; Egloff *et al.* 1997). Once the PP1 docking motif is docked to PP1, the phosphorylated T34 residue of DARPP-32 will be approximately positioned to the active site and inhibit dephosphorylation role of PP1 (Goldberg *et al.* 1995). Inhibition of PP1 by DARPP-32 increased phosphorylated states of neurotransmitter receptors and ion channels, thus enhancing the activities of synaptic function and plasticity (Svenningsson *et al.* 2004; Stipanovich *et al.* 2008).

Beside its main role in dopamine signalling, other recent studies also revealed wide functions of DARPP-32 such as binding to tra2-beta1 in regulating alternative splicing events (Benderska *et al.* 2010). Besides, DARPP-32 is also involved in a number of pathological cases. For instance, it has been reported that the truncated form of DARPP-32 (t-DARPP-32), where amino acid residues encoded by exon 1 are missing, is highly expressed in cancers (Belkhiri *et al.* 2008, Vangamudi *et al.* 2010). Thus, deregulated or loss of DARPP-32 activities can transform normal thyroid cells to tumours. In addition, overexpression of DARPP-32 enhanced interactions between EGFR and ERBB3 and promoted tumour resistance to antitumour drug gefitinib via increased phosphorylation of AKT (Zhu *et al.* 2011).

Although DARPP-32 is a key cellular regulator and is also a potential drug target, there is no comprehensive comparative analysis for this protein across other vertebrates. Furthermore, understanding DARPP-32 function from the evolutionary perspective can shed light on knowing better how this protein was originated and preserved throughout evolution and how its PP1 inhibitory function was associated to its current neuronal regulatory role.

2. Methods

2.1 *Sequence collection for PPP1R1 genes*

Human DARPP-32 amino acid sequence (ENSP00000377813) obtained from the ENSEMBL (*http://asia.ensembl.org/index.html*) was used as query sequence for BLASTP search for vertebrate and invertebrate genomes available in the ENSEMBL. The best BLAST hits were reciprocally BLAST to human genome to ensure the hits were DARPP-32 homologs. Human paralogs for DARPP-32, PPP1R1A (ENSP00000257905) and PPP1R1C (ENSP00000280295) were also obtained from ENSEMBL and were similarly BLAST to other vertebrate and invertebrate genomes to retrieve their respective homologous sequences. The paralogous sequences collected were also examined by using 'Paralogues' option provided by ENSEMBL. A total of 103 sequences of PPP1R1 protein sequences from 39 vertebrate species were collected (supplementary table S1 and supplementary data S1).

2.2 *Multiple sequence alignment*

We used MAFFT (*http://mafft.cbrc.jp/alignment/software/*), which is a rapid alignment algorithm based on fast Fourier transform (Katoh *et al.* 2002) using FFT-NS-i (slow; iterative refinement method) as an alignment option. We also used ClustalW (Chenna *et al.* 2003) implemented in MEGA 5.1 (Tamura *et al.* 2011) as another option for multiple sequence alignment. The multiple aligned results generated from both MAFFT (supplementary data S2) and ClustalW (supplementary data S3) were manually inspected using key phosphorylation residues such as T34, T75, and S102 as well as PP1 docking motif as phylogenetics keys to judge the quality of multiple sequence alignment results. We found that MAFFT generates better alignment results across 39 vertebrate species and thus results from MAFFT were used for our subsequent phylogenetics analyses.

2.3 *Construction of species and gene trees*

Species tree for the selected 39 vertebrate species was generated using the Common Tree from NCBI Taxonomy Browser (*http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi*). The generated species tree is given in supplementary figure S1. The input protein sequences were arranged according to their respective taxonomic order. The ordered multiple alignment sequences generated from MAFFT were then submitted to MEGA 5.1 to reconstruct gene tree using maximum likelihood method. Jones-Taylor-Thornton (JTT) model was used as substitution matrix with Gamma distributed rates among sites. Nearest-

neighbour-interchange (NNI) was used as heuristic method for maximum likelihood and a total of 100 steps of Bootstrap were implemented.

2.4 *Synteny analysis*

Synteny analysis for *DARPP-32*-containing chromosomal block was performed using ENSEMBL genome browser (*http://asia.ensembl.org/index.html*). The chromosomal loci for *DARPP-32* and its neighbor genes were recorded among 39 vertebrates surveyed and outgroups such as chordate and invertebrates whichever relevant. The results are provided in supplementary data S4.

2.5 *Prediction of phosphorylation sites*

GPS v2.1.2 (*http://gps.biocuckoo.org/*), which utilizes group-based prediction method (Zhou *et al.* 2004; Xue *et al.* 2005), was used to predict potential phosphorylation sites on human DARPP-32. Both Ser/Thr and tyrosine kinases were included and high threshold was selected. The resulting predicted phosphorylation sites and the respective kinases are given in supplementary table S2.

2.6 *Template-based secondary structure prediction*

We used YASARA (Krieger *et al.* 2002) for template-based secondary structure prediction which was achieved by running PSI-BLAST to create a target sequence profile and feeding it to the PSI-Pred secondary structure prediction algorithm (Jones 1999). DARPP-32 protein sequences for 39 vertebrate species whenever available were submitted to YASARA for secondary structure prediction.

3. Results and discussion

3.1 *Phylogenetics and synteny analyses suggests plausible origin of PPP1R1 gene family*

We collected amino acid sequences of DARPP-32 together with its paralogs PPP1R1A and PPP1R1C from diverse vertebrates and conducted phylogenetics analysis for these proteins. We used reported key residues such as T34 and motifs such as PP1 docking motif as phylogenetics keys to examine the alignment quality generated from several multiple alignment methods. We found that in our case, MAFFT produced alignment result correctly captured the conservation profiles of the phylogenetics keys used (supplementary data S2). Hence, alignment result from MAFFT was used for subsequent alignment analysis and reconstruction of gene tree.

We did not find any members of *PPP1R1* gene outside vertebrates. Our sequence search against ENSEMBL database indicated that all vertebrates possess at least one member of *PPP1R1* gene (supplementary data S1). Earlier work reported that no homologous sequence of DARPP-32 was found in jawless fish such as lamprey and invertebrates (Hemmings *et al.* 1992). Yger and Girault (2011) proposed that the proto-*PPP1R1* gene might arise in early vertebrate evolution, possibly during divergent of Gnathostomes that encompassed both bony and cartilaginous fish. Our sequence alignment result reveals conservation at the N-terminal regions for all members of *PPP1R1* genes but differ at C-terminal regions. Reconstructed gene tree using maximum likelihood method as shown in figure 1 suggests closer sequence similarity between PPP1R1A and PPP1R1C. This is also consistent with earlier analysis albeit smaller sequence coverage (Yger and Girault 2011).

Synteny analysis revealed conservation of blocks of chromosome across different species. Besides chromosomal block conservation, results from synteny analysis also can shed light on detecting chromosomal rearrangement as well as stabilization of gene order at the genome scale. Our synteny analysis result (figure 2) suggests stable syntenic chromosomal block containing *DARPP-32* gene among eutherians and a marsupial opossum. A number of genes adjacent to *DARPP-32* such as *STARD3*, *NEUROD2*, *CDK12*, *MED1*, *ERBB2*, *FBXL20*, and *PNMT* can be used as 'syntenic markers' to trace chromosomal rearrangement event that accompanied stabilization of '*DARPP-32* syntenic blocks'. No *DARPP-32* gene is detected outside vertebrates such as worms (*Caenorhabditis elegans*), insects (*Drosophila melanogaster*), and chordates (*Ciona savignyi*). However, some of the 'syntenic markers' for *DARPP-32* such as *CDK12*, *STARD3*, *FBXL20*, *ERBB2*, and *PNMT* show synteny in these organisms, suggesting common origin of these syntenic blocks in *DARPP-32*-containing vertebrates.

Based on our gene tree together with phylogenetics results from previous studies (Hemmings *et al.* 1992; Yger and Girault 2011), we propose that the proto-*PPP1R1* gene was originated after the Cambrian period around 542 million years ago (MYA) (figure 3) after first round (1R) and second round (2R) of whole genome duplications of common ancestor for vertebrates (Van de Peer *et al.* 2009). Our synteny results as shown in invertebrates (figure 2) suggests genome rearrangement events may have important consequence to the origin of proto-*PPP1R1* gene. As *STARD3* (StAR-related lipid transfer domain protein 3), which is a late endosomal integral membrane protein involved in cholesterol transport (Alpy and Tomasetto 2006), is always adjacent to *DARPP-32* gene in all vertebrates surveyed (figure 2), and because *DARPP-32* seems to be more primitive than *PPP1R1A* and

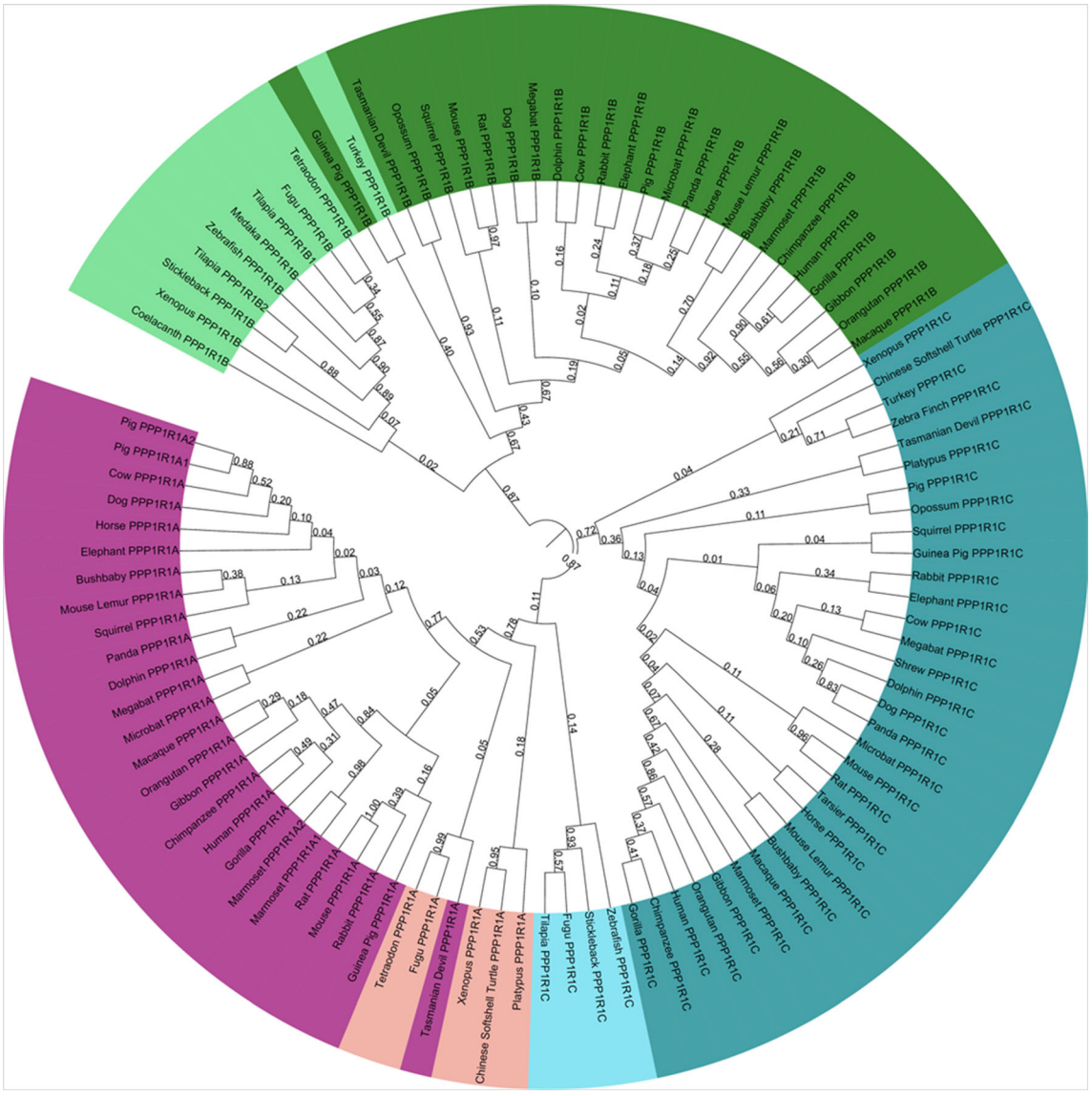

Figure 1. Gene tree of DARPP-32 and its paralogs across 39 vertebrate species generated from maximum likelihood method. DARPP-32, PPP1R1A, and PPP1R1C are shaded with green, purple, and blue respectively. Homologs possessing conserved C-terminal regions in respect to human and other higher vertebrate clades are shaded with their respective darker colours.

PPP1R1C (figure 1) (Yger and Girault 2011), we postulate that proto-*PPP1R1* gene may had been arisen via exonization event in adjacent to *STARD3* during Ordovician period around 510 MYA. The proto-*PPP1R1* gene was retained probably due to its acquired PP1 docking motif that conferred PP1 inhibitory role. This was followed by duplication of proto-*PPP1R1* gene into proto-*DARPP-32* and proto-*PPP1R1A/C* genes and the inhibitory roles of *PP1* by *PPP1R1* gene family were expanded hereafter. Since all vertebrates carry at least one *PPP1R1* gene family, the expansion of PP1 inhibitory function to *DARPP-32*, *PPP1R1A*, and *PPP1R1C* must had been preceded the split of fish and land vertebrates (figure 3).

3.2 *Gain of phosphorylation sites in higher vertebrates suggests extended repertoire of kinase-regulated functions of DARPP-32*

Since DARPP-32 is a phosphorylation-dependent inhibitor of PP1, there are a number of kinases known to phosphorylate DARPP-32 under different circumstances. For instance,

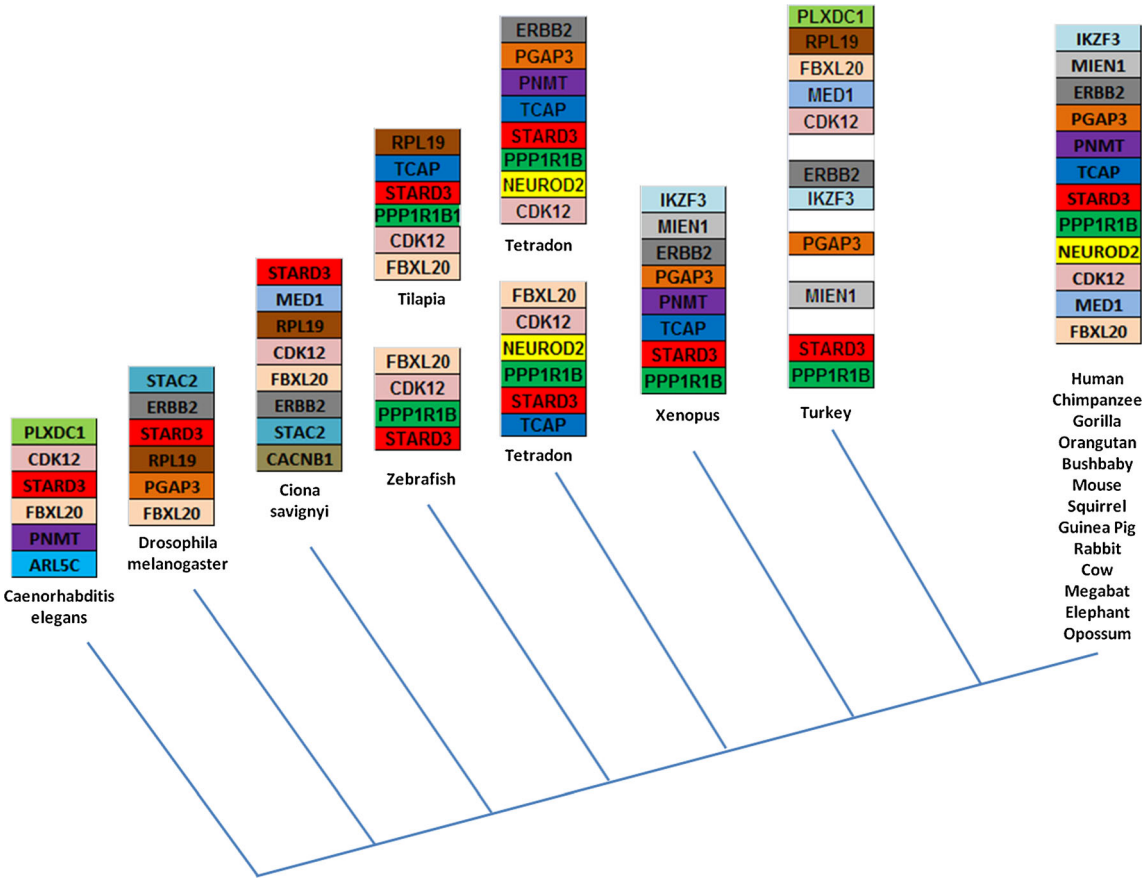

Figure 2. Synteny analysis for *DARPP-32*-containing chromosomal block. No *DARPP-32* (*PPP1R1B*) and other *PPP1R1* genes are found in chordate and invertebrates. The marsupial opossum and other eutherians up to primates show conserved synteny for *DARPP-32* adjacent genes with the same gene order. This conserved synteny block also can be observed even in fish such as zebrafish, tilapia, tetradon and the amphibia xenopus, although there are chromosomal rearrangements in these species. Turkey seems had undergone extensive chromosomal rearrangements where the *DARPP-32*-containing chromosomal block is fragmented to other chromosomes. *DARPP-32* is always associated with *STARD3* gene in all vertebrates surveyed whereas *NEUROD2* gene is associated to *DARPP-32* in tetradon and all mammals. Conserved synteny for *DARPP-32*-containing chromosomal blocks and *DARPP-32* adjacent genes in outgroups such as invertebrate hints the evolutionary origin of *DARPP-32*.

Cdk5 phosphorylates DARPP-32 on T75. The phosphorylated T75 inhibits phosphorylation at T34 by PKA, rescuing the inhibitory function of DARPP-32 on PP1 (Bibb *et al.* 1999). On the other hand, phosphorylation by CK2 increases the action of PKA on T34. Also, phosphorylation by CK1 on S137 decreases dephosphorylation of T34 by calcineurin and hence enhances the inhibitory role of DARPP-32 on PP1 (Fernandez *et al*. 2006). Thus, phosphorylations of DARPP-32 by PKA, CK1, and CK2 have antagonistic effects on PP1, while phosphorylation by Cdk5 has agonistic effect. It appears that DARPP-32 acts as a switch on PP1 function via a number of phosphorylation states.

We used group-based prediction system (GPS) (Zhou *et al.* 2004; Xue *et al.* 2005) to predict potential phosphorylation sites and also their respective kinases on human DARPP-32 where a total of 26 predicted phosphorylation sites are provided in supplementary table S2. We reason that there must be some degree of conservation for a given site if a predicted phosphorylation site is a *bona fide* site to a given kinase. Figure 4 summarizes the molecular features of human DARPP-32 and predicted phosphorylation sites and their respective kinases by GPS. Result from multiple sequence alignment suggests the predicted phosphorylation sites (T34, S45, T75, S102, and S137) are well conserved among vertebrates (figure 5), with gained and loss of certain sites in some species. Besides, eight additional putative phosphorylation sites (S12, S42, S46, Y74, S78, S92 (mouse),

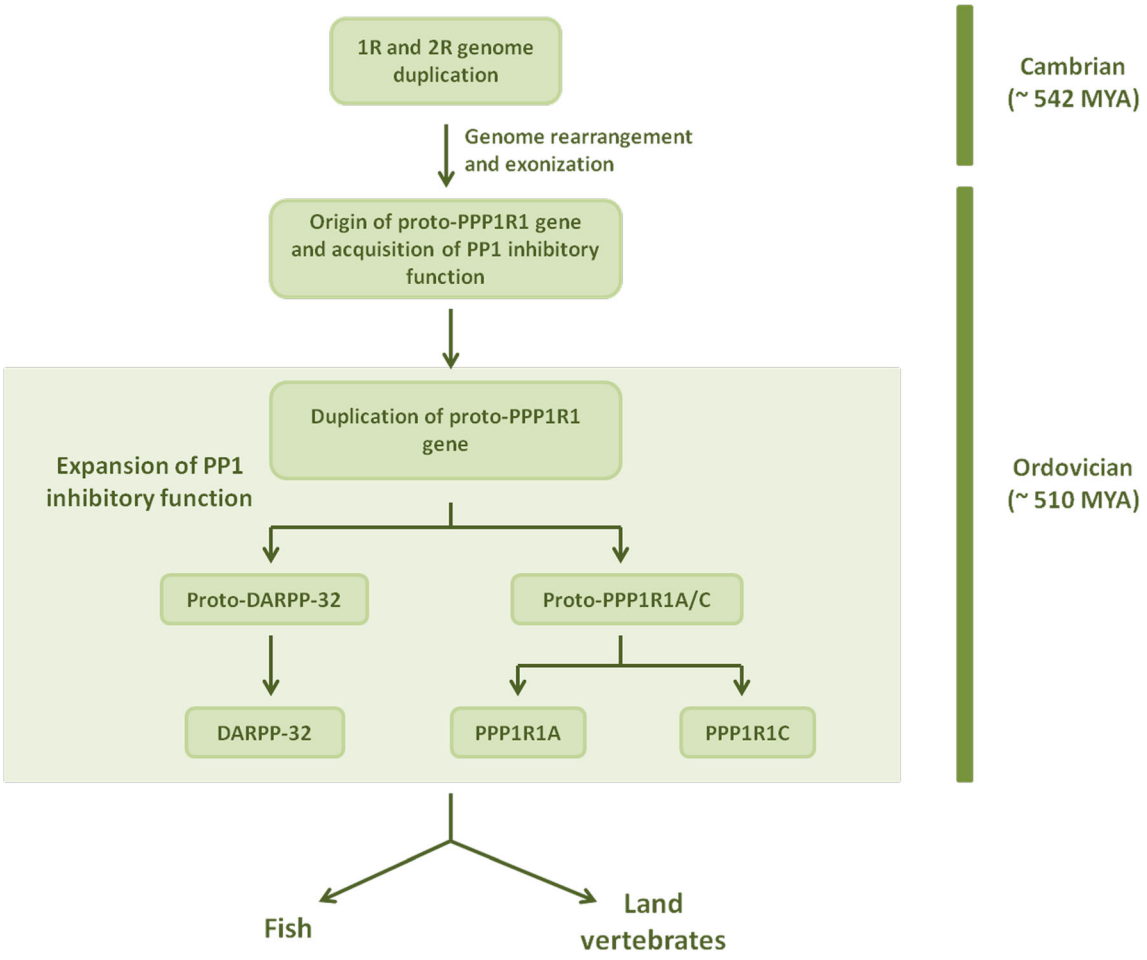

Figure 3. Proposed model for the origin of *PPP1R1* gene family inferred from gene tree and synteny analyses. The proto-*PPP1R1* gene was originated *de novo* from genome rearrangements followed by exonization events around the end of the Cambrian and the beginning of the Ordovivian period. The proto-*PPP1R1* gene may acquire its PP1 inhibitory role by chance and was selected by evolution. Further duplication of the proto-*PPP1R1* gene gave rise to all three paralogous genes, *PPP1R1A*, *DARPP-32*, and *PPP1R1C* detected in extant vertebrates resulting in expansion and diversification of PP1 inhibitory regulations. The paleontological time given is obtained from Van de Peer *et al.* (2009).

Y116, and S198) predicted by GPS (supplementary table S2) are also listed in PhosphoSitePlus (*http://www.phosphosite.org*). The molecular features at the N-terminus including PP1 docking motif and the critical phosphorylation site T34 are conserved in all *PPP1R1* family. Predicted together with known kinases on DARPP-32 shown in figure 4 implicates broad biological roles played by DARPP-32 such as cell cycle (ATR, AUR-B, CDK2/5/6 as kinases), metabolism (mTOR and GSK3A as kinases), cell-cell adhesion and cytoskeleton formation (FAK and ROCK1 as kinases), immunity (JNK, p38 as kinases), and growth factor-induced signal transductions (FGFR1, TYK2, ERK, MAP2K1/2, MAP3K11 as kinases).

It had been found that the activity of extracellular signal-regulated kinase (ERK) that is critical for long-term synaptic plasticity in the brain is also regulated by DARPP-32 (Valjent *et al.* 2005). Our phosphorylation site prediction also suggests that DARPP-32 can be a substrate for kinases such as MAP2K1, MAP2K2, MAP3K11, as well as ERK in the MAPK cascade. Indeed, DARPP-32 is necessary to maintain the activation of MAPK kinase (MEK)/ERK signalling. Conversely, DARPP-32 is depleted by MEK/ERK activation, leading to increased PP1 activity and consequently decreased MEK/MAPK signalling. This auto-regulatory loop represses DARPP-32 levels so that the activation of MEK is regulated through DARPP-32 (Chocarro-Calvo *et al.* 2012).

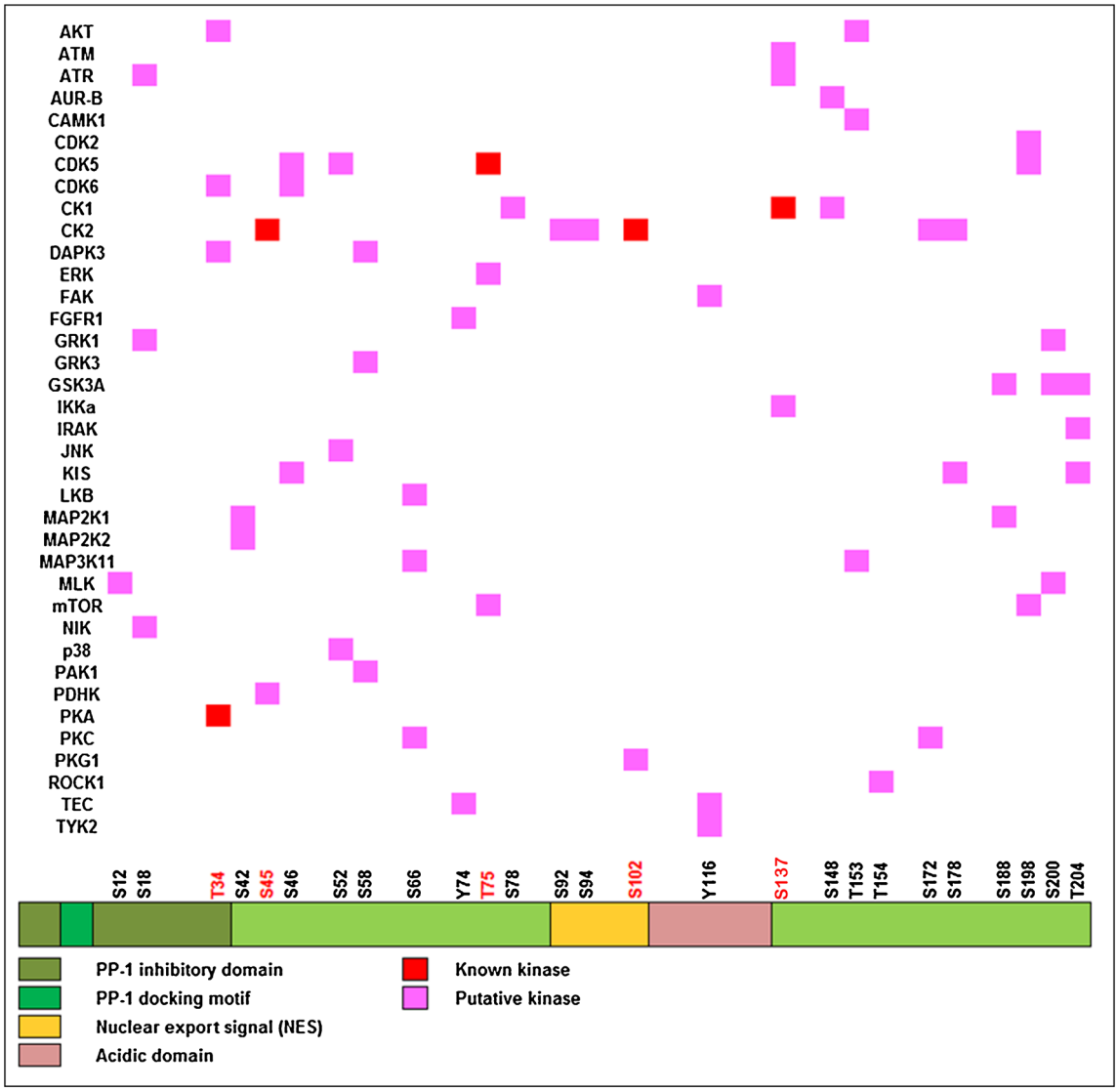

Figure 4. Molecular features of human DARPP-32 and predicted phosphorylation sites and their respective kinases by GPS. Known kinases to their phosphorylation sites are indicated in red.

3.3 *Conservation at regions with predicted transient secondary structures suggests their potential roles in potential protein-kinase interactions*

Both fluorescence and NMR spectroscopy analyses revealed that DARPP-32 has little secondary structure in free solution (Neyroz *et al.* 1993; Dancheck *et al.* 2008) suggesting the intrinsically disordered feature of this protein. However, there are regions showing propensities for α-helical content. Residues 22-29 at the N-terminal to the PKA phosphorylation site at T34, and between residues 92-109 which spans the CK2 phosphorylation site at S102 and the nuclear export signal sequence show some degrees of α-helical structure (Dancheck *et al.* 2008; Marsh *et al.* 2010). Furthermore, the central region from residues 80-115 can exist as a compact core (Dancheck *et al.* 2008). Since these regions are also highly conserved at least in mammals, the transient nature of these regions in adopting an α-helical conformation may play an important role in kinase-substrate recognition.

However, how the structural features determined from NMR studies also applicable to other vertebrates remain unclear. Here we utilized template-based secondary prediction to DARPP-32 across vertebrate lineages. Our prediction results agreed with NMR observation where residues 22–30

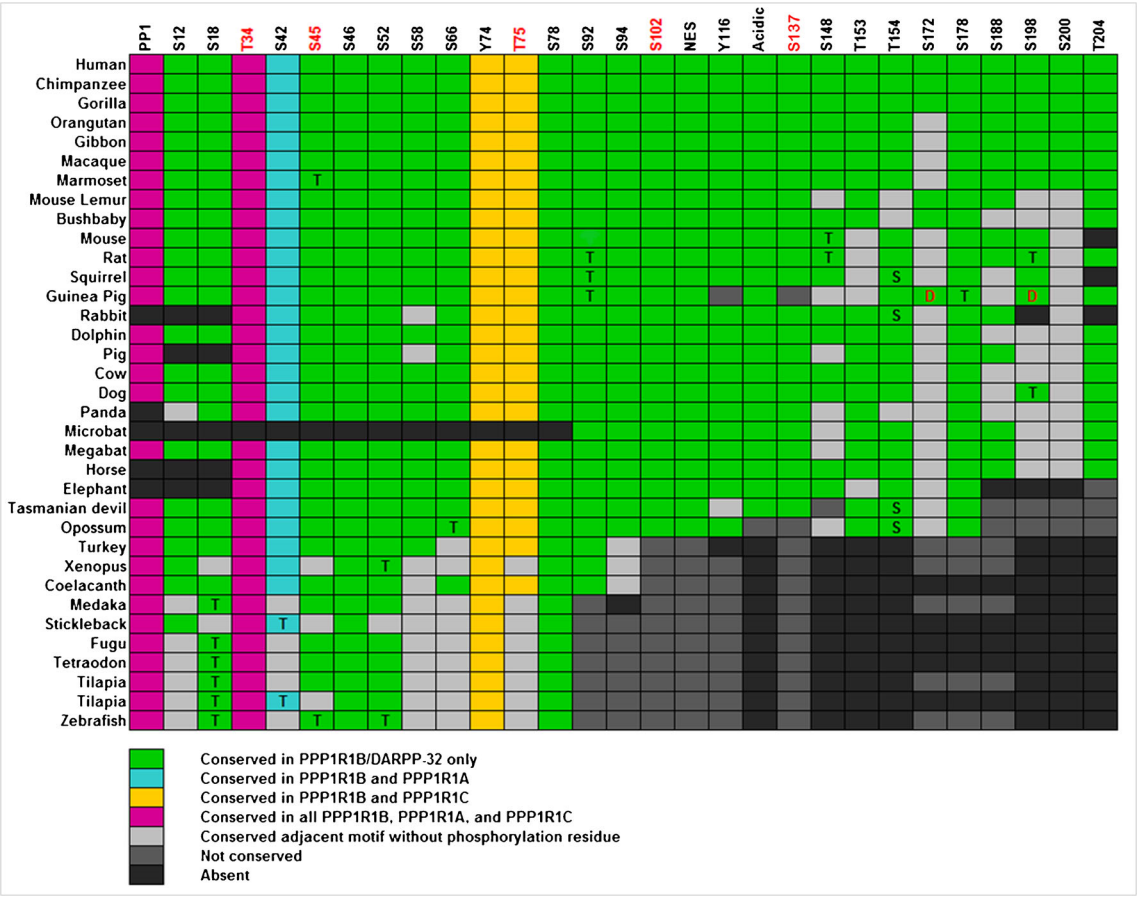

Figure 5. Conservative features of human DARPP-32 across vertebrate clades. Multiple alignment result from MAFFT was used to analyse the conservation features. Modes of conservation are depicted with different colour codes where equivalent phosphorylation sites are given in their respective boxes. Aspartic acid residues (D) which mimic the phosphorylation effect and equivalent residues (S or T) detected at some conservative phosphorylation sites are also illustrated. PP1: PP1 docking motif; NES: nuclear export signal motif; Acidic: acidic-rich motif.

(PRQVEMIRR) of human DARPP-32 adopt an α-helical structure (table 1). In addition, several sites such as residues 79-93 (LKAVQRIAESHLQSI), 103-110 (EEEDELGE) and 128-147 (DEEEEEEEDSQAEVLKVIRQ) are also predicted capable to adopt α-helical structures. These regions cover the central region from residues 80–115, where NMR study suggests its capability of forming a compact core (Dancheck *et al.* 2008). As shown in table 1, regions predicted to adopt a transient secondary structure are conserved across vertebrates, especially residues 22–30 that are highly conserved among vertebrates. This suggests that although DARPP-32 is an intrinsically disordered protein, there are still some regions showing high propensity to adopt transient secondary structures that may be crucial for molecular recognition in protein–protein interactions.

3.4 *Heterogeneity of DARPP-32 C-terminus tail implicates functional expansion via stratified evolution*

Early phylogenetic distribution of DARPP-32 using immunoblotting revealed that DARPP-32 was present in all mammalian species tested, with low levels detected in other peripheral tissues such as choroid plexus, parathyroid cells, adrenal chromaffin cells, posterior pituitary gland, pineal gland, and superior cervical sympathetic ganglion (Hemmings and Greengard 1986). The same study also showed that DARPP-32 was identified in dopaminoceptive brain regions in amniote vertebrates such as birds and reptiles but none was identified in the dopaminoceptive brain regions in anamniote vertebrates such as fish and amphibian. This suggests that DARPP-32 might had acquired its role in regulating PP1 activities in the brain from reptiles and onward.

Table 1. Regions of DARPP-32 predicted to adopt an α-helical structure by YASARA

Species	Predicted motifs with preferred α-helical structure
Human	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 103-110 (EEEDELGE); 128-147 (DEEEEEEEDSQAEVLKVIRQ)
Chimpanzee	22-30 (PRQVEMIRR); 79-92 (LKAVQRIAESHLQS); 129-146 (EEEEEEDSQAEVLKVIRQ)
Gorilla	22-30 (PRQVEMIRR); 90-104 (LKAVQRIAESHLQSI); 114-121 (EEEDELGE); 139-158 (DEEEEEEEDSQAEVLKVIRQ)
Orangutan	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 104-109 (EEDELG); 128-148 (DEEEEEEEDSQAEVLKVIRQS)
Gibbon	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 104-109 (EEDELG); 128-148 (DEEEEEEEDSQAEVLKVIRQS)
Macaque	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 104-109 (EEDELG); 128-148 (DEEEEEEEDSQAEVLKVIRQS)
Marmoset	22-30 (PRQVEMIRR); 79-92 (LKAVQRIAESHLQS); 106-109 (DELG); 129-148 (EEEEEEEDSQAEVLKVIRQS)
Mouse Lemur	22-29 (PQVEMIRR); 78-91 (LKAVQRIAESHLQS); 128-142 (EEEEEEEDSQAEILK)
Bushbaby	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 103-109 (EEEDELG); 128-143 (DEEEEEEEDSQAEILK)
Mouse	22-30 (PRQVEMIRR); 81-88 (AVQHLQTI); 123-136 (EEDEEEDSQAEVLK)
Rat	22-30 (PRQVEMIRR); 79-92 (LKAVQRIAESHLQT); 103-109 (EEEDELG); 129-143 (EEEDEEEDSQAEVLK)
Squirrel	22-31 (PRQVEMIRRR); 84-88 (HLQTI); 98-108 (EEEDELGELRE); 123-140 (EDEEEEEDSQAEVLKSIR); 173-180 (EDQLEDAA)
Guinea Pig	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQTI); 101-111 (EDSQAEVLKAI)
Rabbit	No preferred α-helical regions
Dolphin	22-30 (PRQVEMIRR); 89-104 (LKAVQRIAESHLQSIS); 136-151 (DEEEEEDEDSQAEILK)
Pig	53-68 (LKAVQRIAESHLQSIS); 77-86 (EEEDELGELR); 103-117 (EEEEEEEDSQAEVLK)
Cow	22-30 (PRQVEMIRR); 79-93 (LKAVQRIAESHLQSI); 105-109 (EDELG); 122-141 (EEEEEEDEEEEEDSQAEVLK)
Dog	22-30 (PRQVEMIRR); 79-92 (LKAVQRIAESHLQS); 102-110 (SEEEDELGE); 131-144 (DEEEEEDSQAEVLK); 179-185 (KEQVED)
Panda	13-19 (LPYQIRR); 68-83 (LKAVQRIAESHLQSIS); 92-98 (EEEDELG); 118-133 (EEEEEEEDSQAEVLK)
Microbat	No preferred α-helical regions
Megabat	22-31 (PRQVEMIRRR); 90-103 (LKAVQRIAESHLQS); 139-152 (EEEEEEDSQAEVLK)
Horse	51-67 (LKAVQRIAESHLQSIS); 76-81 (EEEDEL); 97-117 (EEEEDDEEEEEEEDSQAEVLK)
Elephant	62-78 (SLKAVQRIAESHLQSIS); 87-97 (EEEDELGELRA); 106-128 (EEEEEEEEEDEEEDSQAEVLKSS)
Tasmanian devil	22-30 (PRQVEMIRR); 79-91 (LKAVQRIAESHLQ); 120-142 (EEEEEEEEEEEEDDSSQAEIIK); 180-188 (EEKEEAVQ)
Opossum	22-31 (PRQVEMIRRR); 79-92 (LKAVQRIAESHLQS)
Turkey	No preferred α-helical regions
Xenopus	22-31 (PRAVEMIRRR); 77-90 (SLKAVQRIVQSHLQ); 130-134 (EAECS); 149-155 (LENLMHQ); 176-181 (EEQVAC)
Coelacanth	22-30 (PRAVEMIRR); 78-92 (SLKAVQRIVQAHLQS)
Medaka	No preferred α-helical regions
Stickleback	22-30 (PLQVEMIRR), 78-91 (SLKAVQRMALAHLA); 135-144 (KEEAEEREEE)
Fugu	22-30 (PRQVEMIRR); 79-93 (LKAVQKMAEAHMQNL); 115-120 (ATEVQA)
Tetraodon	No preferred α-helical regions
Tilapia	33-41 (PRQVEMIRR); 90-103 (LKAVQKMAEAHMQK); 168-175 (DEEEEDMK)
Tilapia	27-35 (PRQVEMIRR); 85-97 (LKAVQKMAQAHLS); 112-140 (EEEEQHDEERQQKASEESKAQRDQRGLLD); 152-159 (EDAKEREE)
Zebrafish	33-41 (PRQVEMIRR); 90-103 (LKAVQQMAEAQMQK); 157-185 (EEEEEEEEEEEEDEEKEEEEADDMAREES)

The conserved regions across a number of vertebrate species are coloured according to their respective region on DARPP-32.

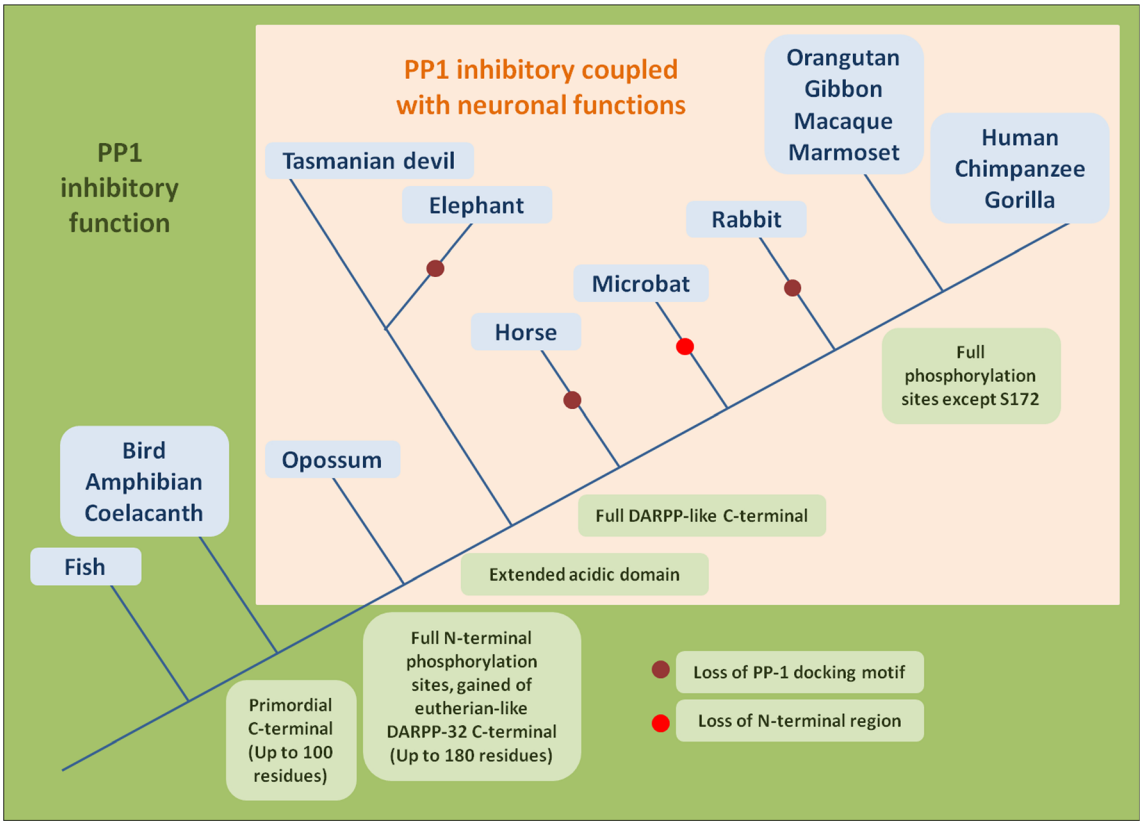

Figure 6. Evolutionary strata of DARPP-32 tail inferred from multiple alignment and conservation of DARPP-32 molecular features including phosphorylation sites. The C-termini of fish are dissimilar to those in higher vertebrates such as primates. Primordial C-terminus with up to ~100 residues in total protein length is detected in coelacanth, amphibian, and bird. This is followed by extension of eutherian-like C-terminus up to 180 residues in total length detected in marsupial opossum. Full extended acidic domains are detected from Tasmanian devil and elephant where full molecular features of human-like C-terminus are observed in vertebrates from horse onwards. There are gain and loss of human-like phosphorylation sites on some species where full set of human-like phosphorylation sites were eventually appeared in primates. Some species such as elephant, horse, and rabbit independently lost their PP1 docking motifs yet retain T34 sites. Microbat has large truncation at N-terminus encompassing many conserved critical residues including T34 and T75, probably due to change of lifestyle that disused DARPP-32.

One of the notable sequence features of DARPP-32 in higher vertebrates is the presence of a highly acidic region spanning 24 glutamic or aspartic acid residues in a total of approximately 32-residual motif. Although this acidic motif covers the phosphorylation site by CK2 at S102 (S97 for mouse) and a nuclear export signal sequence in regulating its nucleo-cytoplasmic localization, the exact role of this acidic motif is still unclear.

Our predicted phosphorylation sites on human DARPP-32 show heterogeneity and stratified hierarchy at the C-terminus (figure 5), indicating that these variable phosphorylation sites evolved at different stages during vertebrate evolution. Our alignment result suggests acquisition of primordial eutherian-like C-terminus that later evolved into a more specific function in regulating PP1 activities in the brain in coelacanth, amphibians, and birds, although it is still not clear whether these primordial C-termini have a direct role in modulating neuronal function in these species. Acquisition of eutherian or mammal-like C-termini with expanded phosphorylation sites implicated broader regulatory role of DARPP-32 in modulating PP1 activity in response to more diverse biological cues.

3.5 *Change of lifestyle during evolution may result in loss of DARPP-32 N-terminus*

Phylogenetics analysis of DARPP-32 in wide vertebrate lineages enables us to identify some anomalous cases in a few species. All of the vertebrate species surveyed in this

study, except the microbat, possess T34 residues. The N-terminus of DARPP-32 in microbat is largely truncated (figure 5), and yet DARPP-32 in megabat is intact. Since DARPP-32 plays a key role in dopaminergic signalling and rewarding stimuli in the brain, we postulate that the truncated isoform of DARPP-32 (t-DARPP-32) N-terminus in microbat is due to change of lifestyle. Microbat lives in dark habitat and uses echolocation to sense its spatial location and to spot its preys such as insects. In contrast, megabat is a fruit bat that does not use echolocation but has large eyes to visually orient itself in forest and inside the cave. Furthermore, the sense of smell of megabat is excellent. These raise the possibility that change of lifestyle habit and the disuse of visual and smell senses had resulted in t-DARPP-32 in microbat, and renders it inactive in inhibiting PP1 function. t-DARPP-32 expression has been linked to β-catenin/TCF signalling, underscoring a novel function of t-DARPP-32 in cells (Vangamudi *et al*. 2011). Other novel functions of this truncated form of DARPP-32 remain unknown. We also observed short truncation of DARPP-32 N-termini in elephant, horse, and rabbit. Nonetheless, the T35 and other known phosphorylation sites in rabbit (Aitken *et al*. 1982) and T34 in elephant and horse for the activation of DARPP-32 are still intact, while what appear to be short N-termini truncations in DARPP-32 for these vertebrates may instead represent incompleteness of the relevant datasets.

Overall, our phylogenetics analysis reveals stratified evolution of phosphorylation site acquisition at the C-termini of DARPP-32 across vertebrates (figure 5). The well-conserved structurally and catalytically important residues in eukaryotic lineages for PP1 are correlated with their regulator-driven functional diversification (Ceulemans *et al*. 2002). Here, we propose overall evolutionary scheme for DARPP-32 (figure 6). Fish have very dissimilar C-termini to those in higher vertebrates. However, primordial C-terminus with up to ~100 residues in total protein length is detected in coelacanth, amphibian, and bird followed by extension of eutherian-like C-terminus up to 180 residues in total length in marsupial opossum. Full extended acidic domains are detected in Tasmanian devil and elephant to primates whereas full molecular features of human-like C-terminus are observed from horse to primates. There are gain and loss of human-like phosphorylation sites on some species where full set of human-like phosphorylation sites were eventually gained in primates. Some species such as elephant, horse, and rabbit lost their N-termini encompassing PP1 docking motif and yet still retain T34 site independently. Microbat is the only species we observed to exhibit deletion at N-terminus encompassing many conserved critical residues including T34 and T75, probably due to change of lifestyle leading to loss of PP1 inhibitory function. Gain of human-like phosphorylation site at the C-termini of DARPP-32 in higher vertebrates implicates expanded complexity of its regulation on PP1 activity in coupled to more diverse cellular cues.

Acknowledgements

This research work was supported by University of Malaya UMRG research grants RG163/09HTM and RG523/13HTM.

References

Aitken A, Bilham T and Cohen P 1982 Complete primary structure of protein phosphatase inhibitor-1 from rabbit skeletal muscle. *Eur. J. Biochem*. **126** 235–246

Albert KA, Hemmings HC Jr, Adamo AI, Potkin SG, Akbarian S, Sandman CA, Cotman CW, Bunney WE Jr and Greengard P 2002 Evidence for decreased DARPP-32 in the prefrontal cortex of patients with schizophrenia. *Arch. Gen. Psychiatry* **59** 705–712

Alpy F and Tomasetto C 2006 MLN64 and MENTHO, two mediators of endosomal cholesterol transport. *Biochem. Soc. Trans.* **34**: 343–345

Belkhiri A, Dar AA, Zaika A, Kelley M and El-Rifai W 2008 t-Darpp promotes cancer cell survival by up-regulation of Bcl2 through Akt-dependent mechanism. *Cancer Res.* **68** 395–403

Benderska N, Becker K, Girault JA, Becker CM, Andreadis A and Stamm S 2010 DARPP-32 binds to tra2-beta1 and influences alternative splicing. *Biochim. Biophys. Acta* **1799** 448–453

Bibb JA, Snyder GL, Nishi A, Yan Z, Meijer L, Fienberg AA, Tsai LH, Kwon YT, Girault JA, Czernik AJ, Huganir RL, Hemmings HC Jr, Nairn AC and Greengard P 1999 Phosphorylation of DARPP-32 by Cdk5 modulates dopamine signalling in neurons. *Nature* **402** 669–671

Bollen M, Peti W, Ragusa MJ and Beullens M 2010 The extended PP1 toolkit: designed to create specificity. *Trends Biochem. Sci.* **35** 450–458

Ceulemans H, Stalmans W and Bollen M 2002 Regulator-driven functional diversification of protein phosphatase-1 in eukaryotic evolution. *BioEssays* **24** 371–381

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG and Thompson JD 2003 Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31** 3497–3500

Chocarro-Calvo A, Zaballos MA, Santisteban P and García-Jiménez C 2012 DARPP-32 is required for MAPK/ERK signaling in thyroid cells. *Mol. Endocrinol*. **26** 471–480

Dancheck B, Nairn AC and Peti W 2008 Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry* **47** 12346–12356

Egloff MP, Johnson DF, Moorhead G, Cohen PT, Cohen P and Barford D 1997 Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *EMBO J* **16** 1876–1887

Fernandez É, Schiappa R, Girault J-A and Le Novère N 2006 DARPP-32 Is a robust integrator of dopamine and glutamate signals. *PLoS Comput. Biol.* **2** e176

Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC and Kuriyan J 1995 Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376** 745–753

Hemmings HC Jr and Greengard P 1986 DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein: regional, tissue, and phylogenetic distribution. *J. Neurosci.* **6** 1469–1481

Hemmings HC Jr, Girault JA, Nairn AC, Bertuzzi G and Greengard P 1992 Distribution of protein phosphatase inhibitor-1 in brain and peripheral tissues of various species: comparison with DARPP-32. *J. Neurochem.* **59** 1053–1061

Hemmings HC Jr, Nairn AC, Aswad DW and Greengard P 1984 DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Purification and characterization of the phosphoprotein from bovine caudate nucleus. *J. Neurosci.* **4** 99–110

Jones DT 1999 Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292** 195–202

Katoh K, Misawa K, Kuma K and Miyata T 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30** 3059–3066

Krieger E, Koraimann G and Vriend G 2002 Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins* **47** 393–402

Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD and Peti W 2010 Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure* **18** 1094–1103

Meister B, Fryckstedt J, Schalling M, Cortés R, Hökfelt T, Aperia A, Hemmings HC Jr, Nairn AC, Ehrlich M and Greengard P 1989 Dopamine- and cAMP-regulated phosphoprotein (DARPP-32) and dopamine DA1 agonist-sensitive Na^+,K^+-ATPase in renal tubule cells. *Proc. Natl. Acad. Sci. USA* **86** 8068–8072

Moorhead GB, Trinkle-Mulcahy L and Ulke-Lemée A 2007 Emerging roles of nuclear protein phosphatases. *Nat. Rev. Mol. Cell Biol.* **8** 234–244

Neyroz P, Desdouits F, Benfenati F, Knutson JR, Greengard P and Girault JA 1993 Study of the conformation of DARPP-32, a dopamine- and cAMP-regulated phosphoprotein, by fluorescence spectroscopy. *J. Biol. Chem.* **268** 24022–24031

Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, Gnad F, Cox J, Jensen TS, Nigg EA, Brunak S and Mann M 2010 Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal* **3** ra3

Ouimet CC, Miller PE, Hemmings HC Jr, Walaas SI and Greengard P 1984 DARPP-32, a dopamine- and adenosine 3':5'-monophosphate-regulated phosphoprotein enriched in dopamine-innervated brain regions. III. Immunocytochemical localization. *J. Neurosci.* **4** 111–124

Stipanovich A, Valjent E, Matamales M, Nishi A, Ahn JH, Maroteaux M, Bertran-Gonzalez J, Brami-Cherrier K, Enslen H, Corbillé AG, Filhol O, Nairn AC, Greengard P, Hervé D and Girault JA 2008 A phosphatase cascade by which rewarding stimuli control nucleosomal response. *Nature* **453** 879–884

Stone RA, Laties AM, Hemmings HC Jr, Ouimet CC and Greengard P 1986 DARPP-32 in the ciliary epithelium of the eye: a neurotransmitter-regulated phosphoprotein of brain localizes to secretory cells. *J. Histochem. Cytochem.* **34** 1465–1468

Svenningsson P, Nairn AC and Greengard P 2005 DARPP-32 mediates the actions of multiple drugs of abuse. *AAPS J* **7** E353–E360

Svenningsson P, Nishi A, Fisone G, Girault JA, Nairn AC and Greengard P 2004 DARPP-32: an integrator of neurotransmission. *Annu. Rev. Pharmacol. Toxicol.* **44** 269–296

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

Ubersax JA and Ferrell JE Jr 2007 Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell. Biol.* **8** 530–541

Valjent E, Pascoli V, Svenningsson P, Paul S, Enslen H, Corvol JC, Stipanovich A, Caboche J, Lombroso PJ, Nairn AC, Greengard P, Hervé D and Girault JA 2005 Regulation of a protein phosphatase cascade allows convergent dopamine and glutamate signals to activate ERK in the striatum. *Proc. Natl. Acad. Sci. USA* **102** 491–496

Van de Peer Y, Maere S and Meyer A 2009 The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10** 725–732

Vangamudi B, Peng DF, Cai Q, El-Rifai W, Zheng W and Belkhiri A 2010 t-DARPP regulates phosphatidylinositol-3-kinase-dependent cell growth in breast cancer. *Mol. Cancer* **9** 240

Vangamudi B, Zhu S, Soutto M, Belkhiri A and El-Rifai W 2011 Regulation of β-catenin by t-DARPP in upper gastrointestinal cancer cells. *Mol. Cancer* **10** 32

Walaas SI, Aswad DW and Greengard P 1983 A dopamine- and cyclic AMP-regulated phosphoprotein enriched in dopamine-innervated brain regions. *Nature* **301** 69–71

Xue Y, Zhou F, Zhu M, Ahmed K, Chen G and Yao X 2005 GPS: a comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.* **33** W184–W187

Yger M and Girault JA 2011 DARPP-32, Jack of All Trades… Master of Which? *Front. Behav. Neurosci.* **5** 56

Zhou FF, Xue Y, Chen GL and Yao X 2004 GPS: a novel group-based phosphorylation predicting and scoring method. *Biochem. Biophys. Res. Commun.* **325** 1443–1448

Zhu S, Belkhiri A and El-Rifai W 2011 DARPP-32 increases interactions between epidermal growth factor receptor and ERBB3 to promote tumour resistance to gefitinib. *Gastroenterology* **141** 1738–1748

MS received 22 November 2013; accepted 08 April 2014

Corresponding editor: STUART A NEWMAN