

RESEARCH ARTICLE

Open Access

# Evolutionarily conserved bias of amino-acid usage refines the definition of PDZ-binding motif

Takahiko Chimura<sup>1,2\*</sup>, Thomas Launey<sup>2</sup> and Masao Ito<sup>1</sup>

## Abstract

**Background:** The interactions between PDZ (PSD-95, Dlg, ZO-1) domains and PDZ-binding motifs play central roles in signal transductions within cells. Proteins with PDZ domains bind to PDZ-binding motifs almost exclusively when the motifs are located at the carboxyl (C-) terminal ends of their binding partners. However, it remains little explored whether PDZ-binding motifs show any preferential location at the C-terminal ends of proteins, at genome-level.

**Results:** Here, we examined the distribution of the type-I (x-x-S/T-x-I/L/V) or type-II (x-x-V-x-I/V) PDZ-binding motifs in proteins encoded in the genomes of five different species (human, mouse, zebrafish, fruit fly and nematode). We first established that these PDZ-binding motifs are indeed preferentially present at their C-terminal ends. Moreover, we found specific amino acid (AA) bias for the 'x' positions in the motifs at the C-terminal ends. In general, hydrophilic AAs were favored. Our genomics-based findings confirm and largely extend the results of previous interaction-based studies, allowing us to propose refined consensus sequences for all of the examined PDZ-binding motifs. An ontological analysis revealed that the refined motifs are functionally relevant since a large fraction of the proteins bearing the motif appear to be involved in signal transduction. Furthermore, co-precipitation experiments confirmed two new protein interactions predicted by our genomics-based approach. Finally, we show that influenza virus pathogenicity can be correlated with PDZ-binding motif, with high-virulence viral proteins bearing a refined PDZ-binding motif.

**Conclusions:** Our refined definition of PDZ-binding motifs should provide important clues for identifying functional PDZ-binding motifs and proteins involved in signal transduction.

## Background

The proteins that contain PDZ domain(s), often called PDZ proteins, play pivotal roles in dynamically organizing molecular architectures at specific intracellular regions in differentiating and differentiated cells [1,2]. Membrane proteins such as cell adhesion molecules, receptors, and channels form functional clusters within selective subcellular regions by binding to PDZ domains [2-5]. Furthermore, some PDZ proteins also anchor specific cytosolic proteins such as protein kinases, cytoskeleton-regulating enzymes and second-messenger-producing enzymes [2,6], and hence, contribute to precise signal transduction between extracellular and intracellular spaces at specific sites such as postsynaptic

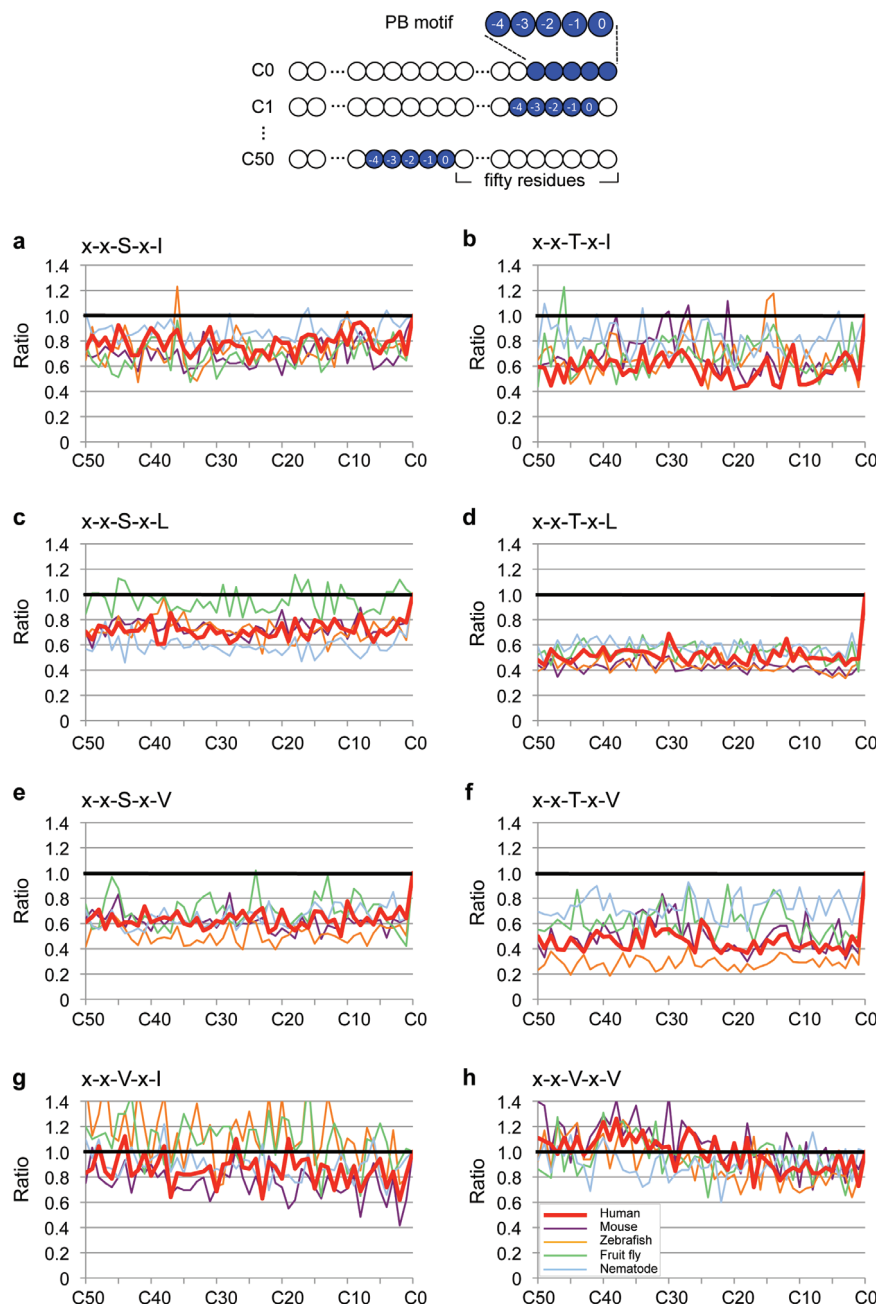
densities in neurons [2,7], immunological synapses in T-lymphocytes [8,9] and tight junctions in endothelial and epithelial cells [1,10].

PDZ domain, an evolutionarily conserved globular structure composed of 80-90 AAs (amino acids) recognizes particular regions of their interactors [6,11,12]. PDZ domains primarily bind to the C-terminal ends of proteins. Interactions between PDZ domains and internal regions of their binding partners have been also reported, though they are less common [2,11,12]. PDZ-binding motifs (hereafter 'PB motifs') have been proposed by sequence similarity in the C-terminal ends of proteins, whose bindings to PDZ domains are mediated by their C-terminal ends. PB motifs are currently categorized into at least three major types on the basis of two AAs located at positions 0 and -2 (Figure 1, upper panel), both being essential for binding to PDZ domains [11,13-15]. The type-I PB motif has the form S/T-x-I/L/

\* Correspondence: takahiko@brain.riken.jp

<sup>1</sup>Laboratory for Memory and Learning, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

Full list of author information is available at the end of the article



**Figure 1 Preferential localization of two-position-specified PB motifs at C-terminal ends of proteins.** Diagram on the top shows how positions (0, -1, -2, -3, -4) in a five-AA long PB motif are defined. Position 0 is defined as the last AA of the motif toward the C-terminal end. C0: the PB motif is located at the C-terminal end, C1: one AA away, and C50: fifty AAs away from the C-terminal end. (a-h) The "ratio" indicates the number of genes identified at each C0 to C50 position divided by the number at C0. The AAs at positions 0 and -2 are specified at the upper left in each graph. AAs are indicated with single letter code (S, serine; T, threonine; I, isoleucine; L, leucine; V, valine). x denotes any AA. Curves for the five species are shown in different colors.

V, in which Serine (S) or Threonine (T) is positioned at -2, any AA (x) at -1, and Isoleucine (I), Leucine (L) or Valine (V) at 0. The Type-II PB motif has the form  $\Phi$ -x- $\Phi$  (where  $\Phi$  denotes any hydrophobic AA). The type-III PB motif has the form D/E-x- $\Phi$  [11,12,14,16].

Although most of the reported PDZ-type interactions are mediated via these canonical C-terminal motifs [17], non-canonical C-terminal motifs are also reported [18-20]. As for the recognition of internal region by PDZ domains, studies based on tertiary structures

revealed the binding mode of PDZ domains with internal amino acid sequences within protein fold resembling a C-terminal end such as beta-hairpin "finger"-like structure [21] and Aspartate residue whose side chain mimics C-terminal end [22], suggesting that the recognition of internal sequences by PDZ domains requires particular strict conditions. Although some internal sequences are similar to those of C-terminal PB motifs [21,23-25], this is not always the case [22,26,27] or even not identified as "motifs" [28-30]. Overall, these internal PB represents less than 5% of PDZ-PB interactions in mammals [31] and are not included in the present study.

Protein-level interaction analysis in recent decades identified a number of protein bindings to PDZ domains *in vitro* and *in vivo* and revealed that most of them possess PB motifs at their C-terminal ends. These C-terminal motifs play an essential role in the bindings to PDZ domains because deletion of C-terminal regions or mutations in these motifs abrogates the PDZ-type interactions. Furthermore, when one or more amino acid residue(s) is added to the C-terminal side of the original "functional" C-terminal PB motifs, PDZ domain cannot recognize such "hidden" PB motifs [32-35]. These results indicate that PDZ domains have a strong positional preference for PB motifs located at the C-terminal ends of proteins. The short PB motif and multiple possible substitutions lead to loose definition of the motif however, making it a very common feature in proteins. Thus its presence at the C-terminal end of a protein cannot be considered as a strong predictor of this protein being involved in a PDZ interaction. Given the central role played by PDZ interactions in signaling and protein localization, it seemed important to revisit the issue of the PB "consensus" motifs and try to refine its current definition.

In order to solve this problem, in this report, we focused on the best-characterized canonical type-I and type-II PB motifs as model systems, and performed genomics-based characterization of these PB motifs. Based on the knowledge that functional PB motifs show positional preference at the C-terminal ends in terms of protein-protein interactions, we hypothesized that, also at the genome-level, PB motifs should show similar positional preferences, because if the encoded protein is genuinely involved in a PDZ interaction important for cell functions, genomes harboring mutations within functional PB motifs at the C-terminal ends would face selective disadvantage. Thus, selection pressure would prevent mutations within the functional PB motifs at the C-terminal ends. In contrast, random creation and disruption of PB motifs by mutations are likely to occur without affecting the cell functions, thus yielding a background scatter of PB motif in proteins. In order to test

this hypothesis, we measured the frequency of occurrence of PB motifs at the C-terminal ends of proteins relative to occurrence in the upstream fifty AAs region. Because PDZ domains are evolutionarily conserved [2,6,11,12,20], it is further expected that these patterns would be conserved to some extent across species. Analysis of the genomes from five species representative of major phylogenetic interval (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*) shows that PB motifs are indeed preferentially located at the C-terminal ends of proteins, which provides novel perspective on PB motifs.

The second point examined below is the preferential AA usage at the position surrounding 0 and -2. For the few PDZ domains proteins that have been previously studied, it has been established that the AA residues surrounding the canonical positions 0 and -2 also play a role for recognition. For example, screen of peptides or proteins binding to the well-characterized mammalian Dlg protein family members, revealed that those proteins preferentially bind to type-I PB motif containing a glutamate at position -3 [19,20,32,36]. Thus, from these particular examples it can be inferred that for most PDZ domain proteins, recognition of a PB motifs involves surrounding AAs in addition to the canonical position 0 and -2. We present a new genomics-based bioinformatics approach to identify such preferred AAs, for the most common variants of type-I and type-II PB motifs. We demonstrate biased AA usage at positions -4, -3 or -1, specifically for the PB motifs located at the C-terminal ends, relative to PB motifs found intra-proteins. In general, hydrophilic (i.e. charged and polar) AAs were favored. Importantly, our genomics-based approach also identified the AAs previously identified by interaction experiments, such as glutamate at position -3, thereby validating the accuracy of our genomics-based sieve. Moreover, the presence of these AAs at positions -4, -3 or -1 identified by genomics strongly correlates with only a few specific ontological classes, most notably transports across plasma membrane and intracellular signaling, suggesting that the presence of PB motifs containing these preferred AAs ('refined PB motifs') at C-terminal ends can serve as indicators for the function of the proteins bearing the refined motif. Consistently, several novel bindings between PDZ proteins and proteins possessing refined PB motifs were identified. We also provide evidence showing clear correlation between the AA compositions of PB motifs in influenza viral proteins and their pathogenicity. Our findings thus refine the "classical" PB motifs definition, as motifs having a strong positional preference at the C-terminal ends of proteins in terms of genomics, and several significant AA bias among the -4, -3 and -1 positions. Experiment, ontological correlation analysis and co-evolution all show that

the presence of a refined sequence at the C-terminus of a protein is predictive of its genuine involvement in a PDZ-type interaction and should accelerate the elucidation of the molecular mechanisms underlying signal transduction.

## Results and discussion

### Preferential positioning of PB motifs at the C-terminal ends

We chose six frequently encountered variants of the type-I PB motif (S/T-x-I/L/V). From the relatively small amount of data on the type-II PB motif ( $\Phi$ -x- $\Phi$ ) [17], we chose the two best-characterized variants, having the form V-x-V/I [13,15,37,38]. From the protein sequences of *H. sapiens* (58192 proteins, 20787 genes), *M. musculus* (37985 proteins, 22271 genes), *D. rerio* (21235 proteins, 18208 genes), *D. melanogaster* (20698 proteins, 14057 genes) and *C. elegans* (27287 proteins, 20003 genes) (Additional files 1 and 2, also see Materials and Methods) obtained from the Ensembl genome database project [39], we identified all the proteins containing these five-AA motifs, 'x-x-S/T-x-I/L/V' or 'x-x-V-x-V/I' within their fifty five AAs C-terminal domain (Figure 1). In this study, five residues (i.e. position 0 up to -4) were taken into consideration, because point mutation analyses have demonstrated that residues playing critical roles in binding to PDZ domains are mostly located within C-terminal 4 or 5 residues [15,40].

The identified proteins were categorized as "C0, C1...C50" based on the position of the motif, with "C0" denoting a PB motif at the C-terminal end, "C1" denoting a PB motif one AA away from the C-terminal end, and "C50" denoting a PB motif fifty AAs away from the C-terminal end (Figure 1 and Additional file 1). Individual proteins in each category (i.e. C0-C50) consist of the proteins derived from different genes and also of the splice variants derived from a single gene. Since our focus is on motif conservation among proteins derived from different genes, we conserved only one record per gene to eliminate the bias that would be introduced by genes with high number of redundant splice variants. Our pilot survey revealed that 84.6%, 82.7%, 80.7%, 77.4% and 79.7% of genes in the genome of the respective species listed above encodes proteins possessing at least one PB motif at any C0-C50 positions, confirming that the PB motifs are highly common AA sequence.

The rationale for our approach is that if functional motifs are preferentially located at the C-terminal end, the number of genes identified at C0 should be significantly larger than those identified at positions other than C0 (i.e. C1-C50). Indeed, we found that this was the case (Figure 1 and Additional file 3). To compare the numbers of genes at C1-C50 to that at C0, the number of genes found at each of the C1-C50 positions was

divided by that at C0. We plotted these relative numbers as 'ratio' in Figure 1. The 'ratio' is found to be lower than unity (i.e. 1) in most cases, as shown in Figure 1a-h for six type-I and two type-II PB motifs, in the five species examined. Furthermore, the ratio decreases abruptly between C0 and C1 rather than gradually, as is evident for type-I PB motifs (Figure 1). These results are consistent with previous reports showing that artificially shifting PB motifs at the C1 position by addition of a single residue at the C-terminal end abolishes the interactions between PB motifs and PDZ domains [33-35]. The genes identified by the C0 position search are shown in the Additional file 4.

To statistically evaluate the overall extent to which PBs are preferentially located at C0 relative to C1-C50, we defined the 'C-index' as the average of the ratios calculated for C1-C50 positions (Figure 2).

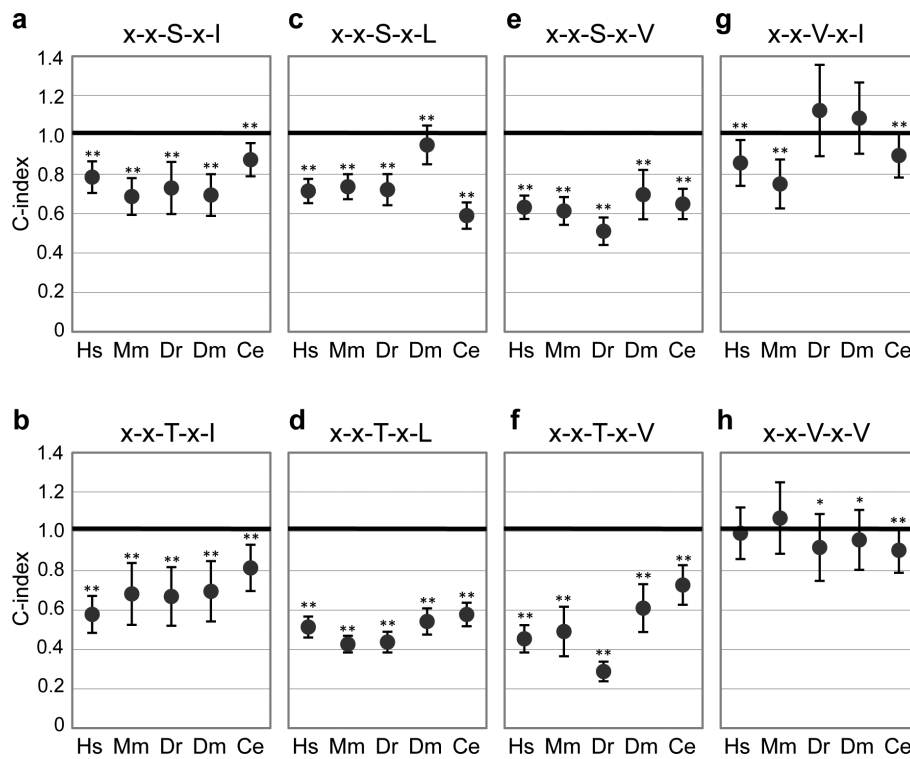
Equation 1:  $C_i$  indicates the number of genes found with a PB at position  $1 \leq i \leq 50$ .

$$C_{index} = \frac{1}{50} \sum_{i=1}^{50} \frac{C_i}{C_0}$$

A C-index lower than 1 implies that the PB motif is more often located at C-terminal end than within C1-C50. For example, the C-index of the x-x-T-x-V motifs in human is approximately 0.45 (Figure 2 and Additional file 3), which indicates that this motif localizes at the C-terminal end with a probability 2.2-fold higher than at C1-C50. As shown in Figure 2a-f, C-indexes are generally significantly lower than unity for type-I motifs. In contrast, for the two type-II PB motifs tested, the C-index is statistically not different from unity in several cases because of the large standard deviations of these C-indexes (Figures 2g and 2h). Therefore, among all the PB motifs considered here, the two type-II PB motifs are less preferentially positioned at the C-terminal ends of proteins than type-I PB motifs.

### Biased amino-acid usage at positions -4, -3 and -1 of the C-terminal PB motifs

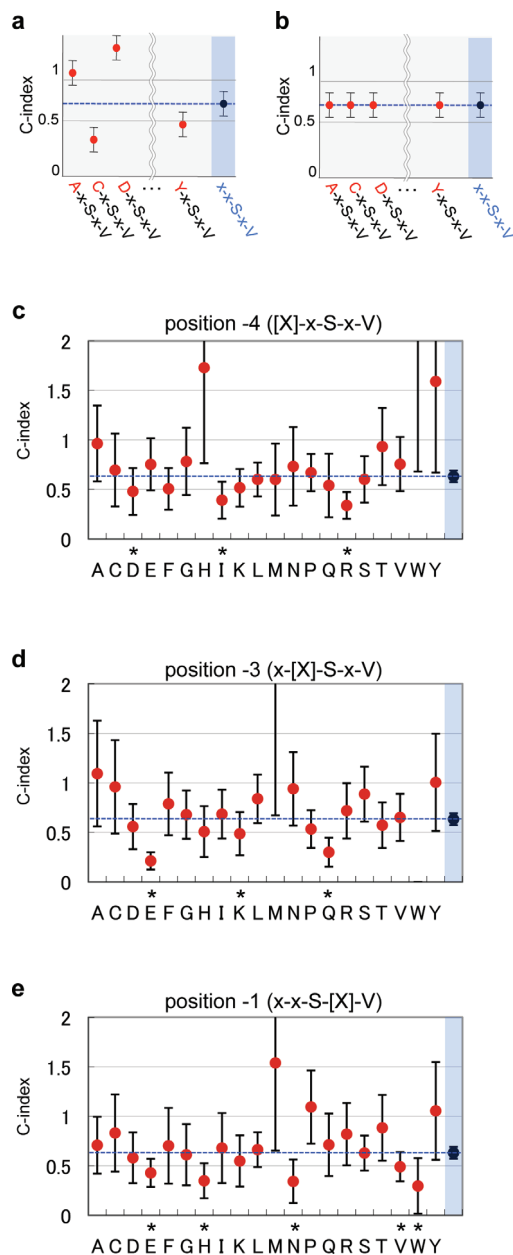
Next we examined whether the AAs present at the positions -4, -3 and -1 in the C0 PB motifs show any biased usage frequency compared to similar positions in PB motifs located within C1-C50. For example, if a particular AA [X] is over-represented at position -4 in x-x-S-x-V motif located at the C0 position but not for C1-C50 PB positions, the C-index of [X]-x-S-x-V would be lower than the C-index of x-x-S-x-V (e.g. C and Y in Figure 3a). In contrast, if such C0-specific over-representations is not the case, the C-index of [X]-x-S-x-V would not vary with the AA-substitution in the [X], showing values similar to that of x-x-S-x-V (Figure 3b). This approach has the advantage of determining usage



**Figure 2 C-indexes of two-position-specified PB motifs in five species.** The averages of the values in C1 to C50 shown in Figure 1 are plotted as C-indexes. Values significantly lower than 1 are indicated by asterisks (Mann-Whitney test, \* $P < 0.01$ , \*\* $P < 0.001$ ). Error bars indicate standard deviations. Hs, *Homo sapiens* (human); Mm, *Mus musculus* (mouse); Dr, *Danio rerio* (zebrafish); Dm, *Drosophila melanogaster* (fruit fly); Ce, *Caenorhabditis elegans* (nematode). Actual values for each plot are shown in Additional file 3.

bias in the context of a particular sequence, independently of the overall frequency of individual AA in proteins. Furthermore, the C-index provides an unbiased normalization that only depends on the position of this sequence along the C-terminal domain. Figure 3c-e shows the actual C-indexes and standard deviations of human x-x-S-x-V motif, in which twenty AAs are substituted in each “x” at positions -4, -3 and -1, respectively. The actual numbers of identified genes, C-indexes and their standard deviations are listed in Additional file 5. It is clearly apparent that the C-indexes of the three-position-specified PB motifs vary with the types of AAs at positions -4, -3 and -1. In the case of human x-x-S-x-V shown as an example here (Figure 3), the AA substitution D, I and R at position -4, E, K and Q at position -3, and E, H, N, V and W at position -1 all result in significantly lower C-index than that of the two-position-specified PB motif. This observation was confirmed for all the variants of the PB motif examined here (Additional file 5). These results indicate that there are specific bias for the AAs surrounding PB motifs located at the C-terminal ends. The results are summarized in the Figure 4, showing all the AA substitutions that result in significantly lower C-indexes than those of two-position

specified PB motifs. Among these substitutions, note that glutamate (E) at -3 shows a very robust usage bias (Figures 3c and 4, Additional file 5), across variants of the Type-I and Type-II motifs as well as across distant species, strongly suggesting that its presence in C-terminal PB motif results from an evolutionary process. Since the PB motifs containing these AAs (‘refined PB motifs’) selectively appear at the C-terminal ends (Figure 3), such PB motifs are probably actually recognized by PDZ proteins. If not, random mutations during evolution would have erased this selective localization at the C-terminal ends of the refined PB motifs. Furthermore, most of the identified AAs (Figure 4) are hydrophilic ones including electrically charged AAs D, E, H, K and R, and polar N, Q, S and T. Meanwhile, hydrophobic AAs such as I, L and V, or AAs with simple side chain such as G and A are rarely identified. This is consistent with the previous reports that have demonstrated the important roles of electrostatic interactions in the bindings between PB motifs and PDZ domains [41-43]. This finding also support the notion that the refined PB motifs are likely to be recognized by PDZ domains. The additional AAs (Figure 4) in the refined PB motif may also contribute to the regulation of PDZ binding.



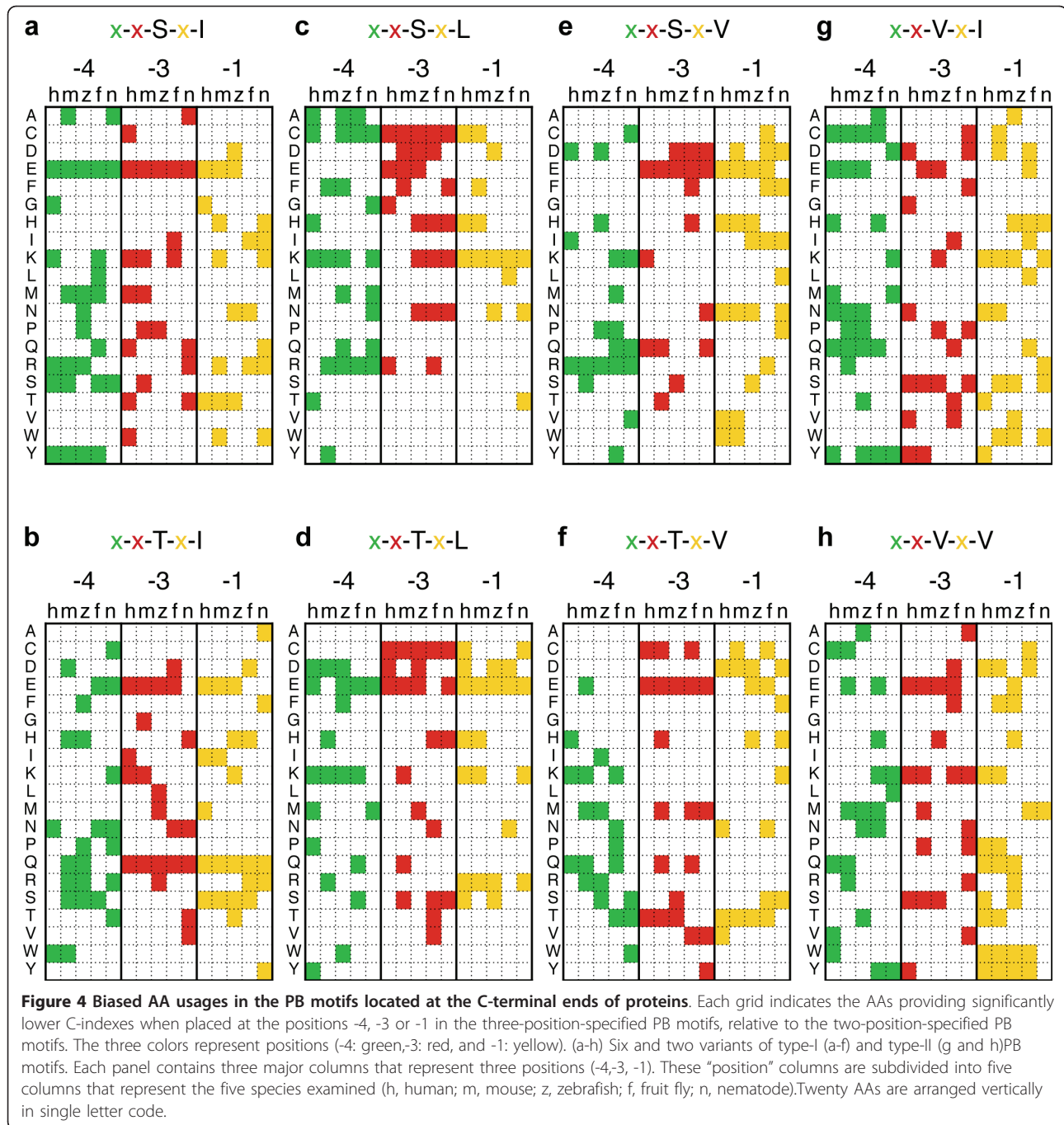
**Figure 3 C-indexes of three-position-specified PB motifs derived from human x-x-S-x-V.** (a and b) Expected patterns of the C-indexes of the three-position specified PB motifs when the AA compositions in the 'x' positions at C0 are different (a) or similar (b) to those at C1-C50, respectively. (c-e) Actual C-indexes of three-position specified PB motifs derived from human x-x-S-x-V. Additionally specified positions (-4 in c, -3 in d, -1 in e) in the three-position-specified PB motifs are indicated at the top of each panel with an [X]. Twenty AAs in single character code are indicated below the panels. The black filled circle in the most right column in each plot indicates the C-index of two-position-specified PB motifs, which is the same value as that shown in Figure 2. Values significantly lower than those of two-position-specified PB motifs are indicated with asterisks (Steel test, \* $P < 1.00E-4$ ). Error bars indicate standard deviations. Actual values for each plot are shown in Additional file 5.

As a case in point, a recent report [44] about the regulation of the NR2B NMDA receptor subunit may shed some light on the physiological role of variations within refined motifs. Sanz-Clemente and collaborators have confirmed that the binding of NR2B to PDZ protein PSD-95 is attenuated by casein kinase 2 (CK2)-mediated phosphorylation of serine at position -2 within the type-I PB motifs of NR2B (-IESDV). A point mutation of E at -3 to Q (i.e. -IQSDV) abolished phosphorylation by CK2 however, while binding to PSD95 was unaffected, which is consistent with our findings that both E and Q at -3 are identified as biased amino acids at position -3 of x-x-S-x-V (Figure 4e, column '-3' and sub-column 'h' and 'm'). Thus, we suggest that selection pressure may have been directed not only toward increasing PDZ binding affinity but also toward creating both constitutive and regulated interactions.

Because amino acids identified in Figure 4 are evolutionarily conserved to some extent, orthologous genes should be commonly identified in the species studied here. As expected, this was the case. For example, our analysis revealed that human and mouse genome encode 1089 genes and 972 genes that contain at least one refined PB motifs at their C-terminal ends, respectively, in which 501 of human 1089 genes and 512 of mouse 972 genes are orthologous to each other (Additional file 6).

#### Over-representations of proteins involved in signal transductions

To date, several approaches using bioinformatics to predict PDZ-type interactions have been implemented. For example, Giallourakis and colleagues used evolutionary conservation, similarity of coexpression profiles relative to PDZ proteins in different tissues as well as the presence of PB motifs as criteria for their predictions [31]. Several strategies based on genome-wide proteomics have also been formulated to identify binding sequences of particular PDZ domains by screening of random or genome-encoded peptide libraries *in vitro* [19,20,45]. Using such quasi-exhaustive analyses, computational simulations to calculate binding affinity between any PDZ domains and genome-encoded proteins have been performed [46-48]. Although predictions can be made as whether a protein may bind to PDZ domains, *in vitro* screening of peptides is not exempt of artifacts that may negatively affect prediction accuracy [49]. Similarly, prediction algorithms that are trained against a set of known PDZ-PB interactions may be influenced by any error or omission in the training set. Our approach is orthogonal to these studies since it does not depend *a priori* on experimental evidence and may thus provide complementary information and help identify both false positive and false negative. We used three different strategies to evaluate the quality of our predictions: An



ontological analysis, a comparison with some published PDZ-type interaction data and an *in vivo* test of interactions predicted by our analysis.

If the refined PB motifs correspond to genuine evolution-selected sequences, it is expected that the proteins possessing these PB motifs would present some similarities of function. Furthermore, considering the literature on proteins with identified PDZ-type interactions [1,2,6,11,12], it is expected that proteins involved in signal

transductions and membrane proteins should be over-represented in this subset of C-terminal PB motifs. In order to test *a posteriori* the quality of our predictions, we determined whether any ontological category is over-represented among the human proteins (1089 genes) that possess the refined PB motifs shown in Figure 4 based on the GO (Gene Ontology at <http://www.geneontology.org/>) [50] molecular function term (Table 1). This analysis showed that some functional categories are clearly

**Table 1 GO terms over-represented in the proteins possessing PB motifs with the amino acids identified in Figure 4**

category: Molecular Function	WITH the AAs			WITHOUT the AAs	
	GO term	size	hit	p-value	hit
transmembrane transporter activity	880	86	4.06E-13	23	5.55E-02
signal transducer activity	2116	149	2.26E-10	53	1.05E-02
substrate-specific transporter activity	936	80	2.08E-09	24	6.09E-02
peptide binding	177	16	4.00E-03	7	4.49E-02
nucleoside-triphosphatase regulator activity	406	29	5.14E-03	10	2.14E-01

Significantly ( $p < 0.01$ ) over-represented GO terms (Molecular Function) in the proteins possessing PB motif WITH the amino acids shown in Figure 4 are identified by web-based tool PIPE2. The identified GO terms are indicated in the 'GO term' column. The number of genes categorized in the GO terms among the proteins possessing PB motifs is indicated in the 'hit' column. The extent of statistical significance is indicated in the 'p-value' column. Although significantly over-represented GO terms were not detected in the proteins possessing PB motifs WITHOUT the amino acids shown in Figure 4, the number of genes and their p-values are also shown in the right side ('WITHOUT AAs' column) for comparison. The total number of genes in each category in the PIPE2 database is indicated in the 'size' column. AA; amino acid.

over-represented among the proteins with refined PB. These include receptors of growth factors, G-protein-coupled receptors, cell adhesion molecules, ion channels, kinases, proteins involved in trafficking, GTPase-activating proteins (GAPs) and guanine nucleotide exchange factors (GEFs), all of which are important players in signal transductions. In contrast, however, no obvious function over-representation was detected for the human proteins (488 genes in total) that did not possess any of the -4, -3, -1 biases shown in Figure 4, despite the presence of a C-terminal classical PB motifs. Taking that PDZ proteins play central roles in trafficking and organizing molecular architectures containing membrane proteins and signaling proteins, the results shown in Table 1 are consistent with our refined PB motif definitions being better predictor of molecular function than the classical PB motif.

#### Binding activities of PB motifs containing the identified amino acids

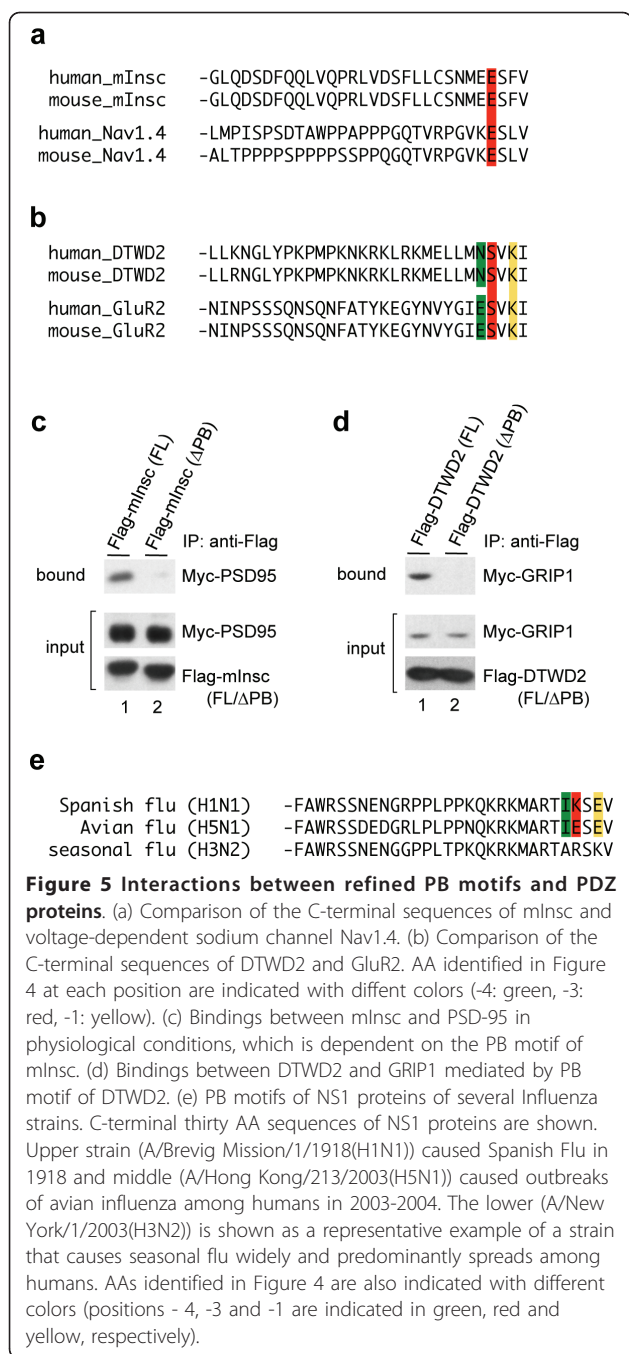
Next, we further asked whether the presence of refined PB motifs serves as a reliable indicator for the identification of PDZ-type interactions. For this purpose, we surveyed published quasi-exhaustive interaction data [19,45], in which interactions between 157 mouse PDZ domains and 217 genome-encoded C-terminal peptides were tested *in vitro*. We found that 68 out of 217 peptides possess refined PB motifs (Additional file 7). According to the aforementioned study, 61 out of 68 are experimentally validated PDZ-binding peptides. Furthermore, 4 (Kir3.3, GluR3, EphA4 and ErbB2) out of 7 negative peptides, have already been shown to be PDZ-binding proteins in other works [33,51-55]. Thus, combined with these results, 95% (65 out of 68) were true positives.

These results prompted us to test *in vivo* some of the predictions produced by *in silico* sequence analysis and to examine bindings between PDZ proteins and proteins possessing refined PB motifs shown in Figure 4. The

mammalian homologue of *Drosophila* Inscuteable (mInsc) play important roles in defining cell polarity during development [56]. mInsc possess a refined type-I PB motif, x-E-S-x-V (Figure 5a), as identified in Figure 4e. However, it has not been shown whether the PB motif is functional and allows mInsc to bind to PDZ domains. This three-position specified PB motif is also observed at the C-terminal ends of voltage-dependent sodium channel Nav1.4 proteins (Figure 5a). Because Nav1.4 binds to a PDZ protein PSD-95 [52], we predicted that mInsc would also bind to PSD-95. In order to test this possibility, Flag-tagged mInsc and Myc-tagged PSD-95 were coexpressed in COS-7 cells and co-immunoprecipitation analyses were performed (Figure 5c). As expected, PSD-95 was coimmunoprecipitated by mInsc (lane 1 in Figure 5c). Furthermore, deletion of C-terminal four AAs of mInsc disrupted this binding, indicating that the binding between PSD-95 and mInsc is mediated by the PB motif of mInsc. We also tried to identify functional type-II PB motifs. As shown in Figure 5b, DTWD2, an unknown-function protein, possess three refined type-II PB motifs N-x-V-x-I, x-S-V-x-I and x-x-V-K-I, all of which are identified in Figure 4. Interestingly, two of them, x-S-V-x-I and x-x-V-K-I are also found at the C-terminal end of GluR2, a subunit of AMPA-type glutamate receptors. Considering that a PDZ protein GRIP1 binds to the type-II PB motif of the GluR2 [37,52,57], it is expected that DTWD2 also binds to GRIP1. As shown in Figure 5d, the interactions between DTWD2 and GRIP1 was indeed observed, in which the type-II PB motifs of DTWD2 was essential. Thus, we successfully identified functional PB motifs based on the three-position specified PB motifs identified in our study (Figure 4).

Finally, we tested the hypothesis that the refined PB motifs correspond to evolutionary selected sequences by examining the co-evolution of virus pathogens. Several types of virus express viral proteins that possess type-I PB motifs at their C-terminal ends and bind to cellular





flu in Asia in 2003-2004 possess I-K-S-E-V and I-E-S-E-V motifs at their C-terminal ends, respectively [62,63]. These correspond to some of the refined PB motifs identified here, I-x-S-x-V, x-K-S-x-V and x-x-S-E-V in Spanish flu NS1 (Figure 5e, top row) and I-x-S-x-V, x-E-S-x-V and x-x-S-E-V in Avian flu NS1 (Figure 5e, middle row). In contrast, the PB motifs of the NS1 proteins derived from low-pathogenic strains producing seasonal flu (H3N2) correspond to a non-refined A-R-S-K-V (compare to Figure 5e, bottom row). Interestingly, two of the three-positions specified PB motifs, I-x-S-x-V and x-K-S-x-V, found in highly pathogenic strain are specifically found in human (Figure 4e, column 'h'), which may suggest that these strains of highly pathogenic influenza viruses have evolved to efficiently bind to human PDZ proteins. These results suggest that the three-position specified PB motifs should be evaluated as potential indicators of viral pathogenicities.

## Conclusions

We did a genome-level comprehensive study of the PB motif variants present in five phylogenetically distant species. We have shown that PB motifs are preferentially located at the C-terminal ends of proteins, in line with experimental results showing that PDZ interactions preferentially take place with C-terminal PB motifs. Our analysis identified specific AA usage bias for the -4, -3 and -1 positions surrounding the "classical" two-position-specified PB motifs, x-x-S/T-x-I/L/V and x-x-Φ-x-Φ. Ontological analysis of the proteins presenting this refined C-terminal PB motifs revealed very specific bias toward signaling and transport proteins. PDZ-type interactions are known to play key roles in these cellular processes, suggesting that the protein subset with refined PB motif are likely to be engaged in genuine PDZ-type interaction. By correlating motif position with sequence variation, the innovative analysis method presented here allows to detect fine variations in protein motifs, across variants and across species, while not requiring any training set. Being orthogonal with previously described strategies, we have shown that it provides a complementary approach to refine *in silico* predictions. Because these *in silico* analyses are applicable to any species whose protein sequences are comprehensively registered into databases, the methodology shown here has general applicability in discovering and evaluating any protein motif with an identified positional biases.

## Methods

### Bioinformatics

We downloaded protein sequences ('dataset\_1' in Additional file 1) assigned by 'protein\_coding\_genes' with gene ID numbers and protein ID numbers from

PDZ proteins [58]. The PB motif sequences of NS1 proteins, viral proteins of influenza, vary with the isolates of influenza strains, whose pathogenicity can correlate with the binding activity of the PB motif of NS1 with cellular PDZ proteins [59-61]. These results prompted us to test whether NS1 derived from highly pathogenic strain possess the refined PB motifs shown in Figure 4. As shown in Figure 5e, the NS1 proteins of the highly pathogenic influenza viruses H1N1 that caused the "Spanish Flu" in 1918 and H5N1 that caused several outbreaks of Avian

Ensembl project <http://www.ensembl.org/index.html>[39] by using BioMart (Additional file 1). A version of the dataset was Release 55. Because each dataset contains the information of single species, the following procedures were separately done for the five species. After removing extraneous characters, each text line contains a single gene ID, a single protein ID and a single protein sequence. We further extracted protein sequences that contain asterisks (\*) denoting stop codons and more than fifty-four AA long to perform the C0 to C50 searches ('dataset\_2' in Additional file 1). Specifically for human datasets, proteins encoded in the haplotypic chromosomal regions, denoted by chromosome name HSCHR6\_MHC\_APD, HSCHR6\_MHC\_COX, HSCHR6\_MHC\_DBB, HSCHR6\_MHC\_MANN, HSCHR6\_MHC\_MCF, HSCHR6\_MHC\_QBL, HSCHR6\_MHC\_SSTO, HSCHR4\_1 and HSCHR17\_1, were removed to avoid multiple identifications of the same genes [64]. The numbers of proteins and genes in dataset\_1 and \_2 are shown in Additional file 2. Fifty-one data subsets ('dataset\_C0' to 'dataset\_C50' in Additional file 1) were generated for each species, based on the position of the motif within C0-C50, then C0-C50 searches were performed. All the Perl and UNIX scripts corresponding to these steps are available upon request to the author.

#### Data analysis and statistics

All the statistical tests were performed using KyPlot 5.0 software (KyenceLab Inc. Japan). Non-parametric Mann-Whitney test or Steel test was used to examine statistical difference. *P*-values are indicated in each figure.

#### Detection of over-represented GO molecular function term

The Ensembl gene IDs were converted to Entrez Gene ID using web-based tool Clone/Gene ID Converter, version 2.0 <http://idconverter.bioinfo.cnio.es/>[65]. The over-represented ontological categories were identified using PIPE2 <http://pipe2.systemsbio.net/PIPE2/>[66] with the Entrez Gene IDs.

#### Co-immunoprecipitation assay

The cDNAs encoding full-length mInsc, DTWD2, PSD-95 and GRIP1 were amplified from mouse brain cDNA libraries by PCR and subcloned into pCMV-Tag2 or pCMV-Tag3 (Clontech) for the expression of Flag-tagged or Myc-tagged proteins, respectively. As for mInsc and DTWD2, deletion mutants lacking C-terminal four AAs were also constructed. Transfection of these plasmids into COS-7 cells (RIKEN Cell Bank) was performed using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocols. Transfected cells were lysed in Tris buffer (120 mM NaCl, 1 mM EDTA, 20 mM Tris-Cl pH 7.5, 0.5% (v/v) Triton X-100, protease inhibitors

cocktail) and briefly sonicated. Lysates were centrifugated (10 min; 15,000 × g) to remove insoluble matter. Anti-Flag-M2 agarose (Sigma-Aldrich) were added to the supernatant fraction and incubated for 2 hrs at 4°C. After washing, all precipitated complex were denatured in SDS sample buffer and subjected to SDS-PAGE followed by Western blot analysis using anti-Myc antibody (Santa Cruz Biotechnology) or anti-Flag antibody (Sigma-Aldrich), and chemiluminescence-based detection system ECL plus (GE Healthcare).

#### Additional material

**Additional file 1: Schema of bioinformatics used in this study.** The blue filled squares indicate datasets that contain gene IDs and protein IDs (black filled circles) and protein sequences (horizontal lines). Asterisk (\*) in each line end denotes a stop codon. Details are described in Materials and Methods.

**Additional file 2: The number of proteins and genes in the datasets used in the bioinformatics analyses.**

**Additional file 3: The number of genes identified by C0 to C50 searches using two-position-specified PB motifs as queries.**

**Additional file 4: Gene IDs and five-amino-acid sequences located at the C-terminal ends of proteins isolated by each C0 search.**

**Additional file 5: The number of genes that encode proteins possessing three-position-specified PB motifs located at C0 to C50 positions.**

**Additional file 6: Orthologous genes between human and mouse that encode proteins possessing refined PB motifs.**

**Additional file 7: 68 C-terminal peptides of mouse genome-encoded proteins possessing refined PB motifs and their bindings to PDZ domains.**

#### Abbreviations

aa: amino acid; PB motif: PDZ (PSD-95,Dlg, ZO-1)-binding motif; GO: gene ontology; AMPA: alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid;

#### Acknowledgements

This work was supported by RIKEN Brain Science Institute and Grant-in-Aid for Young Scientists from Japanese Ministry of Education, Culture, Sports, Science and Technology [19700304 to T.C.]. We thank Dr. Kenta Nakai and members of his laboratory for critical reading of the manuscript.

#### Author details

<sup>1</sup>Laboratory for Memory and Learning, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. <sup>2</sup>Launey Research Unit for Molecular Neurocybernetics, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

#### Authors' contributions

TC and MI designed the project; TC performed the experiments; TC, TL and MI performed the data analysis; TC, TL and MI wrote the manuscript. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 19 January 2011 Accepted: 8 June 2011

Published: 8 June 2011

#### References

1. Suzuki A, Ohno S: The PAR-aPKC system: lessons in polarity. *J Cell Sci* 2006, **119**:979-987.

2. Feng W, Zhang M: **Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density.** *Nat Rev Neurosci* 2009, **10**:87-99.
3. Kim E, Niethammer M, Rothschild A, Jan YN, Sheng M: **Clustering of Shaker-type K<sup>+</sup> channels by interaction with a family of membrane-associated guanylate kinases.** *Nature* 1995, **378**:85-88.
4. Ehlers MD, Tingley WG, Huganir RL: **Regulated subcellular distribution of the NR1 subunit of the NMDA receptor.** *Science* 1995, **269**:1734-1737.
5. Kornau HC, Schenker LT, Kennedy MB, Seeburg PH: **Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95.** *Science* 1995, **269**:1737-1740.
6. Kim E, Sheng M: **PDZ domain proteins of synapses.** *Nat Rev Neurosci* 2004, **5**:771-781.
7. Cho KO, Hunt CA, Kennedy MB: **The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein.** *Neuron* 1992, **9**:929-942.
8. Ludford-Menting MJ, Oliaro J, Sacirbegovic F, Cheah ET, Pedersen N, Thomas SJ, Pasam A, Iazzolino R, Dow LE, Waterhouse NJ, Murphy A, Ellis S, Smyth MJ, Kershaw MH, Darcy PK, Humbert PO, Russell SM: **A network of PDZ-containing proteins regulates T cell polarity and morphology during migration and immunological synapse formation.** *Immunity* 2005, **22**:737-748.
9. Yeh JH, Sidhu SS, Chan AC: **Regulation of a late phase of T cell polarity and effector functions by Crtam.** *Cell* 2008, **132**:846-859.
10. Willott E, Balda MS, Fanning AS, Jameson B, Van Itallie C, Anderson JM: **The tight junction protein ZO-1 is homologous to the Drosophila discs-large tumor suppressor protein of septate junctions.** *Proc Natl Acad Sci USA* 1993, **90**:7834-7838.
11. Sheng M, Sala C: **PDZ domains and the organization of supramolecular complexes.** *Annu Rev Neurosci* 2001, **24**:1-29.
12. Noury C, Grant SG, Borg JP: **PDZ domain proteins: plug and play!** *Sci STKE* 2003, **2003**:RE7.
13. Lin SH, Arai AC, Wang Z, Nothacker HP, Civelli O: **The carboxyl terminus of the prolactin-releasing peptide receptor interacts with PDZ domain proteins involved in alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor clustering.** *Mol Pharmacol* 2001, **60**:916-923.
14. Hung AY, Sheng M: **PDZ domains: structural modules for protein complex assembly.** *J Biol Chem* 2002, **277**:5699-5702.
15. Wiedemann U, Boisguerin P, Leben R, Leitner D, Krause G, Moelling K, Volkmer-Engert R, Oschkinat H: **Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides.** *J Mol Biol* 2004, **343**:703-718.
16. Maximov A, Südhof TC, Bezprozvanny I: **Association of neuronal calcium channels with modular adaptor proteins.** *J Biol Chem* 1999, **274**:24453-24456.
17. Beumung T, Skrabanek L, Niv MY, Mukherjee P, Weinstein H: **PDZBase: a protein-protein interaction database for PDZ-domains.** *Bioinformatics* 2005, **21**:827-828.
18. Zhang Y, Yeh S, Appleton BA, Held HA, Kausalya PJ, Phua DC, Wong WL, Lasky LA, Wiesmann C, Hunziker W, Sidhu SS: **Convergent and divergent ligand specificity among PDZ domains of the LAP and zonula occludens (ZO) families.** *J Biol Chem* 2006, **281**:22299-22311.
19. Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaja LA, MacBeath G: **PDZ domain binding selectivity is optimized across the mouse proteome.** *Science* 2007, **317**:364-369.
20. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin X, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS: **A specificity map for the PDZ domain family.** *PLoS Biol* 2008, **6**:e239.
21. Hillier BJ, Christopherson KS, Prehoda KE, Bredt DS, Lim WA: **Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex.** *Science* 1999, **284**:812.
22. Penkert RR, DiVittorio HM, Prehoda KE: **Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex.** *Nat Struct Mol Biol* 2004, **11**:1122-1127.
23. Paasche JD, Attramadal T, Kristiansen K, Oksvold MP, Johansen HK, Huitfeldt HS, Dahl SG, Attramadal T: **Subtype-specific sorting of the ETA endothelin receptor by a novel endocytic recycling signal for G protein-coupled receptors.** *Mol Pharmacol* 2005, **67**:1581-1590.
24. Trejo J: **Internal PDZ ligands: novel endocytic recycling motifs for G protein-coupled receptors.** *Mol Pharmacol* 2005, **67**:1388-1390.
25. Slattery C, Jenkin KA, Lee A, Simcocks AC, McAinch AJ, Poronnik P, Hryciw DH: **Na<sup>+</sup>-H<sup>+</sup> exchanger regulatory factor 1 (NHERF1) PDZ scaffold binds an internal binding site in the scavenger receptor megalin.** *Cell Physiol Biochem* 2011, **27**:171-178.
26. Wong HC, Bourdelas A, Krauss A, Lee HJ, Shao Y, Wu D, Mlodzik M, Shi DL, Zheng J: **Direct binding of the PDZ domain of Dishevelled to a conserved internal sequence in the C-terminal region of Frizzled.** *Mol Cell* 2003, **12**:1251-1260.
27. Tuomi S, Mai A, Nevo J, Laine JO, Vilki V, Ohman TJ, Gahmberg CG, Parker PJ, Ivaska J: **PKCepsilon regulation of an alpha5 integrin-ZO-1 complex controls lamellae formation in migrating cancer cells.** *Sci Signal* 2009, **2**:ra32.
28. van Huizen R, Miller K, Chen DM, Li Y, Lai ZC, Raab RW, Stark WS, Shortridge RD, Li M: **Two distantly positioned PDZ domains mediate multivalent INAD-phospholipase C interactions essential for G protein-coupled signaling.** *EMBO J* 1998, **17**:2285-2297.
29. Runyon ST, Zhang Y, Appleton BA, Sazinsky SL, Wu P, Pan B, Wiesmann C, Skelton NJ, Sidhu SS: **Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3.** *Protein Sci* 2007, **16**:2454-2471.
30. Lenfant N, Polanowska J, Bamps S, Omi S, Borg JP, Reboul J: **A genome-wide study of PDZ-domain interactions in C. elegans reveals a high frequency of non-canonical binding.** *BMC Genomics* 2010, **11**:671.
31. Giallourakis C, Cao Z, Green T, Wachtel H, Xie X, Lopez-Illasaca M, Daly M, Rioux J, Xavier R: **A molecular-properties-based approach to understanding PDZ domain proteins and PDZ ligands.** *Genome Res* 2006, **16**:1056-1072.
32. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC: **Recognition of unique carboxyl-terminal motifs by distinct PDZ domains.** *Science* 1997, **275**:73-77.
33. Xia J, Zhang X, Staudinger J, Huganir RL: **Clustering of AMPA receptors by the synaptic PDZ domain-containing protein PICK1.** *Neuron* 1999, **22**:179-187.
34. Fujii K, Maeda K, Hikida T, Mustafa AK, Balkissoon R, Xia J, Yamada T, Ozeki Y, Kawahara R, Okawa M, Huganir RL, Ujike H, Snyder SH, Sawa A: **Serine racemase binds to PICK1: potential relevance to schizophrenia.** *Mol Psychiatry* 2006, **11**:150-157.
35. Bassan M, Liu H, Madsen KL, Armsen W, Zhou J, Desilva T, Chen W, Paradise A, Brasch MA, Staudinger J, Gether U, Irwin N, Rosenberg PA: **Interaction between the glutamate transporter GLT1b and the synaptic PDZ domain protein PICK1.** *Eur J Neurosci* 2008, **27**:66-82.
36. Kurakin A, Swistowski A, Wu SC, Bredesen DE: **The PDZ domain as a complex adaptive system.** *PLoS One* 2007, **2**:e953.
37. Osten P, Khatri L, Perez JL, Köhr G, Giese G, Daly C, Schulz TW, Wensky A, Lee LM, Ziff EB: **Mutagenesis reveals a role for ABP/GRIP binding to GluR2 in synaptic surface accumulation of the AMPA receptor.** *Neuron* 2000, **27**:313-325.
38. Perez JL, Khatri L, Chang C, Srivastava S, Osten P, Ziff EB: **PICK1 targets activated protein kinase Calpha to AMPA receptor clusters in spines of hippocampal neurons and reduces surface levels of the AMPA-type glutamate receptor subunit 2.** *J Neurosci* 2001, **21**:5417-5428.
39. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M: **The Ensembl genome database project.** *Nucleic Acids Res* 2002, **30**:38-41.
40. Schultz J, Hoffmüller U, Krause G, Ashurst J, Macias MJ, Schmieder P, Schneider-Mergener J, Oschkinat H: **Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels.** *Nat Struct Mol Biol* 1998, **5**:19-24.
41. Harris BZ, Lau FW, Fujii N, Guy RK, Lim WA: **Role of electrostatic interactions in PDZ domain ligand recognition.** *Biochemistry* 2003, **42**:2797-2805.
42. Sugi T, Oyama T, Muto T, Nakanishi S, Morikawa K, Jingami H: **Crystal structures of autoinhibitory PDZ domain of Tamalin: implications for metabotropic glutamate receptor trafficking regulation.** *EMBO J* 2007, **26**:2192-2205.
43. Dominiak PM, Volkov A, Li X, Messerschmidt M, Coppens P: **A theoretical databank of transferable aspherical atoms and its application to**

- electrostatic interaction energy calculations of macromolecules. *J Chem Theory Comput* 2007, **3**:232-247.
44. Sanz-Clemente A, Matta JA, Isaac JT, Roche KW: **Casein kinase 2 regulates the NR2 subunit composition of synaptic NMDA receptors.** *Neuron* 2010, **67**:984-996.
  45. Chen JR, Chang BH, Allen JE, Stiffler MA, MacBeath G: **Predicting PDZ domain-peptide interactions from primary sequences.** *Nat Biotechnol* 2008, **26**:1041-1045.
  46. Hui S, Bader GD: **Proteome scanning to predict PDZ domain interactions using support vector machines.** *BMC Bioinformatics* 2010, **11**:507.
  47. Shao X, Tan CS, Voss C, Li SS, Deng N, Bader GD: **A regression framework incorporating quantitative and negative interaction data improves quantitative prediction of PDZ domain-peptide interaction from primary sequence.** *Bioinformatics* 2011, **27**:383-390.
  48. te Velthuis AJ, Sakalis PA, Fowler DA, Bagowski CP: **Genome-wide analysis of PDZ domain binding reveals inherent functional overlap within the PDZ interaction network.** *PLoS One* 2011, **6**:e16047.
  49. Luck K, Travé G: **Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions.** *Bioinformatics* 2011, **27**:899-902.
  50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
  51. Lunn ML, Nassirpour R, Arrabit C, Tan J, McLeod I, Arias CM, Sawchenko PE, Yates JR, Slesinger PA: **A unique sorting nexin regulates trafficking of potassium channels via a PDZ domain interaction.** *Nat Neurosci* 2007, **10**:1249-1259.
  52. Dong H, O'Brien RJ, Fung ET, Lanahan AA, Worley PF, Huganir RL: **GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors.** *Nature* 1997, **386**:279-284.
  53. Borg JP, Marchetto S, Le Bivic A, Ollendorff V, Jaulin-Bastard F, Saito H, Fournier E, Adélaïde J, Margolis B, Birnbaum D: **ERBIN: a basolateral PDZ protein that interacts with the mammalian ERBB2/HER2 receptor.** *Nat Cell Biol* 2000, **2**:407-414.
  54. Jaulin-Bastard F, Saito H, Le Bivic A, Ollendorff V, Marchetto S, Birnbaum D, Borg JP: **The ERBB2/HER2 receptor differentially interacts with ERBIN and PICK1 PSD-95/DLG/ZO-1 domain proteins.** *J Biol Chem* 2001, **276**:15256-15263.
  55. Richter M, Murai KK, Bourgin C, Pak DT, Pasquale EB: **The EphA4 receptor regulates neuronal morphology through SPAR-mediated inactivation of Rap GTPases.** *J Neurosci* 2007, **27**:14205-14215.
  56. Zigman M, Cayouette M, Charalambous C, Schleiffer A, Hoeller O, Dunican D, McCudden CR, Firnberg N, Barres BA, Siderovski DP, Knoblich JA: **Mammalian inscuteable regulates spindle orientation and cell fate in the developing retina.** *Neuron* 2005, **48**:539-545.
  57. Matsuda S, Mikawa S, Hirai H: **Phosphorylation of serine-880 in GluR2 by protein kinase C prevents its C terminus from binding with glutamate receptor-interacting protein.** *J Neurochem* 1999, **73**:1765-1768.
  58. Javier RT: **Cell polarity proteins: common targets for tumorigenic human viruses.** *Oncogene* 2008, **27**:7031-7046.
  59. Jackson D, Hossain MJ, Hickman D, Perez DR, Lamb RA: **A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity.** *Proc Natl Acad Sci USA* 2008, **105**:4381-4386.
  60. Thomas M, Kranjec C, Nagasaka K, Matlashewski G, Banks L: **Analysis of the PDZ binding specificities of Influenza A virus NS1 proteins.** *Viral J* 2011, **8**:25.
  61. Liu H, Golebiewski L, Dow EC, Krug RM, Javier RT, Rice AP: **The ESEV PDZ-binding motif of the avian influenza A virus NS1 protein protects infected cells from apoptosis by directly targeting Scribble.** *J Virol* 2010, **84**:11164-11174.
  62. Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG: **Initial genetic characterization of the 1918 "Spanish" influenza virus.** *Science* 1997, **275**:1793-1796.
  63. Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, Finkelstein DB, Xu X, Wang J, Ma J, Fan Y: **Large-scale sequence analysis of avian influenza isolates.** *Science* 2006, **311**:1576.
  64. Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E, Howe K, Jackson DK, Palmer S, Roberts AN, Sims S, Stewart CA, Traherne JA, Trevanion S, Wilming L, Rogers J, de Jong PJ, Elliott JF, Sawcer S, Todd JA, Trowsdale J, Beck S: **Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project.** *Immunogenetics* 2008, **60**:1-18.
  65. Alibés A, Yankilevich P, Cañada A, Díaz-Uriarte R: **IDconverter and IDcLight: conversion and annotation of gene and protein IDs.** *BMC Bioinformatics* 2007, **8**:9.
  66. PIPE2. [<http://pipe2.systemsbio.net/PIPE2/>].

doi:10.1186/1471-2164-12-300

**Cite this article as:** Chimura et al.: Evolutionarily conserved bias of amino-acid usage refines the definition of PDZ-binding motif. *BMC Genomics* 2011 **12**:300.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

