BIOINFORMATICS DISCOVERY NOTE Vol. 27 no. 7 2011, pages 899–902
doi:10.1093/bioinformatics/btr060

Sequence analysis Advance Access publication February 7, 2011

Phage display can select over-hydrophobic sequences that may impair prediction of natural domain–peptide interactions

Katja Luck* and Gilles Travé*
Oncoproteins, Unité CNRS-UDS UMR 7242, Ecole Supérieure de Biotechnologie de Strasbourg, 1, Bd Sébastien Brant, BP 10413, 67412 Illkirch - Cedex, France
Associate Editor: Alex Bateman

ABSTRACT

Motivation: The phage display peptide selection approach is widely used for defining binding specificities of globular domains. PDZ domains recognize partner proteins via C-terminal motifs and are often used as a model for interaction predictions. Here, we investigated to which extent phage display data that were recently published for 54 human PDZ domains can be applied to the prediction of human PDZ–peptide interactions.

Results: Promising predictions were obtained for one-third of the 54 PDZ domains. For the other two-thirds, we detected in the phage display peptides an important bias for hydrophobic amino acids that seemed to impair correct predictions. Therefore, phage display-selected peptides may be over-hydrophobic and of high affinity, while natural interaction motifs are rather hydrophilic and mostly combine low affinity with high specificity. We suggest that potential amino acid composition bias should systematically be investigated when applying phage display data to the prediction of specific natural domain–linear motif interactions.

Contact: katja.luck@unistra.fr; gilles.trave@unistra.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 22, 2010; revised on January 28, 2011; accepted on February 1, 2011

1 INTRODUCTION

Many protein complexes that function in cellular regulation and signalling are assembled by multiple linear motif–globular domain interactions, which are mostly specific, yet of low affinity (Diella *et al.*, 2008). One well studied example of such interactions consists of the PDZ domains, which mainly recognize linear motifs at the extreme C-terminus of partner proteins (Doyle *et al.*, 1996). PDZs are implicated in the regulation of cell polarity, tight junctions, intercellular communication and neuronal synapses (Nourry *et al.*, 2003). The last residue (referred to as position p0) in PDZ-binding motifs usually is Val or Leu. The third last peptide residue (position p-2) can be either Thr or Ser, hydrophobic or Glu or Asp, thereby defining three main categories of PDZ-binding motifs (Songyang *et al.*, 1997) (Stricker *et al.*, 1997). These characteristics make PDZ–binding motifs relatively easy to predict. However, the correct prediction of PDZ domain binding specificities, i.e. prediction of which PDZ-binding motif will bind to which PDZ domain, remains challenging and numerous approaches have been proposed to tackle this problem (Brannetti *et al.*, 2001; Chen *et al.*, 2008; Hui and Bader, 2010; Kalyoncu *et al.*, 2010; Schillinger *et al.*, 2009; Smith and Kortemme, 2010).

Most predictors rely on prior experimental knowledge about binding preferences between peptides and globular domains. Phage peptide display has been widely used to provide such information (Sidhu *et al.*, 2003). This approach is based on selecting, out of a library of billions of peptides expressed on the surface of bacteriophages, a limited number of peptides that bind strongly to a given protein attached to a solid support. Several phage display studies have been performed on particular PDZ domains derived from the proteins MAGI1 (Fuh *et al.*, 2000), INADL (Vaccaro *et al.*, 2001), PDZRhoGEF and LARG (Smietana *et al.*, 2008), MUPP1 and DLG4 (Sharma *et al.*, 2009), PTP-BL (van den Berk *et al.*, 2007), Erbin (Skelton *et al.*, 2003), HtrA1 and HtrA3 (Runyon *et al.*, 2007). Tonikian *et al.* (2008) applied phage display in a high-throughput manner to determine and compare binding preferences of 28 *Caenorhabditis elegans* and 54 *Homo sapiens* PDZ domains. The data obtained in this study represent a highly valuable resource that allows to test the general application of phage display data to predictions of natural PDZ–protein interactions using position-specific scoring matrices (PSSMs). This approach was validated on a few PDZ domains in the study of Tonikian *et al.* (2008), and the phage display data were subsequently used in several recent studies for predictions of natural PDZ–peptide interactions (Hui and Bader, 2010; Smith and Kortemme, 2010). PDZ phage display data have also been used as test data in the 'DREAM4 Peptide Recognition Domain Specificity Prediction' challenge (Smith and Kortemme, 2010). Thus, phage display is supposed to capture accurately the binding specificities of domain–linear motif interactions.

Here, we performed and evaluated predictions of human PDZ–peptide interactions using the phage display data of Tonikian *et al.* (2008). Promising predictions were obtained for one-third of the 54 PDZ domains. In contrast, for the other two-thirds of the PDZ domains we detected important bias for hydrophobic amino acids in the phage display peptides that will probably impair the correct prediction of naturally occurring PDZ-binding peptides. We suggest that utilization of phage display data for prediction of natural binders should systematically involve prior analysis of potential sequencial bias in the data.

2 RESULTS

2.1 Prediction of natural PDZ–peptide interactions using phage display data

We searched the human proteome for C-termini of five residues in length that are likely to bind to the 54 human PDZ domains

*To whom correspondence should be addressed.

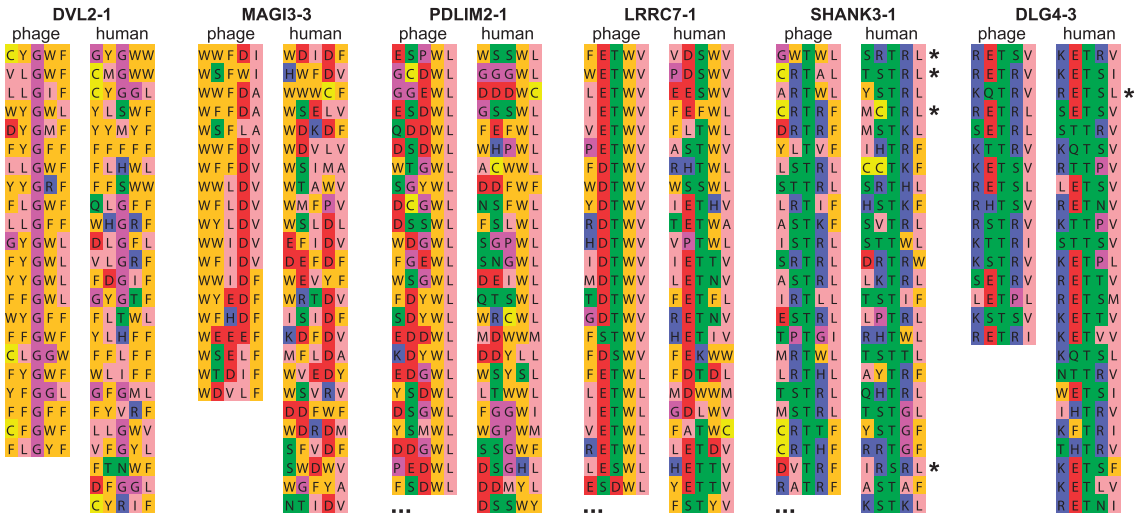

Fig. 1. Prediction of natural binders to PDZ domains using phage display data of Tonikian *et al.* (2008). The last five residues of phage display (PD) peptides together with predicted best-matching human C-terminal peptides are shown for six PDZ domains, ordered from the left to the right from most hydrophobic to most hydrophilic PD peptides. PD lists that were too long for being entirely displayed are indicated by '…'. Asterisks indicate human C-termini that are identical to C-termini of corresponding PD peptides. Colour code: gold = aromatic, light pink = hydrophobic, pink = G or P, green = polar, red = acidic, blue = basic, yellow = C. [Figure made with Jalview (Waterhouse *et al.*, 2009).]

for which Tonikian *et al.* (2008) obtained phage peptides. To this aim, we constructed a PSSM (see Supplementary Material) for each list of peptides selected by the 54 human PDZ domains. A PSSM captures the occurrence of each amino acid at each position within a list of aligned sequences. This allowed us to describe, for each PDZ domain, a sequence profile defined by the phage peptides that bound to it. Using each of the 54 PSSMs obtained in that way, we selected the 25 C-termini of human proteins that matched best to the sequence profile of the corresponding phage peptide list (reported in Supplementary Dataset S1). Within these sets of 25 most similar human C-termini, a number of peptides were actually found to be identical to the corresponding phage peptides (reported in Supplementary Dataset S2). Several instances of this search are shown in Figure 1.

Some of the phage peptide lists seemed to be anomalously enriched in hydrophobic amino acids (such as DVL2-1 and MAGI3-3 in Fig. 1). We used the hydrophobicity index of Kidera *et al.* (1985) to compute the average hydrophobicity (see Supplementary Material) of each list of phage peptides and ranked these lists from the most hydrophobic to the most hydrophilic (Fig. 2A). We observed that more identical human C-termini were returned for the hydrophilic phage peptide lists than for the hydrophobic ones (Fig. 2B, compare left side and right side of the plot, *P*-value <1.0E-6).

Next, we calculated an additional PSSM for each list of 25 human C-termini and determined its distance to the PSSM of the corresponding phage peptides (see Supplementary Material). The better the 25 human C-termini match to the sequence profile of the phage peptides, the more similar (less distant) the corresponding two PSSMs should be to each other and the more likely the 25 human C-termini would be to bind the corresponding PDZ domain. We observed that the more similar the PSSMs, the more hydrophilic the corresponding phage peptides

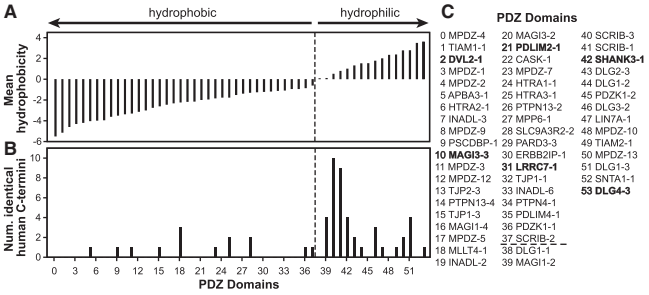

Fig. 2. Analysis of PDZ–peptide interaction predictions. (**A**) The 54 PDZ domains used by Tonikian *et al.* (2008) were ranked based on the mean hydrophobicity of their corresponding phage display (PD) peptides, from the most hydrophobic to the most hydrophilic. This ranking is conserved for plot B. The vertical dashed line separates hydrophobic from hydrophilic peptide lists. (**B**) Numbers of human C-termini that are identical to PD peptides are plotted for each PDZ domain. (**C**) PDZ domains (named as in Tonikian *et al.*) are listed based on the hydrophobicity of the PD peptide lists with numbers that were used in diagram A and B. Names in bold indicate PDZs that are shown in Figure 1.

(Pearson correlation coefficient of −0.51, *P*-value = 7.5E-5, see Supplementary Figure S1). This analysis indicates that the 25 best-matching human C-termini seem to better reproduce the sequence profile of the corresponding phage peptides when they are hydrophilic (see instances in Fig. 1).

2.2 Analysing the amino acid composition of phage display peptides

The above-mentioned analysis has also revealed that about two-thirds of the human PDZ domains used in the study of Tonikian *et al.* (2008) preferentially selected peptides of rather hydrophobic

Table 1. Comparison of mean hydrophobicity and W content of different peptide datasets

Source	Mean hydrophobicity[a]	% W at p-1[b]	Num peptides[c]
Tonikian *et al.*	−1.41	45.7	1390
C-terminome	0.64	1.9	26 904
PDZbase	0.92	5.6	233
Chen *et al.*	0.93	3.7	108

All peptides were reduced to a length of five residues.
[a]Calculated with index of Kidera *et al.* (1985), value of most hydrophilic peptide of length five: 9.35, most hydrophobic = −7.85.
[b]Percentage of peptides with Trp at peptide position p-1.
[c]Number of peptides.

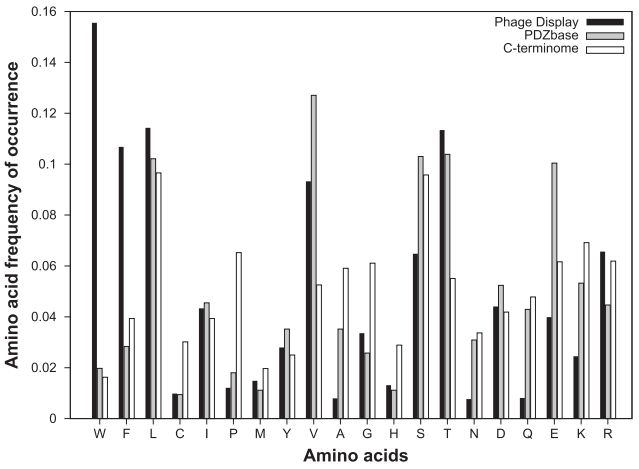

Fig. 3. Amino acid composition of phage peptides versus the human C-terminome and PDZ-binding peptides from the PDZbase. Amino acids are sorted from the most hydrophobic (left) to the most hydrophilic (right) according to the hydrophobicity scale of Kidera *et al.* (1985). All sequences were cut to a length of five residues.

character (Fig. 2A). We compared the mean hydrophobicity of the phage peptides to that of different peptide sets (Table 1, column 2): the C-terminome (all C-termini from the human proteome, assumed to reflect the general hydrophobicity of human C-terminal sequences, see Supplementary Material); the PDZbase [containing experimentally validated PDZ-binding peptides originating from various proteomes (Beuming *et al.*, 2005)]; and the mouse PDZ-binding peptides published by Chen *et al.* (2008). All three sets, which represent naturally occurring sequences, display a hydrophilic character in contrast to the phage peptides, which are in average markedly hydrophobic. In particular, the natural PDZ-binding peptides (derived both from Chen *et al.* and PDZbase) are significantly more hydrophilic than the phage peptides (2-sample *t*-test, *P*-value = 5E-49).

We further analysed this discrepancy by calculating the frequency of occurrence of each of the 20 amino acids in the phage peptides, the PDZ-binding peptides from the PDZbase and the human C-terminome. The phage sequences are strongly enriched in the aromatic amino acids W and F (Fig. 3). We also computed the amino acid frequencies in these three datasets for each of the five peptide positions separately (Supplementary Figure S2). All positions show an enrichment in hydrophobic amino acids, in particular aromatic residues, for phage peptides. This trend could not be observed for natural PDZ-binding peptides from the PDZbase or in general human C-terminal sequences. For instance, position p-1 is occupied by W in almost 50% of the phage peptides versus only 2% of the human C-terminal peptides, and 6% of the PDZ-binding peptides of the PDZbase (Table 1, column 3, Fisher's exact test *P*-value <2.2E-16). Interestingly, positions p-1 and p-3 and to a lesser extent p-4 seem even to be under-represented for polar or charged residues in phage peptides in contrast to the two other sets (Supplementary Fig. S2).

These results indicate that the hydrophobic character of phage PDZ-binding peptides does not correspond to sequence properties observed in natural PDZ-binding peptides and general human C-terminal sequences. This might explain why our search for best-matching human C-termini to the sequence profile of phage peptides seems to perform better for PDZ domains that preferentially select hydrophilic phage peptides.

3 DISCUSSION

Here, we addressed the problem of predicting natural PDZ–peptide interactions using phage display data. We observed that phage peptide lists of Tonikian *et al.* (2008) can be classified on the basis of their hydrophobic character. Human C-termini matched better to the sequence profiles defined by the phage peptides, when they were hydrophilic. More specifically, human C-termini that are identical to phage peptides could be found more frequently for hydrophilic phage peptides. In addition, we realised that in average the phage peptides were much more hydrophobic than published natural PDZ-binding sequences as well as human C-termini in general. In particular, the phage display data showed a very strong preference for the largest aromatic amino acid Trp at peptide position p-1. All these results indicate that prediction of interactions between PDZs and naturally occurring peptides perform better when based on hydrophilic phage peptides.

It should be noted that short linear interaction motifs (Slims) have been found to display a particular amino acid composition, which distinguishes them from both folded and disordered regions (Fuxreiter *et al.*, 2007). The least conserved positions in Slims are usually non-hydrophobic, whereas the highest conserved positions are very often occupied by hydrophobic and charged amino acids. Indeed, published PDZ ligands generally agree with this trend, since the canonical PDZ-binding motif pattern consists of a hydrophobic (usually not aromatic) amino acid at peptide position p0 and Thr/Ser, hydrophobic (usually not aromatic) or Glu/Asp at position p-2. The phage display procedure of Tonikian *et al.* often selected such characteristics at positions p0 and p-2, but the other less conserved positions (p-1, p-3 and p-4) were, for two-thirds of the PDZ domains tested, very frequently hydrophobic, thereby deviating from sequence characteristics of Slims.

Biological interactions are characterized both by their affinity and specificity. Affinity represents absolute interaction strength, whereas specificity is a relative property derived from the comparison of interaction strengths of different interacting partners. For instance, if a PDZ domain binds with higher (but not necessarily high) affinity to a few peptides than to all others, it will be specific. Molecular dynamic studies (Basdevant *et al.*, 2006) have indicated that hydrophobic interactions are the most important force contributing

to PDZ–peptide affinity, and Beuming *et al.* (2009) have suggested that Trp at p-1 contributes strongly to the affnity of C-terminal peptides to Erbin PDZ domain via hydrophobic effects. In this regard, the phage display procedure, being mainly affinity driven, may have selected hydrophobic and especially aromatic amino acids at the least conserved positions of the PDZ-binding motif. However, transient interactions are required for PDZ-mediated cell signalling. In such a context, PDZ-binding hydrophobic sequences might turn out to be counter-productive due to an excessively high affinity. In addition, interactions involved in signalling also require specificity that might not be conferred by hydrophobic binders. Indeed, by examining SPOT data from Wiedemann *et al.* (2004), we observed that a 'super-binding peptide' with Trp at p-1 displaying high affinity for Erbin PDZ domain seemed to be robust against mutations at other peptide positions indicating a strong contribution of Trp to the binding affinity. Hence, the Trp at p-1, and hydrophobic residues at least conserved positions in general, would probably allow for more putative interaction partners to a PDZ domain and would make specific recognition impossible. In summary, it seems that the phage display approach has a tendency to select high affinity binders presenting artificial sequence features in contrast to evolution rather selecting for specific binders in the context of Slims. We notice that a similar conclusion has independently been drawn in a recent phage display study by Ernst *et al.* (2010). While this property of phage display may be useful for drug design or synthetic biology, it may limit its application for predicting natural domain–motif interactions. Recently, a promising approach was proposed to modify the phage display experimental protocol towards a procedure that will rather select specific than high affinity peptides (Hoffmann *et al.*, 2010).

Our study indicates that PDZ–peptide interaction predictions based on hydrophobic phage peptides should be considered carefully, especially with regard to specific, natural interactions, whereas predictions of interaction networks based on hydrophilic phage peptides are promising. We hypothesize that similar constraints in phage display data might also arise in the context of other types of domain–linear motif interactions. Given the wide use of phage display for the determination of binding specificities of domain–linear motif interactions, the problems addressed here might apply to many other studies as well.

ACKNOWLEDGEMENTS

We thank N. Davey, B. Kieffer and T. Gibson for helpful discussion of the manuscript.

Funding: Région Alsace (K.L.); CDE (K.L.); CNRS; ARC (Grant nr. 3171); and ANR (project ANR-MIME-2007 EPI-HPV-3D).

Conflict of Interest: none declared.

REFERENCES

Basdevant,N. *et al*. (2006) Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J. Am. Chem. Soc.*, **128**, 12766–12777.

Beuming,T. *et al*. (2005) PDZbase: a protein-protein interaction database for PDZ-domains. *Bioinformatics*, **21**, 827–828.

Beuming,T. *et al*. (2009) High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci.*, **18**, 1609–1619.

Brannetti,B. *et al*. (2001) iSPOT: a web tool for the analysis and recognition of protein domain specificity. *Comp. Funct. Genomics*, **2**, 314–318.

Chen,J.R. *et al*. (2008) Predicting PDZ domain-peptide interactions from primary sequences. *Nat. Biotechnol.*, **26**, 1041–1045.

Diella,F. *et al*. (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci.*, **13**, 6580–6603.

Doyle,D.A. *et al*. (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell*, **85**, 1067–1076.

Ernst,A. *et al*. (2010) Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol. Biosyst.*, **6**, 1782–1790.

Fuh,G. *et al*. (2000) Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J. Biol. Chem.*, **275**, 21486–21491.

Fuxreiter,M. *et al*. (2007) Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, **23**, 950–956.

Hoffmann,S. *et al*. (2010) Competitively selected protein ligands pay their increase in specificity by a decrease in affinity. *Mol. Biosyst.*, **6**, 126–133.

Hui,S. and Bader,G.D. (2010) Proteome scanning to predict PDZ domain interactions using support vector machines. *BMC Bioinformatics*, **11**, 507.

Kalyoncu,S. *et al*. (2010) Interaction prediction and classification of PDZ domains. *BMC Bioinformatics*, **11**, 357.

Kidera,A. *et al*. (1985) Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J. Protein Chem.*, **4**, 23–55.

Nourry,C. *et al*. (2003) PDZ domain proteins: plug and play! *Sci. STKE*, **2003**, RE7.

Runyon,S.T. *et al*. (2007) Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3. *Protein Sci.*, **16**, 2454–2471.

Schillinger,C. *et al*. (2009) Domain interaction footprint: a multi-classification approach to predict domain-peptide interactions. *Bioinformatics*, **25**, 1632–1639.

Sharma,S.C. *et al*. (2009) T7 phage display as a method of peptide ligand discovery for PDZ domain proteins. *Biopolymers*, **92**, 183–193.

Sidhu,S.S. *et al*. (2003) Exploring protein-protein interactions with phage display. *Chembiochem*, **4**, 14–25.

Skelton,N.J. *et al*. (2003) Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the erbin PDZ domain. *J. Biol. Chem.*, **278**, 7645–7654.

Smietana,K. *et al*. (2008) Degenerate specificity of PDZ domains from RhoA-specific nucleotide exchange factors PDZRhoGEF and LARG. *Acta Biochim. Pol.*, **55**, 269–280.

Smith,C.A. and Kortemme,T. (2010) Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J. Mol. Biol.*, **402**, 460–474.

Songyang,Z. *et al*. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science*, **275**, 73–77.

Stricker,N.L. *et al*. (1997) PDZ domain of neuronal nitric oxide synthase recognizes novel c-terminal peptide sequences. *Nat. Biotechnol.*, **15**, 336–342.

Tonikian,R. *et al*. (2008) A specificity map for the PDZ domain family. *PLoS Biol.*, **6**, e239.

Vaccaro,P. *et al*. (2001) Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from Drosophila melanogaster. *J. Biol. Chem.*, **276**, 42122–42130.

van den Berk,L.C.J. *et al*. (2007) An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry*, **46**, 13629–13637.

Waterhouse,A.M. *et al*. (2009) Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

Wiedemann,U. *et al*. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol.*, **343**, 703–718.