Published in final edited form as:
J Am Chem Soc. 2006 May 3; 128(17): 5913–5922. doi:10.1021/ja060943h.

Uncovering Quantitative Protein Interaction Networks for Mouse PDZ Domains using Protein Microarrays

Michael A. Stiffler, **Viara P. Grantcharova**[§], **Mark Sevecka**, and **Gavin MacBeath**[*]
Department of Chemistry and Chemical Biology Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138

Abstract

One of the principle challenges in systems biology is to uncover the networks of protein-protein interactions that underlie most biological processes. To date, experimental efforts directed at this problem have largely produced only qualitative networks that are replete with false positives and false negatives. Here, we describe a domain-centered approach – compatible with genome-wide investigations – that enables us to measure the equilibrium dissociation constant (K_D) of recombinant PDZ domains for fluorescently-labeled peptides that represent physiologically-relevant binding partners. Using a pilot set of 22 PDZ domains, 4 PDZ domain clusters, and 20 peptides, we define a gold standard dataset by determining the K_D for all 520 PDZ-peptide combinations using fluorescence polarization. We then show that microarrays of PDZ domains identify interactions of moderate to high affinity ($K_D \leq 10$ μM) in a high-throughput format with a false positive rate of 14% and a false negative rate of 14%. By combining the throughput of protein microarrays with the fidelity of fluorescence polarization, our domain/peptide-based strategy yields a quantitative network that faithfully recapitulates 85% of previously reported interactions and uncovers new biophysical interactions, many of which occur between proteins that are co-expressed. From a broader perspective, the selectivity data produced by this effort reveal a strong concordance between protein sequence and protein function, supporting a model in which interaction networks evolve through small steps that do not involve dramatic rewiring of the network.

Introduction

Most eukaryotic proteins that receive and process signals are constructed in a modular fashion from a combination of interaction and catalytic domains.[1] Interaction domains mediate the formation of multiprotein complexes that confine signaling proteins to appropriate subcellular locations and help determine the specificity of enzyme-substrate interactions. One of the primary challenges of systems biology is to define the function of protein interaction domains on a genome-wide scale and thus uncover the networks of protein-protein interactions that underlie complex biological processes.

To date, most protein interaction networks that have been defined experimentally are Boolean in nature: proteins are reported either to ‘interact’ or ‘not interact’. For example, the yeast two-hybrid assay has been used on a large scale to identify interactions in the proteomes of yeast[2,3] and *Caenorhabditis elegans*,[4] as well as to focus on interactions mediated by coiled-coil domains.[5] Multiprotein complexes have also been uncovered on a large scale by identifying proteins that co-purify with selected bait proteins using mass spectrometry.[6–8] In every case, the resulting network does not provide information about the strength of the protein-

E-mail: macbeath@chemistry.harvard.edu.
[§]Present address: Merrimack Pharmaceuticals, 101 Binney Street, Cambridge, MA 02142.

protein interactions and hence cannot be used to predict how the network is likely to change as a function of cell state.

To obtain quantitative protein interaction networks, it is necessary not only to identify interactions, but to measure affinities as well. This additional step serves at least four purposes. First, the additional rigor required to quantify interactions minimizes the amount of incorrect information in the resulting dataset. Most high-throughput methods suffer from alarmingly high rates of false positives and false negatives,[9–12] limiting their usefulness in generating biological hypotheses and calling into question conclusions about network topology that are based on these data.[13,14] Second, determining binding affinities helps to prioritize which interactions are more likely to be physiologically relevant among a series of biophysical interactions. Third, quantitative information can serve as a powerful training set for computational studies aimed at predicting protein-protein interactions,[15] as well as for modeling biological processes such as signal transduction.[16,17] Finally, large-scale quantitative investigations can reveal the dynamic nature of protein interaction networks. For example, we recently described a quantitative network for the human ErbB receptors and found that the extent to which each receptor becomes more promiscuous (more interconnected) when over-expressed correlates with its oncogenic potential.[18] This link between network dynamics and cancer is only revealed by measuring binding affinities.

To uncover quantitative interaction networks, we need technologies that enable large-scale investigations, but at the same time yield reliable data. Since microarray technology enables a large number of proteins to be queried with a large number of probes, it is ideally suited to studying the binding selectivity of entire families of protein interaction domains with respect to large collections of binding targets.[19] We have recently used microarrays of human Src homology 2 (SH2) and Phosphotyrosine Binding (PTB) domains to identify and quantify interactions with phosphopeptides representing physiological sites of tyrosine phosphorylation on receptor tyrosine kinases.[18] It therefore seemed reasonable to explore a similar strategy to study interactions between mouse PDZ domains and their physiological targets.

Among the many interaction domains identified in the past decade, PDZ's are one of the most frequently encountered. PDZ domains are approximately 90 residues long and were first identified as regions of sequence homology in diverse signaling proteins.[20,21] The primary function of PDZ domains is to mediate protein-protein interactions by recognizing the C-termini of their target proteins in a sequence-specific fashion.[22–24] They are often found in combination with other interaction modules (such as WW, SH3, and PTB domains) and help direct the specificity of receptor tyrosine kinases, establish and maintain cell polarity, direct protein trafficking, and coordinate synaptic signaling.[25–28] Their importance is underscored by the severe neuronal and developmental phenotypes observed in PDZ knockout mice[29–33] and by their implication in human congenital diseases.[34–36]

The enormous diversity of PDZ domain function is manifest in their abundance; at current count there are over 250 such domains in the mouse genome.[37,38] This implicates PDZ domains in the wiring of a large number of proteins in molecular networks from the membrane to the nucleus. With the ultimate goal of uncovering a genome-wide, quantitative interaction network for mouse PDZ domains, we set out to determine if protein microarray technology could form the backbone for these studies. Most importantly, we wished to define the fidelity of this technique since this information is critical for all future investigations, not only of PDZ domains, but of all other classes of protein interaction modules as well.

Results and Discussion

In order to assess the suitability of using protein microarray technology to study PDZ-mediated interactions, we began by constructing a list of 22 PDZ domains for which one or more cellular ligands have been identified (Table 1 and Figure S1 in the Supporting Information). PDZ domains have been classified into three general classes based on their preference for the three C-terminal residues of their binding partners: class I domains recognize the consensus sequence S/T-X-V/L (where X is any amino acid); class II domains prefer ψ-X-ψ (where ψ is a hydrophobic amino acid); and class III domains prefer E/D-X-ψ.[28,39] To ensure diverse representation, we included members of all three classes. Since many PDZ-containing proteins feature more than one PDZ domain, it is fairly common to find two or more PDZ's clustered together in the linear sequence of their host protein with very little intervening sequence. For some PDZ domains that occur in such clusters, it has been found that their adjacent domains contribute to their structural stability or have an effect on their binding selectivity.[40–43] We therefore included in our list four clusters of two or three adjacent domains, as well as the individual domains.

With the intended goal of extending these studies to all PDZ domains, we designed a general strategy to clone, express, and purify each domain or domain cluster. To abstract individual domains from their full-length proteins, we determined their boundaries by aligning their sequences and by using available structural information (see Materials and Methods). Cloning of the PDZ domains was accomplished by amplifying their coding regions from mouse cDNA using the polymerase chain reaction (PCR). We obtained products with the expected molecular weight for all 22 individual PDZ's, as well as for each cluster. The PCR products were transferred into a suitable vector by topoisomerase I-mediated directional cloning and each clone was verified by DNA sequencing. The coding regions were then transferred by λ-recombinase-mediated directional subcloning into an *Escherichia coli* expression vector that appends thioredoxin and His_6-tags to the N-terminus of the resulting protein. All of these steps are completely general and easily applied to entire families of protein domains using simple automation.

Focusing on protein interaction domains, rather than full-length proteins, greatly increases the likelihood of obtaining reasonable quantities of soluble, monomeric, recombinant protein from bacteria. All 26 PDZ domains and domain clusters were produced in *E. coli* and purified to homogeneity in a single step by immobilized metal affinity chromatography. Intact, pure protein was recovered for all 26 constructs as judged by SDS-polyacrylamide gel electrophoresis (Figure S2 in the Supporting Information) and all 26 constructs were monomeric as judged by analytical size exclusion column chromatography (Table S1 in the Supporting Information). The success of this effort highlights the advantage of domain-oriented functional proteomics.

Based on this representative set of PDZ domains, we assembled a list of 20 proteins that have been shown to interact, via their C-termini, with one or more of our PDZ domains (Table 2 and Table S2 in the Supporting Information). Previous studies have shown that some PDZ domains exhibit binding selectivity out to the −8 position of their target protein (ninth residue from the C-terminus).[39] We therefore synthesized peptides whose sequences correspond to the last 10 residues of their parent protein. The tripeptide sequence NNG was added to the N-terminus of each peptide to increase its water solubility and each peptide was capped with 5-(and-6)-carboxytetramethylrhodamine [5(6)-TAMRA] at its N-terminus prior to deprotection and cleavage from the resin. 5(6)-TAMRA serves both as a chromophore for quantification purposes and as a fluorophore for visualization on the protein microarrays.

As with the PDZ domains, we sought to make our set of PDZ ligands as diverse as possible. Contained within our set of peptides are ligands for each class of PDZ domains: Claudin1 and Mel1a/b bind to class III PDZ's; AN2, Parkin, EphrinB1/2, Nrxn1/2, and GluR2 bind to class II PDZ's; and the remaining 13 peptides bind to class I PDZ's. The PDZ ligands include promiscuous proteins and highly selective proteins. For example, the protein Cript is known to interact with nine different PDZ domains included in our study, while the protein Dlgap1/2/3 has only been described to bind one PDZ domain – that of Shank3 (Table S1). Our peptides also represent proteins with diverse functionality. They include two glutamate receptor subunits (NMDAR2A and GluR2), two G-protein coupled receptors (Frizzled and Mel1a/b), and several ion channel subunits (Scn4a, Scn5a, Stargazin, Kv1.4, and Kir2.1). The proteins Nrxn1/2, Claudin1 and AN2 play a role in cell adhesion and mobility while the remaining proteins are cytoplasmic or membrane-bound adapter and regulatory proteins (Cript, KIF1B, Mapk12, L-glutaminase, Cnksr2, EphrinB1/2, and Dlgap1/2/3).

PDZ domain microarrays

We have previously described methods to prepare microarrays of functionally active proteins on aldehyde-presenting glass surfaces.[19] In order to make these methods compatible with high-throughput investigations, we developed a strategy to array proteins in individual wells of 96-well microtiter plates. Although some commercial arrayers are able to print directly into microtiter plates, this process is slow and the width of the microarraying pins or glass tips limits access to only the central region of each well. We therefore developed a method in which we array our proteins onto flat glass substrates that are cut to a size that spans all the wells of a microtiter plate (112.5 mm × 74.5 mm). 96 identical microarrays are printed on each substrate in the pattern of a 96-well plate (Figure 1A). The glass is then permanently attached to the bottom of a bottomless microtiter plate using an intervening silicone gasket coated on both sides with a strong, biocompatible adhesive (Figure 1B). The result is a microtiter plate containing a protein microarray in each well.

Using this technology, we arrayed our 26 PDZ constructs, as well as recombinant thioredoxin (negative control), on aldehyde-presenting glass plates. The proteins were all spotted at a concentration of 40 μM in a buffer that contained 20% glycerol (v/v) to prevent evaporation. Samples were printed in triplicate and a small amount (100 nM) of cyanine-5 (Cy5)-labeled bovine serum albumin (BSA) was included in each sample to facilitate image analysis. After a 1-h incubation at room temperature, the unreacted aldehydes were quenched and the surfaces blocked by the addition of buffer containing BSA. The arrays were then probed with a single concentration of each peptide (1 μM) and incubated at room temperature for one hour. The arrays were washed, dried, and scanned for both Cy5 and 5(6)-TAMRA fluorescence. The Cy5 image was used to define the location of each spot and the mean fluorescence of replicate spots in the 5(6)-TAMRA image was determined for each PDZ construct (see, for example, Figure 2). This value was then divided by the mean intensity of control spots (thioredoxin) to determine the 'fold-over-background' ratio, or FOB (Table S3 in the Supporting Information).

Having identified specific interactions in this manner, we sought to quantify the strength of each interaction. We have recently shown that apparent K_D's can be measured by probing protein microarrays with different concentrations of a fluorescent ligand and fitting the resulting data to an equation that describes saturation binding.[18] Dissociation constants obtained in this way agree very well with those obtained using surface plasmon resonance (typically within two-fold). This approach worked well to quantify interactions mediated by SH2 and PTB domains, which tend to bind their targets with submicromolar affinities. We were unable to measure K_D's above 2 μM, however, since nonspecific binding of the fluorescent probe to the array surface becomes prohibitively high at probe concentrations above 5 μM. Unlike SH2 and PTB domains, most physiologically relevant interactions mediated by

PDZ domains fall in the low micromolar range: often between 1 and 10 μM.[28] We noticed, however, that although we were unable to quantify weak interactions using protein microarrays, we were able to detect them. Based on this observation, we designed the following two-step strategy for defining quantitative PDZ interaction networks:

1. screen microarrays of PDZ domains with fluorescent peptides to identify domain-peptide interactions in high-throughput; and
2. measure the strength of interactions detected on the arrays using a solution-phase, fluorescence polarization assay.

This strategy exploits the power of microarray technology to screen every possible PDZ-peptide combination in a rapid and economical fashion. It then takes advantage of the already assembled reagents to quantify interactions in a high-fidelity but lower throughput solution-phase assay. The success of this strategy depends on the fidelity of the initial screen. How successful are protein microarrays in identifying relatively low-affinity interactions? If the assay is too insensitive, many interactions will be missed (high rate of false negatives). If, on the other hand, the assay is too permissive, false positives will abound and the advantage afforded by the microarrays will be lost. We therefore set out to determine, in a rigorous fashion, the rate of false positives and false negatives on the protein microarrays, and how these rates vary with binding affinity.

High-throughput fluorescence polarization

To determine the rate of false positives and false negatives, we need a 'gold standard' against which the arrays can be measured. Since our strategy relies on fluorescence polarization (FP) to confirm and quantify interactions, this solution-phase assay serves as an appropriate standard. We therefore developed an FP-based assay compatible with large-scale investigations. Equilibrium dissociation constants were determined in 384-well microtiter plates by introducing purified domains into the top row of each plate and preparing two-fold serial dilutions down each column. A fixed, low concentration of fluorescent peptide was then introduced into each well of the plate using a 96-channel pipetting robot. In this way, 24 affinity constants were determined in each plate, with 16 points per curve. The final concentrations of the PDZ domains ranged from 20 μM down to 0.6 nM. Following this strategy, we collected FP data for all 520 PDZ-peptide combinations (8,320 separate measurements). We then fit the data for each PDZ-peptide combination (FP, recorded as millipolarization units) to eq 1,[44]

$$FP = FP_{\max}\frac{K_D + [pep] + [PDZ] - \sqrt{(K_D + [pep] + [PDZ])^2 - 4[pep][PDZ]}}{2[pep]} + FP_0 \quad (1)$$

where FP_{max} is the maximum signal at saturation, FP_0 is the signal in the absence of PDZ domain, [*PDZ*] is the total concentration of PDZ domain, [*pep*] is the total concentration of peptide (20 nM), and K_D is the calculated equilibrium dissociation constant (see, for example, Figure 2; K_D's for all interactions are provided in Table S4 of the Supporting Information). An interaction was scored as 'specific' if it fit well to eq 1 ($R^2 \geq 0.95$), with a $K_D \leq 20$ μM and high signal ($FP \geq 15$ mP at 20 μM PDZ).

With these quantitative measurements in hand, we compared our microarray results, obtained by probing each array with a single concentration of fluorescent peptide (1 μM), with those obtained by FP. Since the microarrays provide a fold-over-background (FOB) ratio for each PDZ-peptide pair, molecules that 'interact' are defined as those with an FOB ratio that exceed some arbitrary threshold. At a given threshold, 'false positives' are interactions that exceed the threshold but are not present in the gold standard dataset; 'false negatives' are interactions that do not exceed the threshold, but are nevertheless found in the gold standard dataset. Using the

FP affinity data as our gold standard, we calculated the false positive and false negative rates for the PDZ microarrays at different FOB thresholds, ranging from 1.0 to 2.0 (Figure 3A). To assess the affinity limits of the microarrays, we performed this analysis using three different gold standard datasets: (1) interactions with $K_D \leq 5$ μM; (2) interactions with $K_D \leq 10$ μM; and (3) interactions with $K_D \leq 20$ μM (Figure 3A). As expected, the false positive rate decreased and the false negative rate increased as the FOB threshold was raised. Importantly, the false positive rate dropped steeply between FOB thresholds of 1.0 and 1.4, while the false negative rate exhibited a lag phase before climbing steadily. If we consider interactions with $K_D \leq 5$ μM (keeping in mind that we probed our microarrays with 1 μM peptide), we find that an FOB threshold of 1.4 produces 19% false positives and 6% false negatives. In addition, almost half (47%) of the 'false positives' in this analysis were, in fact, *bona fide* interactions with K_D's between 5 μM and 20 μM. These false positive/false negative rates compare very favorably with the yeast two-hybrid assay, where estimates of 50% false positives and 90% false negatives have been reported.[3,9–12,45] The low rates that we see here can be attributed to several factors. First, our assay is performed under very controlled *in vitro* conditions and so does not suffer from the noise introduced by biological systems. Second, the concentrations of both the proteins and the peptides are controlled and normalized in our assay, whereas protein expression levels vary substantially in cell-based assays. Finally, our assay is focused on a family of structurally and functionally related domains, while the rates reported above were estimated from studies of proteins with diverse structure and function. This again underscores the value of adopting a domain-oriented approach to functional proteomics.

Given that many physiological interactions mediated by PDZ domains fall in the 1–10 μM range, we were encouraged to see that a threshold of 1.4 was still able to identify 86% of interactions with $K_D \leq 10$ μM, with only 14% false positives (Figure 3A). When we extend our analysis to the very weak interactions ($K_D \leq 20$ μM), however, the false negative rate jumps to 34% (Figure 3A). This is not too surprising; it shows that protein microarrays, when probed with 1 μM ligand, miss a substantial number of interactions with K_D between 10 μM and 20 μM.

It is reasonable to think that our ability to detect weak interactions would be improved by probing the arrays with a higher concentration of fluorescent peptide. We therefore repeated the entire microarray experiment, but with the peptides at a concentration of 5 μM rather than 1 μM. Comparison with the FP dataset (Figure 3B) shows that, contrary to this prediction, the rate of false negatives actually increases, rather than decreases, while the rate of false positives remains largely unchanged. The reason for this seemly paradoxical observation is that a higher concentration of peptide results in elevated levels of nonspecific binding to the control protein (thioredoxin) and to the slide surface. The increase in nonspecific binding outpaces the increase in specific binding to the cognate PDZ domains. As a result, FOB values generally decrease, rather than increase, as the ligand concentration is raised from 1 μM to 5 μM. It is possible that improved surfaces would alleviate this effect to some extent. Nonspecific binding of the peptides to the control protein and to the non-active site regions of the target proteins, however, does not depend on surface chemistry and ultimately limits all assays of this nature.

It is often assumed that the intensity of spots on a protein microarray correlates with the affinity of interactions. To test this hypothesis, we plotted all of our interaction data, with K_D on the *y*-axis (as determined by FP) and microarray spot intensity on the *x*-axis (Figure 4). While it is true that bright spots generally represent high affinity interactions, little can be concluded about low intensity spots. The most reasonable explanation is that the intensity of spots on a microarray is a function not only of binding affinity, but also of the amount of active protein in the spot. Some proteins are destabilized on the glass surface, resulting in a low percentage of folded protein in the spot. Others may preferentially attach to the surface in a way that blocks their function. As a result, high affinity interactions can show up as weak spots. What is most

sobering about these results is that they were obtained under the most ideal of circumstances. All of the PDZ domains share a common fold and were printed at the same concentration (40 μM). If the concentrations of the proteins are not normalized prior to printing, and if proteins of diverse size, structure and function are studied together, it is likely that spot intensity will correlate even less with binding affinity and that the rate of false positives and false negatives will increase.

Comparison with previously reported interactions

So far, we have shown that microarrays of recombinant PDZ domains can be used to identify PDZ-peptide interactions with high fidelity, and that these interactions can be retested and quantified rapidly using fluorescence polarization. We have not yet addressed if our domain-based strategy effectively captures information about physiological protein-protein interactions. How well do abstracted domains substitute for full-length proteins and how well do synthetic peptides substitute for PDZ ligands? To address these questions, we performed an extensive search of the scientific literature to identify PDZ-mediated interactions that involve the domains and ligands used in our study (Table S2). In total, we found 85 such interactions, all but three of which have been narrowed down to the PDZ domain of interest. We then compared our biophysical interaction data, obtained by microarrays and FP, with this list (Figure 5). For this comparison, we used an FOB cutoff of 1.4 for the microarrays and a K_D cutoff of 20 μM for the FP data. Of the 85 interactions that were previously reported, we observed 72 interactions (85%) either by microarray or by FP. There were only seven cases where a previously reported interaction was observed by microarray technology but not by FP, supporting our assumption that FP serves as an appropriate standard by which to judge the microarrays. Interestingly, many of the microarray false positives (17 of 46 at the 20 μM threshold) arose from interactions with a single PDZ domain: PDZ3 of CIPP. If we exclude this ‘sticky’ domain from our analysis, the false positive rate drops from 12% to 9%. One of the advantages of screening multiple domains against multiple ligands is that trends like this can easily be identified. Domains that are prone to nonspecific interactions can be recognized as a densely populated row on the interaction matrix (Figure 5), while ligands that are prone to nonspecific interactions show up as densely populated columns.

In several cases, interactions were detected with tandem PDZ constructs but not with any of the corresponding isolated domains. For example, we detected specific binding between the AN2 peptide and the three domain cluster of SAP97, but did not detect any interaction between this peptide and the corresponding isolated domains. In instances like this, it is likely that the isolated domains lack structural stability, but the larger construct does not.[43] Cloning clusters of tightly coupled domains provides a way to study PDZ’s that do not behave well when abstracted from their native context. We anticipate that this strategy will also be important in other domain-based functional proteomics efforts.

In addition to identifying interactions that have previously been observed, we identified and quantified 56 PDZ-peptide interactions that have not previously been reported. Since, in several cases, more than one PDZ domain from a given protein recognized the same peptide, these interactions represent 39 new protein-protein interactions. It is important to emphasize that these are *bona fide* biophysical interactions; their physiological relevance, however, remains to be determined. Since two proteins must be co-expressed in order to interact in a physiological context, we can use information from large-scale gene expression studies as a first-pass filter for physiological relevance. Using high density oligonucleotide arrays, Hogenesch and coworkers have measured expression levels in 61 different mouse tissues for most of the protein-encoding genes in the mouse genome.[46] Of the 12 PDZ-containing proteins and 20 PDZ ligands used in our study, good quality expression data are available for 9 of the PDZ proteins and 11 of the PDZ ligands. This means that, of the 39 new protein-protein interactions

that we identified, we were able to evaluate 17 of them for co-expression using the Hogenesch dataset. For this analysis, we used a fairly conservative approach: genes were deemed to be expressed in a tissue if their transcript levels were at least three-fold higher than the median transcript level for that gene across all 61 tissues. By this criterion, some proteins were designated 'expressed' in only one tissue (Scn4a, for example, is found only in skeletal muscle), while other proteins were found to be expressed in several tissues (PSD-95, for example, is expressed in 11 different tissues of the central nervous system). Even with this rather conservative cutoff, we found evidence of co-expression for 12 of the 17 novel interactions (Table 3). It should be noted, however, that absence of co-expression by this criterion does not rule out the possibility that two proteins interact *in vivo*. Of the 13 previously reported interactions that could be evaluated using the Hogenesch expression data, two did not meet our stringent criterion for co-expression.

One particularly compelling new interaction that we observed is that between Stargazin and the multiple PDZ-containing protein CIPP. Stargazin has been shown to interact with α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, and this interaction is essential for the delivery of these receptors to the membrane of granule cells.[47] It has also been shown that the C-terminal tail of Stargazin is required for targeting AMPA receptors to synapses[47] and that mice with mutations in the gene encoding Stargazin lack functional AMPA receptors on cerebellar granule cells and exhibit ataxia and epilepsy.[48,49] *In situ* hybridization studies confirm the co-expression of Stargazin and CIPP in several brain tissues and CIPP is expressed at very high levels in the granule cell layer of the cerebellum.[48,50] That Stargazin and CIPP interact in our *in vitro* assay suggests that CIPP may play a role in targeting AMPA receptors to the synapse. It will be exciting to see what other hypotheses emerge from a more global analysis of PDZ-mediated interactions.

Sequence-Function Relationships

In addition to identifying specific interactions, investigations of this nature provide a broad perspective on molecular recognition within biological systems. In order to investigate the relationship between PDZ sequence and function, we subjected our quantitative interaction data, expressed as equilibrium association constants (K_A's), to correlation-based hierarchical clustering in both dimensions (Figure 6). From the perspective of the peptides, the algorithm brings together ligands that are recognized by similar sets of PDZ domains. Not surprisingly, this unsupervised analysis sorted the peptides according to their class (peptide names are colored by class in Figure 6). A notable exception was the Dlgap1/2/3 peptide, a class I peptide that clustered with the class II peptides. Based on their three C-terminal residues, we would expect the class II peptides derived from EphrinB1/2, Nrxn1/2, and AN2 to behave similarly and the class I peptide derived from Dlgap1/2/3 to cluster tightly with the Mapk12 peptide. Instead, Dlgap1/2/3 and AN2 are both bound tightly by the PDZ domain of Shank3 and are not recognized by any of the other PDZ domains. Dlgap1/2/3 and AN2 both feature a bulky residue in position −1. Although this residue is solvent-exposed in most PDZ-peptide complexes, a crystal structure of the Shank3 PDZ domain complexed with the C-terminal hexapeptide of GKAP shows tight contacts with the −1 residue.[51] It is not obvious how the Shank3 PDZ accommodates a tryptophan residue at this position (instead of an arginine) and it is unlikely that the Shank3-Dlgap1/2/3 interaction would have been predicted based on our current understanding of domain selectivity. This observation underscores the importance of focusing on physiologically relevant sequences and taking an unbiased approach to the discovery of protein-protein interactions.

From the perspective of the PDZ domains, the clustering algorithm that we used brings together domains that exhibit similar sequence *selectivity*. As with the peptides, a dendrogram (tree) is generated that provides a relative measure of the similarity between elements in the matrix. In

the PDZ dendrogram, the similarity of two PDZ domains decreases as the horizontal length of the path that connects them to their closest common branch point (node) increases. Importantly, since the biological role of PDZ domains is to recognize and bind their target proteins, this dendrogram reflects protein *function*. Completely independently, we can also generate a dendrogram for these PDZ domains based on their primary amino acid sequence. By using the full sequence of the domains, rather than just their active site residues, the resulting dendrogram offers a view of their evolutionary history. In this case, nodes in the tree represent the most recent common ancestor linking two PDZ domains and the horizontal distance between the node and the modern day domains provides an approximate measure of the time since these two domains diverged. Interestingly, the tree derived from protein sequence is remarkably similar to the tree derived from protein function (Figure 6). It has previously been shown that the binding selectivity of a PDZ domain can be substantially altered by a few, or even a single, point mutation.[52–54] Despite the ease with which protein function can be altered by small perturbations in protein sequence, no such mutations were, in fact, retained throughout the evolution of these 22 PDZ domains; their sequence similarity tracks closely with their functional similarity. This outcome suggests that it is difficult to substantially rewire a protein interaction network without incurring deleterious effects. Although single protein-protein interactions may be added or subtracted, it is much more difficult to conceive of changes that dramatically alter network connectivity and yet confer a selective advantage. Our data are consistent with a model in which protein interaction networks evolve through small incremental steps. It will be interesting to see if this model is supported by interaction data that encompass the entire family of mouse PDZ domains, as well as other interaction modules.

Summary and Conclusion

With the availability of whole genome sequencing information, it is now possible to identify every member of an entire family of protein interaction domains. This provides, for the first time, the opportunity to study, in a comprehensive and unbiased way, the recognition properties of functionally related protein modules. In addition to generating biological hypotheses, efforts of this nature provide an opportunity to study fundamental aspects of molecular selectivity in a biological context and thus can provide insight into molecular evolution, as well as a more global perspective on how cells insulate parallel pathways from each other in some circumstances, exploit cross-talk between pathways in other circumstances, or use some of the same proteins in distinct pathways to produce different physiological outcomes. To achieve these goals, it is crucial that we obtain high quality, quantitative data using physiologically relevant molecules. Here, we have outlined a strategy that combines the efficiency and economy of protein microarray technology with the high fidelity of an automatable, solution-phase assay to uncover quantitative protein interaction networks for mouse PDZ domains. We have shown that protein microarrays are able to identify relative weak protein-protein interactions ($K_D \leq 10$ μM) with a false negative rate of 14% and a false positive rate of 14%; that these interactions can be retested and quantified efficiently; and that the resulting dataset recapitulates 85% of known interactions while highlighting new interactions of potential biological significance. We have also shown that, even with a relatively small interaction matrix comprising 520 measurements, there is a tight link between protein sequence and function that supports a model in which protein interaction networks evolve through small, incremental steps. We are currently expanding our study of mouse PDZ domains to include every member of this family. We anticipate that the resulting data will teach us much about biological selectivity and provide further insight into the relationship between sequence and function.

Experimental Section

Cloning of PDZ domains

To abstract individual domains from their full-length proteins, we determined their boundaries by aligning their sequences with ClustalW, a general purpose multiple sequence alignment program (http://www.ebi.ac.uk/clustalw/). We then used available structural information to confirm that the boundaries were defined appropriately. We defined the N-terminus of the PDZ domains at 15 residues before the GLGF motif and the C-terminus at 30 residues after the predicted end of the domain. When the PDZ domain occurred at the C-terminus of a protein or was adjacent to another domain, no C-terminal tail was included. We then used Web Primer (http://genome-www2.stanford.edu/cgi-bin/SGD/web-primer) to design primers for the amplification of predicted PDZ domains from mouse cDNA. Whenever Web Primer could not find suitable primers due to high GC content or secondary structure formation, we extended the PDZ domain by a few amino acids either on the N- or C-terminus until the primers met the specified criteria. We included the sequence CACC in our 5′ primers for directional cloning using TOPO vectors (Invitrogen, Carlsbad, CA) and incorporated a TAG stop codon into our 3′ primers.

Coding sequences were amplified from mouse cDNA (BD Biosciences, Palo Alto, CA) using the polymerase chain reaction (PCR) with the following cycling parameters: 95 °C, 5 min; followed by 38 cycles of 95 °C, 30 sec; 54 °C, 30 sec; 72 °C, 1 min; followed by a final 10 min incubation at 72 °C. PCR products were separated by agarose gel electrophoresis, and bands of the appropriate size were excised and purified using an agarose gel purification kit (Qiagen, Valencia, CA). The resulting products were transferred into the vector pENTR/D-TOPO by topoisomerase I-mediated directional cloning (Invitrogen). Each clone was verified by DNA sequencing.

Production and purification of recombinant proteins

The coding region for each PDZ domain or domain cluster was transferred into the *E. coli* expression vector pET-32-DEST[18] via λ-recombinase-mediated directional subcloning and confirmed by restriction enzyme digestion. Expression vectors were transformed into BL21 (DE3)pLysS *E. coli* and cells from a single ampicillin and chloramphenicol-resistant colony were grown at 37 °C in 500 mL Luria-Bertani medium supplemented with 100 μg/mL ampicillin and 30 μg/mL chloramphenicol to an OD_{600} of approximately 0.7. Isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to a final concentration of 1 mM and the cultures were shaken at 25 °C for 15 h. Cells were recovered by centrifugation and resuspended in 20 mL lysis buffer (300 mM NaCl, 10 mM imidazole, 50 mM NaH_2PO_4, pH 8) containing 1 mM each of benzamidine and phenylmethylsulphonylfluoride (PMSF). Cells were lysed by freeze/thaw, followed by sonication (2 min). Samples were then centrifuged at 30,000 *g* for 30 min at 4 °C to remove insoluble material. His_6-tagged proteins were purified by immobilized metal affinity chromatography using Ni-NTA agarose beads (Qiagen). Beads were washed with 50 mL of lysis buffer, followed by 50 mL of wash buffer (300 mM NaCl, 20 mM imidazole, 50 mM NaH_2PO_4, 0.1% (v/v) Triton X-100, pH 8). Proteins were eluted from the beads with 1 mL of elution buffer (300 mM NaCl, 200 mM ethylenediaminetetraacetic acid (EDTA), 50 mM NaH_2P0_4, pH 8) and dialyzed against Buffer A (100 mM KCl, 10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4). Protein concentrations were determined based on their absorbance at 280 nm. For samples used in the protein microarray experiments, glycerol was added to each sample to a final concentration of 20% (v/v). Proteins were divided into aliquots and stored at −80 °C.

Peptide synthesis

Peptides were synthesized on an Apex 396 single-probe fast wash peptide synthesizer (Advanced ChemTech, Louisville, KY). Peptides were synthesized in *N*,*N*-dimethylformamide (DMF) on the solid phase at a 50 μmol scale using standard Fmoc chemistry. All amino acids were coupled twice at 5-fold molar excess. Amino acids were activated *in situ* with 0.95 equivalents of 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 2 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) and coupled for 1 h at room temperature. Resin was standard polystyrene Wang resin (0.8 mmol/g) charged with the appropriate C-terminal residue. All peptides were synthesized with an additional NNG sequence at their N-terminus to improve solubility. Following their synthesis but before deprotection and cleavage, peptides were labeled for 1 h with 2 equivalents of 5-(and-6)-carboxytetramethylrhodamine (Molecular Probes, Eugene, OR), activated with an equimolar amount of HBTU. Peptides were cleaved from the resin using Reagent K [82.5% TFA (v/v), 5% phenol (v/v), 5% water (v/v), 5% thioanisole (v/v), 2.5% 1,2-ethanedithiol (v/v)] and precipitated in cold diethylether. Peptides were purified by reverse phase HPLC using a C_{18} semipreparative column (Grace Vydac, Bodman Industries, Aston, PA). Fractions containing the correct product were identified by MALDI-TOF mass spectrometry using a Voyager DE Pro (Applied Biosystems, Foster City, CA). Solvent was removed by lyophilization and the purified peptides stored at −80 °C.

Manufacture and processing of protein microarrays

Purified recombinant PDZ domains were spotted at a concentration of 40 μM onto 112.5 mm × 74.5 mm × 1 mm aldehyde-presenting glass substrates (Erie Scientific Company, Portsmouth, NH) using a Biochip Arrayer (PerkinElmer, Boston, MA). 96 identical microarrays were printed in a 12 by 8 pattern on the glass plates, with a pitch of 9 mm. Each microarray consisted of a 9 by 9 pattern of spots, with a center-to-center spacing of 250 μm. Proteins were spotted in triplicate. Following a 1 h-incubation, the glass was attached to the bottom of a bottomless 96-well microtiter plate (Greiner Bio-one, Kremsmünster, Austria) using an intervening silicone gasket (Grace Bio-Labs, Bend, OR). Immediately before use, the plates were quenched with Buffer A containing 1% bovine serum albumin (BSA) (w/v) for 1 h at room temperature, followed by incubation in Buffer A containing 1% BSA (w/v) and 50 mM glycine. The arrays were rinsed briefly in Buffer A containing 0.1% Tween 20 (v/v) and probed with either 1 or 5 μM of 5(6)-TAMRA-labeled peptides, dissolved in Buffer B (100 mM KCl, 1% BSA (w/v), 0.1% Tween 20 (v/v), 1 mM dithiothreitol, 10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4). Following a 1-h incubation at room temperature, the peptide solution was removed and the arrays washed with 300 μL of Buffer A containing 0.1% Tween 20 (v/v). The arrays were rinsed twice with 300 μL ddH_2O and spun upside down in a centrifuge for 60 sec to remove residual water.

Scanning and analysis of protein microarrays

PDZ microarrays were scanned at 10 μm resolution using a Tecan LS400 microarray scanner (Tecan, Männedorf, Switzerland). Cyanine-5 fluorescence was imaged using a 633 nm laser and 5(6)-TAMRA fluorescence was imaged using a 543 nm laser. Images were analyzed using Array-Pro Analyzer 4.5 (Tecan). Microarray spots were identified based on the cyanine-5 image and the mean 5(6)-TAMRA fluorescence of each protein was calculated from the three replicate spots. Fold over background (FOB) values were determined by dividing the mean 5 (6)-TAMRA fluorescence for each protein by the mean 5(6)-TAMRA fluorescence of nine replicate control spots (thioredoxin). For those arrays probed with 1 μM of peptide, FOB values for each protein-peptide interaction were averaged over three independent trials.

Fluorescence polarization

PDZ domains were introduced into separate wells in row A of black 384-well microtiter plates (Corning, Corning, NY) at a concentration of 25 μM (80 μL per well). The remaining wells of the plate were filled with 40 μL of Buffer A and two-fold serial dilutions of each domain were prepared down each column, resulting in 40 μL per well. Fluorescent peptides were dissolved at a concentration of 100 nM in Buffer C (100 mM KCl, 0.1% BSA (w/v), 5 mM dithiothreitol, 10 mM NaH_2PO_4/Na_2HPO_4, pH 7.4) and introduced into a separate 384-well plate. Each well in a given column contained the same peptide solution. 10 μL of peptide solution was then transferred from every well of the peptide plate to the corresponding well in the PDZ assay plate using a Biomek FX 96-channel pipetting robot (Beckman Coulter, Fullerton, CA). Plates were incubated at room temperature for 1 h. 5(6)-TAMRA fluorescence was detected using an Analyst AD fluorescence plate reader (Molecular Devices, Sunnydale, CA), with excitation at 525 nm and emission at 590 nm. Fluorescence polarization, in millipolarization units (mP), was defined as$10^3 \cdot (I^{=} - I^{\perp}) (I^{=} + I^{\perp})$, where $I^{=}$ and $I^{\perp}$ are the fluorescence intensities parallel and perpendicular to the plane of incident light, respectively. To determine the equilibrium dissociation constant (K_D) for each PDZ-peptide interaction, the fluorescence polarization data were fit to eq 1 in an automated fashion using Origin (Origin Lab Corporation, Northampton, MA).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the Bauer Center for Genomics Research at Harvard University for support with instrumentation and automation. This work was supported by an award from the Smith Family Foundation, an award from the Arnold and Mabel Beckman Foundation, and a grant from the National Institutes of Health (1 RO1 GM072872-01). M.A.S. was supported in part by the National Institutes of Health Molecular, Cellular, and Chemical Biology Training Grant (5 T32 GM07598-25). M.S. is the recipient of an Alfred and Isabel Bader fellowship and a Jacques-Émile Dubois fellowship.

References

1. Pawson T, Nash P. Science 2003;300:445–452. [PubMed: 12702867]
2. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. Nature 2000;403:623–627. [PubMed: 10688190]
3. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. Proc Natl Acad Sci U S A 2001;98:4569–4574. [PubMed: 11283351]
4. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. Science 2004;303:540–543. [PubMed: 14704431]
5. Newman JR, Wolf E, Kim PS. Proc Natl Acad Sci U S A 2000;97:13203–13208. [PubMed: 11087867]
6. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G. Nature 2002;415:141–147. [PubMed: 11805826]
7. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M,

Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M. Nature 2002;415:180–183. [PubMed: 11805837]

8. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Nature. 2006
9. Aloy P, Russell RB. FEBS Lett 2002;530:253–254. [PubMed: 12387902]
10. Bader JS, Chaudhuri A, Rothberg JM, Chant J. Nat Biotechnol 2004;22:78–85. [PubMed: 14704708]
11. Phizicky E, Bastiaens PI, Zhu H, Snyder M, Fields S. Nature 2003;422:208–215. [PubMed: 12634794]
12. Deane CM, Salwinski L, Xenarios I, Eisenberg D. Mol Cell Proteomics 2002;1:349–356. [PubMed: 12118076]
13. Cesareni G, Ceol A, Gavrila C, Palazzi LM, Persico M, Schneider MV. FEBS Lett 2005;579:1828–1833. [PubMed: 15763559]
14. Deeds EJ, Ashenberg O, Shakhnovich EI. Proc Natl Acad Sci U S A 2006;103:311–316. [PubMed: 16384916]
15. Obenauer JC, Yaffe MB. Methods Mol Biol 2004;261:445–468. [PubMed: 15064475]
16. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Nat Biotechnol 2002;20:370–375. [PubMed: 11923843]
17. Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R, Westerhoff HV. Oncogene 2005;24:5533–5542. [PubMed: 16007170]
18. Jones RB, Gordus A, Krall JA, MacBeath G. Nature 2006;439:168–174. [PubMed: 16273093]
19. MacBeath G, Schreiber SL. Science 2000;289:1760–1763. [PubMed: 10976071]
20. Cho KO, Hunt CA, Kennedy MB. Neuron 1992;9:929–942. [PubMed: 1419001]
21. Woods DF, Bryant PJ. Mech Dev 1993;44:85–89. [PubMed: 8155583]
22. Kim E, Niethammer M, Rothschild A, Jan YN, Sheng M. Nature 1995;378:85–88. [PubMed: 7477295]
23. Kornau HC, Schenker LT, Kennedy MB, Seeburg PH. Science 1995;269:1737–1740. [PubMed: 7569905]
24. Sato T, Irie S, Kitada S, Reed JC. Science 1995;268:411–415. [PubMed: 7536343]
25. Garner CC, Nash J, Huganir RL. Trends Cell Biol 2000;10:274–280. [PubMed: 10856930]
26. Bilder D. Trends Genet 2001;17:511–519. [PubMed: 11525834]
27. Sheng M, Sala C. Annu Rev Neurosci 2001;24:1–29. [PubMed: 11283303]
28. Nourry C, Grant SG, Borg JP. Sci STKE 2003;2003:RE7. [PubMed: 12709532]
29. Hildebrand JD, Soriano P. Cell 1999;99:485–497. [PubMed: 10589677]
30. Zhadanov AB, Provance DW Jr, Speer CA, Coffin JD, Goss D, Blixt JA, Reichert CM, Mercer JA. Curr Biol 1999;9:880–888. [PubMed: 10469590]
31. Bladt F, Tafuri A, Gelkop S, Langille L, Pawson T. Proc Natl Acad Sci U S A 2002;99:6816–6821. [PubMed: 11983858]
32. Laverty HG, Wilson JB. Genomics 1998;53:29–41. [PubMed: 9787075]
33. Caruana G, Bernstein A. Mol Cell Biol 2001;21:1475–1483. [PubMed: 11238884]
34. Boeda B, El-Amraoui A, Bahloul A, Goodyear R, Daviet L, Blanchard S, Perfettini I, Fath KR, Shorte S, Reiners J, Houdusse A, Legrain P, Wolfrum U, Richardson G, Petit C. Embo J 2002;21:6689–6699. [PubMed: 12485990]
35. Verpy E, Leibovici M, Zwaenepoel I, Liu XZ, Gal A, Salem N, Mansour A, Blanchard S, Kobayashi I, Keats BJ, Slim R, Petit C. Nat Genet 2000;26:51–55. [PubMed: 10973247]
36. Boerkoel CF, Takashima H, Stankiewicz P, Garcia CA, Leber SM, Rhee-Morris L, Lupski JR. Am J Hum Genet 2001;68:325–333. [PubMed: 11133365]

37. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinsci F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E. Nucleic Acids Res 2005;33:D447–453. [PubMed: 15608235]
38. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. Nucleic Acids Res 2006;34:D257–260. [PubMed: 16381859]
39. Songyang Z, Fanning AS, Fu C, Xu J, Marfatia SM, Chishti AH, Crompton A, Chan AC, Anderson JM, Cantley LC. Science 1997;275:73–77. [PubMed: 8974395]
40. Grootjans JJ, Zimmermann P, Reekmans G, Smets A, Degeest G, Durr J, David G. Proc Natl Acad Sci U S A 1997;94:13683–13688. [PubMed: 9391086]
41. Grootjans JJ, Reekmans G, Ceulemans H, David G. J Biol Chem 2000;275:19933–19941. [PubMed: 10770943]
42. Long JF, Tochio H, Wang P, Fan JS, Sala C, Niethammer M, Sheng M, Zhang M. J Mol Biol 2003;327:203–214. [PubMed: 12614619]
43. Zhang Q, Fan JS, Zhang M. J Biol Chem 2001;276:43216–43220. [PubMed: 11553623]
44. Roehrl MH, Wang JY, Wagner G. Biochemistry 2004;43:16056–16066. [PubMed: 15610000]
45. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Nature 2002;417:399–403. [PubMed: 12000970]
46. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. Proc Natl Acad Sci U S A 2004;101:6062–6067. [PubMed: 15075390]
47. Chen L, Chetkovich DM, Petralia RS, Sweeney NT, Kawasaki Y, Wenthold RJ, Bredt DS, Nicoll RA. Nature 2000;408:936–943. [PubMed: 11140673]
48. Letts VA, Felix R, Biddlecome GH, Arikkath J, Mahaffey CL, Valenzuela A, Bartlett FS 2nd, Mori Y, Campbell KP, Frankel WN. Nat Genet 1998;19:340–347. [PubMed: 9697694]
49. Hashimoto K, Fukaya M, Qiao X, Sakimura K, Watanabe M, Kano M. J Neurosci 1999;19:6027–6036. [PubMed: 10407040]
50. Kurschner C, Mermelstein PG, Holden WT, Surmeier DJ. Mol Cell Neurosci 1998;11:161–172. [PubMed: 9647694]
51. Im YJ, Lee JH, Park SH, Park SJ, Rho SH, Kang GB, Kim E, Eom SH. J Biol Chem 2003;278:48099–48104. [PubMed: 12954649]
52. Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C. Nat Struct Biol 2002;9:621–627. [PubMed: 12080331]
53. Schneider S, Buchert M, Georgiev O, Catimel B, Halford M, Stacker SA, Baechi T, Moelling K, Hovens CM. Nat Biotechnol 1999;17:170–175. [PubMed: 10052354]
54. Gee SH, Quenneville S, Lombardo CR, Chabot J. Biochemistry 2000;39:14638–14646. [PubMed: 11087420]

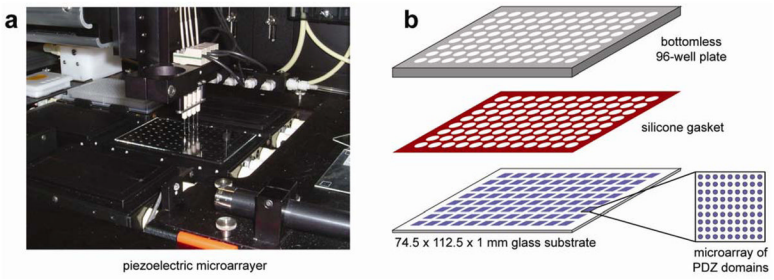

Figure 1.

Microarrays in microtiter plates. (a) Piezoelectric microarrayer spotting PDZ domains on aldehyde-presenting glass. (b) Attachment of microarrays to bottomless microtiter plate using an intervening silicone gasket.

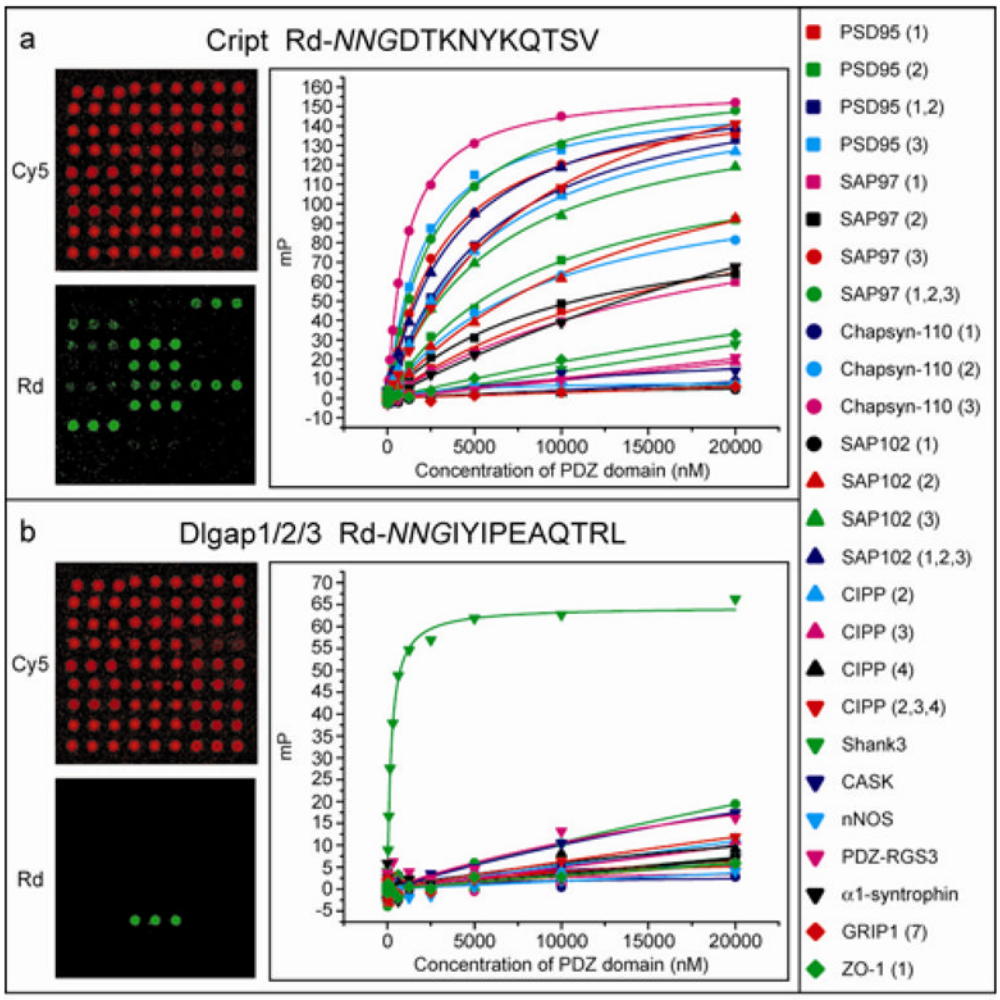

Figure 2.

Detection and quantification of PDZ-peptide interactions. (a) A peptide derived from Cript interacts with many PDZ domains. The two panels on the left show a PDZ microarray, probed with 1 μM Cript peptide. The Cy5 image (red) shows the placement of the spots while the 5 (6)-TAMRA image (Rd; green) shows binding of the peptide to the immobilized PDZ domains. The PDZ's were spotted in triplicate from left to right, top to bottom, in the order shown on the far right. The last three spots on the microarray are thioredoxin. The graph to the right of the microarrays shows fluorescence polarization curves obtained by incubating the Cript peptide with varying concentrations of each PDZ domain. (b) A peptide derived from Dlgap1/2/3 is selective for only one PDZ domain.

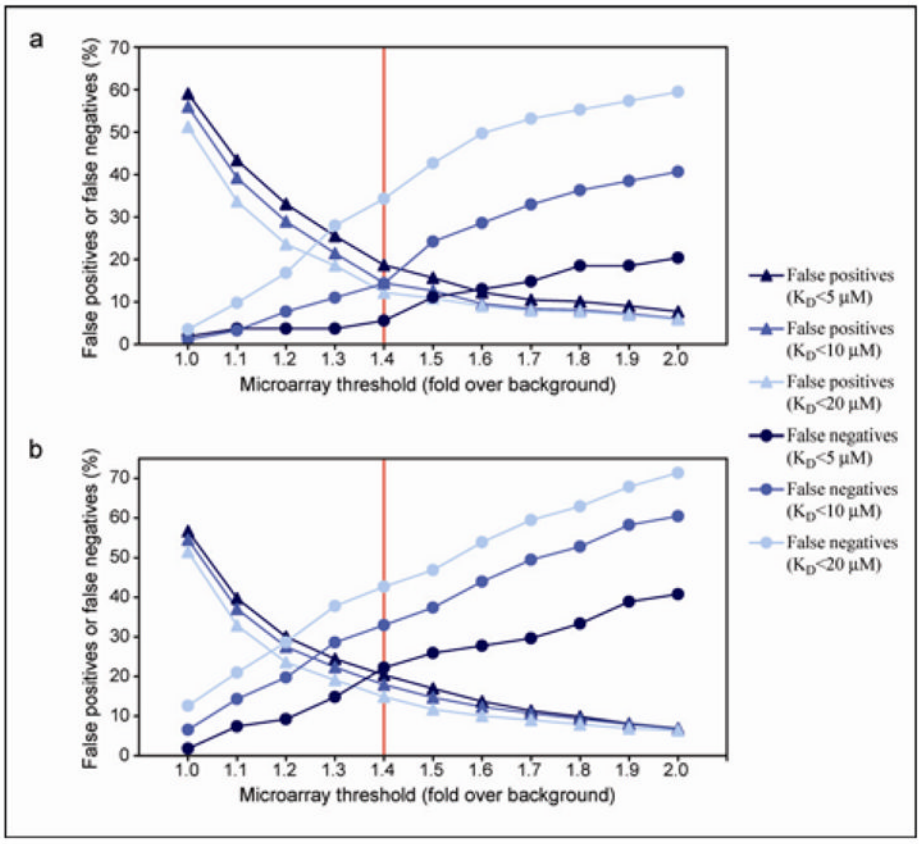

Figure 3.

False positive/false negative rates of protein microarrays. (a) False positive and false negative rates were determined at microarray thresholds ranging from 1.0 to 2.0 fold over background (FOB). The red line indicates a threshold of 1.4. Results relative to three different gold standard datasets are shown: interactions with $K_D \leq 5$ μM; interactions with $K_D \leq 10$ μM; and interactions with $K_D \leq 20$ μM. (a) Results obtained by probing the microarrays with 1 μM peptide. (b) Results obtained by probing the microarrays with 5 μM peptide.

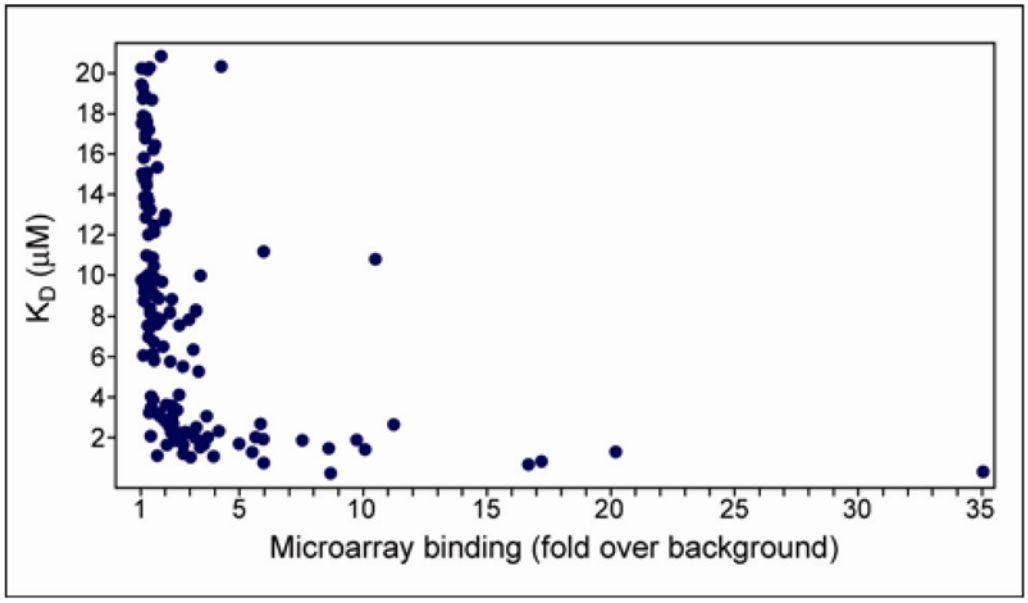

Figure 4.
Correlation between microarray spot intensity (fold over background) and K_D (as determined by fluorescence polarization).

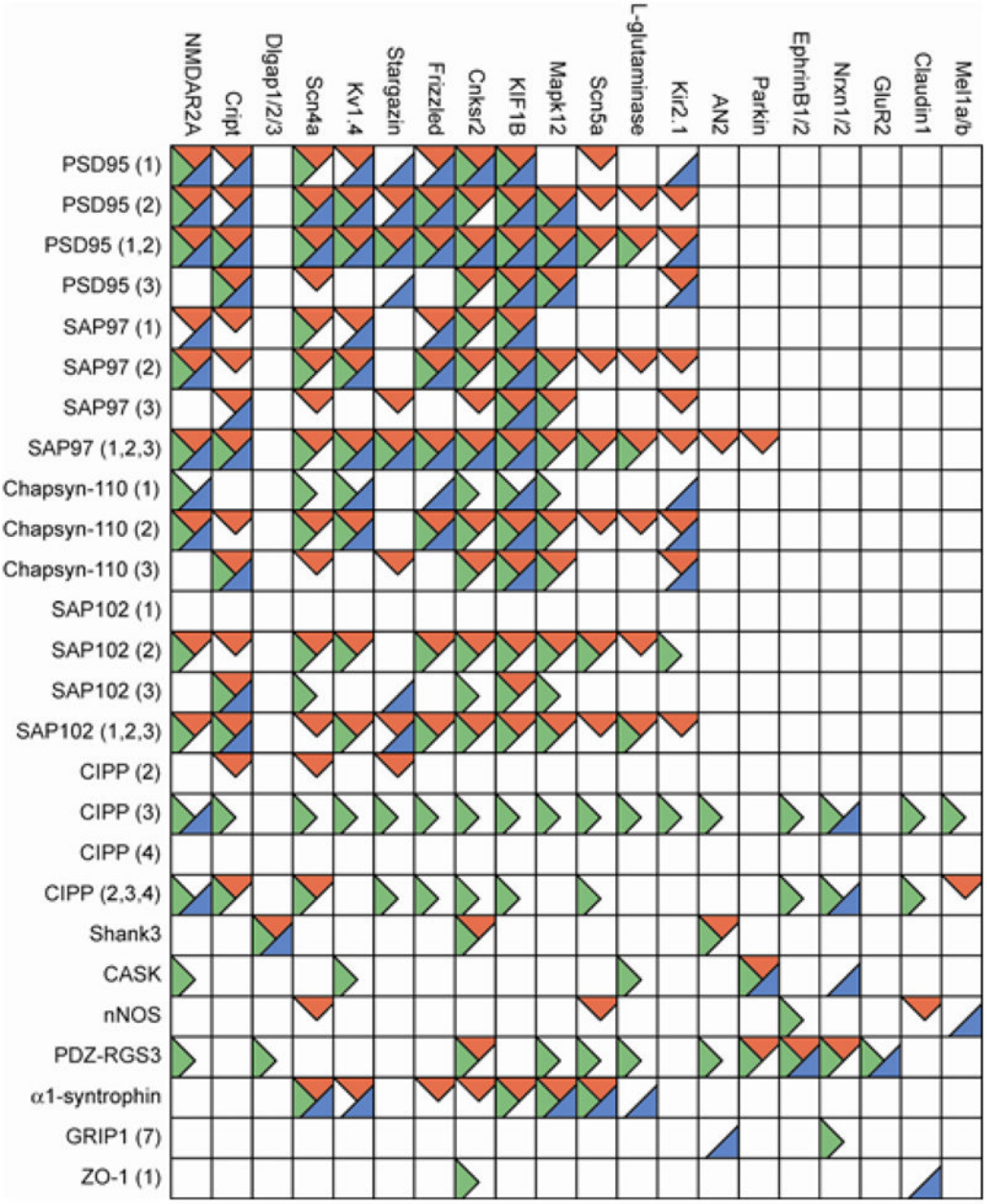

Figure 5.

A comparison of observed and reported interactions. Rows are PDZ domains and columns are PDZ interaction partners. Blue triangles are previously reported interactions; red triangles are interactions observed by fluorescence polarization with $K_D \leq 20$ μM; and green triangles are interactions observed on the microarrays with FOB > 1.4.

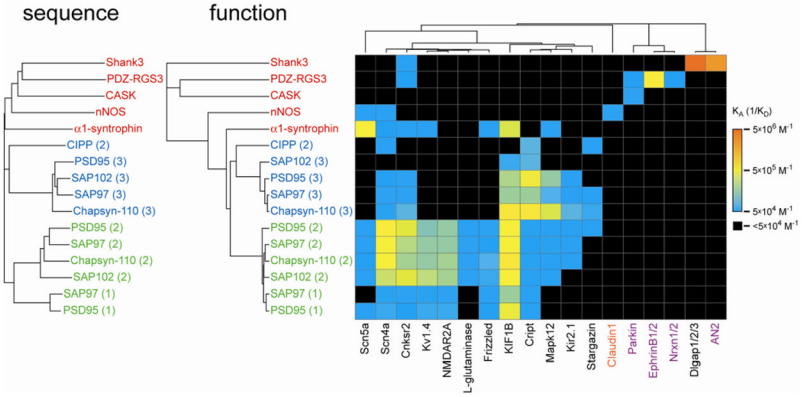

Figure 6.

Relationship between PDZ domain sequence and function. (Right) Correlation-based hierarchical clustering of PDZ-peptide affinity constants. All interactions with $K_D \leq 20$ μM ($K_A \geq 5 \times 10^4$ M^{-1}) were used. Peptide names are colored according to class (class I peptides are black, class II peptides are purple, class III peptides are orange) and PDZ names are colored to highlight major clusters. (Left) Dendrogram resulting from a multiple sequence alignment of the PDZ domains.

Table 1

PDZ Domains

PDZ Domain[a]	Ensembl Gene ID
PSD95 (1)	ENSMUSG00000020886
PSD95 (2)	ENSMUSG00000020886
PSD95 (1,2)	ENSMUSG00000020886
PSD95 (3)	ENSMUSG00000020886
SAP97 (1)	ENSMUSG00000022770
SAP97 (2)	ENSMUSG00000022770
SAP97 (3)	ENSMUSG00000022770
SAP97 (1,2,3)	ENSMUSG00000022770
Chapsyn-110 (1)	ENSMUSG00000052572
Chapsyn-110 (2)	ENSMUSG00000052572
Chapsyn-110 (3)	ENSMUSG00000052572
SAP102 (1)	ENSMUSG00000000881
SAP102 (2)	ENSMUSG00000000881
SAP102 (3)	ENSMUSG00000000881
SAP102 (1,2,3)	ENSMUSG00000000881
CIPP (2)	ENSMUSG00000061859
CIPP (3)	ENSMUSG00000061859
CIPP (4)	ENSMUSG00000061859
CIPP (2,3,4)	ENSMUSG00000061859
Shank3	ENSMUSG00000022623
CASK	ENSMUSG00000031012
nNOS	ENSMUSG00000029361
PDZ-RGS3	ENSMUSG00000059810
α1-syntrophin	ENSMUSG00000027488
GRIP1 (7)	ENSMUSG00000034813
ZO-1 (1)	ENSMUSG00000030516

[a]The name of the protein containing each PDZ domain is provided. For proteins that contain more than one PDZ domain, the domain number is provided in parentheses.

Table 2

PDZ Binding Partners

Protein	Peptide Sequence
NMDAR2A	NNGKKMPSIESDV
Cript	NNGDTKNYKQTSV
Dlgap1/2/3	NNGIYIPEAQTRL
Scn4a	NNGVRPGVKESLV
Kv1.4	NNGSNAKAVETDV
Stargazin	NNGNTANRRTTPV
Frizzled	NNGTNSKQGETTV
Cnksr2	NNGHTHSYIETHV
KIF1B	NNGNLKAGRETTV
Mapk12	NNGGARVPKETAL
Scn5a	NNGSPDRDRESIV
L-glutaminase	NNGLSKENLESMV
Kir2.1	NNGPRPLRRESEI
AN2	NNGPALRNGQYWV
Parkin	NNGACMGDHWFDV
EphrinB1/2	NNGQSPANIYYKV
Nrxn1/2	NNGKKNKDKEYYV
GluR2	NNGNVYGIESVKI
Claudin1	NNGPTPSSGKDYV
Mel1a/b	NNGNNNLIKVDSV

Table 3

Co-expression of Interacting Proteins

PDZ Domain	Interacting Partner	Tissues in which the genes for both proteins are expressed[a]
PSD95	Scn5a	Preoptic
Chapsyn-110	Cnksr2	Amygdala[b], frontal cortex[b], preoptic, cerebellum[b], cerebral cortex[b], dorsal root ganglia, dorsal striatum[b], hippocampus[b], hypothalamus, olfactory bulb
Chapsyn-110	Mapk12	Cerebellum, dorsal root ganglia, main olfactory epithelium
Chapsyn-110	Scn5a	Preoptic, main olfactory epithelium
SAP102	Kv1.4	Olfactory bulb
SAP102	Cnksr2	Amygdala, frontal cortex, preoptic, cerebral cortex, dorsal striatum, hippocampus, olfactory bulb
SAP102	KIF1B	Amygdala, frontal cortex, preoptic, cerebral cortex, dorsal striatum, hippocampus, olfactory bulb
Scn5a	Scn5a	Preoptic
CIPP	Stargazin	Cerebellum[b], upper spinal cord, lower spinal cord
Shank3	Cnksr2	Amygdala, frontal cortex, preoptic, cerebellum[b], cerebral cortex[b], dorsal striatum[b], hippocampus[b], hypothalamus, olfactory bulb
PDZ-RGS3	Cnksr2	Cerebellum, dorsal root ganglia
PDZ-RGS3	Nrxn1/2	Trigeminal, cerebellum, dorsal root ganglia, main olfactory epithelium
Chapsyn-110	Scn4a	None
SAP102	Scn4a	None
SAP102	Mapk12	None
CIPP	Scn4a	None
Shank3	AN2	None

[a] Genes were considered ‘expressed’ if their transcript levels were at least three-fold greater than the median transcript levels for that gene across all mouse tissues.

[b] In these tissues, the genes for both proteins were expressed at levels that were more than ten-fold over the median.