



Published in final edited form as:

ACS Synth Biol. 2019 May 17; 8(5): 918–928. doi:10.1021/acssynbio.8b00499.

A Multireporter Bacterial 2-Hybrid Assay for the High-Throughput and Dynamic Assay of PDZ Domain–Peptide Interactions

David M. Ichikawa^{†,‡,§}, Carles Corbi-Verge^{†,‡,§}, Michael J. Shen[†], Jamie Snider[‡], Victoria Wong[‡], Igor Stagljar^{‡,§,||}, Philip M. Kim^{*,‡,§,⊥}, Marcus B. Noyes^{*,†}

[†]Department of Biochemistry Molecular Pharmacology and Institute for Systems Genetics, NYU Langone Health, New York, New York 10016, United States

[‡]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada

[§]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada

^{||}Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 3E1, Canada

[⊥]Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3E1, Canada

Abstract

The accurate determination of protein–protein interactions has been an important focus of molecular biology toward which much progress has been made due to the continuous development of existing and new technologies. However, current methods can have limitations, including scale and restriction to high affinity interactions, limiting our understanding of a large subset of these interactions. Here, we describe a modified bacterial-hybrid assay that employs combined selectable and scalable reporters that enable the sensitive screening of large peptide libraries followed by the sorting of positive interactions by the level of reporter output. We have applied this tool to characterize a set of human and *E. coli* PDZ domains. Our results are consistent with prior characterization of these proteins, and the improved sensitivity increases our ability to predict known and novel *in vivo* binding partners. This approach allows for the recovery of a wide range of affinities with a high throughput method that does not sacrifice the scale of the screen.

Graphical Abstract

* **Corresponding Authors** pm.kim@utoronto.ca., marcus.noyes@nyumc.org.

Author Contributions

M.B.N. and P.M.K. conceived of the project. D.M.I. and M.J.S. carried out the bacterial experiments, while C.C.V. executed all computational analysis. J.S. and V.W. carried out all MaMTH construct generation and assays. I.S. oversaw all MaMTH assays and provided critical analysis of results. D.M.I., M.B.N., and P.M.K. wrote the manuscript.

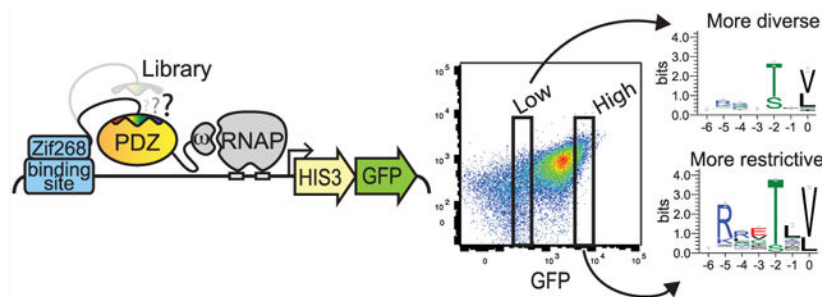
[‡]D.M.I. and C.C.V. contributed equally.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acssyn-bio.8b00499](https://doi.org/10.1021/acssyn-bio.8b00499).

Figures S1–S7 that detail all of the PDZ selection results, peptide analysis, and the base expression cassettes used in our vectors; Table S1 of the PDZ-peptide affinities used to calibrate our systems; Supporting Methods (PDF)

The authors declare no competing financial interest.



Keywords

bacteria hybrid assay; PDZ domain; protein interaction; synthetic biology

Protein interactions are at the foundation of cellular processes, yet our understanding of how variation across the proteome may influence these interactions is cursory at best. In fact, it is estimated that there are over 500000 protein–protein interactions (PPIs) predicted to occur in a human cell,¹ yet we only understand the mechanism behind a very small fraction of these interactions. Understanding these mechanisms may help us to predict the functional consequence of protein variations in these PPIs, a critical point to comprehend as we now know from studies such as the 1000 Genomes Project that variation is prevalent from one human genome to the next.^{2,3} These variations might alter the coding sequence of interacting partners⁴ as well as influence expression levels through modification of regulatory sequences. While we already know that common PPIs play important roles in cellular programs,⁵ and dysfunctional interactions are at the heart of many diseases,⁶ our limited understanding of the mechanisms behind most PPIs severely restricts our ability to understand how protein variants might influence these interactions.

Understanding the mechanism of any protein interaction can be difficult *a priori*. However, some of the most common interactions are dictated by conserved domain families that have evolved across Archaea, Bacteria, and Eukaryota. Some of these protein domains mediate interactions with short conserved peptide sequences commonly referred to as short-linear motifs (SLiMs), typically comprised of <12 amino acids in disordered regions.⁷ Such interactions are of somewhat simpler nature than those of two globular protein domains and are particularly common in signaling pathways. SLiM interaction domains (SLiDs) are common in metazoan proteomes with over 1000 human proteins containing at least one of the top 5 most common (PDZ, SH3, SH2, WD40, and TPR). While some SLiDs require post-translational modification of their target peptides,⁸ many do not. For example, the PDZ domain (named after the 3 proteins for which it was first characterized PSD-95/Dlg/ZO-1) primarily specifies the unaltered C-terminal residues of its binding partner, though phosphorylation can block or modify this interaction.⁹ Unlike eukaryotic proteomes, SLiDs appear far less common in bacteria with the minor exception of the PDZ domain that is found in many prokaryotic species.¹⁰ However, prokaryotes that do express proteins with PDZ domains typically have a small number of these proteins or none at all. Nevertheless, the sheer number of prokaryotic species leads to an extremely large total number of prokaryotic PDZs.¹¹ In fact, there are more PDZ domains found in 96 common microbes of

the human microbiome¹² (368) than human PDZ domains¹³ (266). The prevalence of the PDZ domain in humans and its simplified interaction with the C-terminus has led to several focused studies,¹⁴⁻¹⁷ but many of these works have been biased by affinity, limited in scale, and largely overlook the prokaryotic PDZs.

Many high-throughput screens have been employed to define PPIs;¹⁸⁻²⁴ however, genetic screens and hybrid assays, in particular driven in a high-throughput manner, have had relatively low sensitivity in detecting motif-mediated interactions. Likewise, affinity-purification mass spectrometry (AP/MS) is focused by design on high-affinity interactions and detects few SLiD interactions. Other methods are geared directly toward domain-motif interactions, but suffer from other shortcomings. For instance, peptide arrays²³ are limited by scale; as it is still relatively expensive to synthesize the peptides on the array, only a limited number of peptides can be measured. Similarly, the hold-up assay offers quantitative affinity data, but is limited by the number of peptides measured.²⁵ Peptide phage display is another commonly employed method naturally suited to characterize domain-motif interactions.^{14,15,26} However, as a method based on enrichment over multiple rounds, the results can be biased by affinity and specific amino acid properties,²⁷ limiting their predictive power of lower affinity interactions that might be common *in vivo*. In addition, as purifying each individual protein to screen can be laborious, any method that requires protein expression and purification would be quite challenging to employ for an exhaustive screen of PPIs and their variants. Moving forward, we will need a simple, fast, and accurate method to address protein variations in either member of an interacting pair and provide a predictive model of protein function.

Here we describe a simple, sensitive, multireporter bacterial 2-hybrid system (MR-B2H) capable of capturing a dynamic range of protein interactions and enabling the accurate prediction of known and novel *in vivo* protein partners. Using libraries that represent both naturally occurring as well as purely random peptides, we characterize PDZ-peptide interactions that represent all the C-termini a PDZ domain will encounter in a cell as well as all mutant variants of those termini. We apply the MR-B2H system to characterize several PDZ domains from human and *E. coli* proteins that have been investigated by other common methods, allowing us to benchmark our results. We find the MR-B2H results demonstrate an improved ability to predict known *in vivo* interacting partners. The MR-B2H is a robust, rapid, and accurate method that could be adopted by almost any lab. Some of the main benefits of the MR-B2H detailed here are (1) facile workflow circumventing the need for protein purification and using only basic molecular biology reagents, (2) parallel and rapid characterization of protein domains to facilitate screening >15 domains in 2–3 days, (3) accurate prediction of protein interactions, and (4) sensitive fluorescent output that captures a broad range of PPI affinities and allows an approximation of relative activity. While we have applied this technique toward the PDZ domain, it could be easily modified to investigate other SLiDs as well as more complex protein interactions.

RESULTS AND DISCUSSION

Adapting the Multireporter Bacterial 1-Hybrid (B1H) to a B2H.

The omega (ω)-based B1H is a well-established, robust, and rapid method for characterizing protein–DNA interactions.^{28–30} The strength of this system is that assays can be done in a matter of days and multiple domains can be investigated in parallel. In addition, the system employs the ω -subunit (RpoZ) of bacterial RNA-polymerase (RNAP) as the activation domain and assays are conducted in an *E. coli* strain lacking the endogenous omega, as this RNAP subunit is nonessential. In this way, recruitment of the polymerase takes place in the absence of competition from the endogenous omega,³⁰ resulting in a highly sensitive assay of the protein–DNA interaction. Prior works have employed PPIs in bacterial hybrid assays but these projects were not modular nor to the scale of the omega-based system and they largely focused on understanding protein–DNA interactions or engineering PPIs.^{31–33} In order to adapt our system to focus on the characterization of PPIs and take advantage of the sensitivity of the omega recruitment, we made the following modifications. First, the PDZ domain of interest is expressed as a fusion to the omega subunit from the low copy reporter plasmid (Figure 1A). Second, a 7-amino acid peptide is fused to the C-terminus of the Zif268 DNA-binding domain *via* a 10-residue linker. The consensus Zif268 binding-site is placed 21 base pairs upstream of a weak minimal promoter that drives the expression of HIS3 and GFP of the reporter plasmid. In this way, activation of the MR-B2H system and expression of the selectable markers HIS3 and GFP is dependent on the PDZ interacting with the peptide: the DNA-binding domain interacts with its target and omega will interact with the RNAP, but the two components will only be brought together on the reporter plasmid through a positive PDZ-peptide interaction. Further, the strength of the PPI required to survive selective pressure can be titrated by different concentrations of 3-amino triazole (3AT)—a competitive inhibitor of the HIS3 gene product.

To test the functionality and demonstrate the sensitivity and dynamic range of the MR-B2H, we tested nine previously characterized peptides paired with the PDZ domain of Erbin (ERBIN) for their combined ability to activate our GFP reporter. These peptides were chosen from prior work as they had demonstrated a wide-range of IC₅₀ values (3300 fold)³⁴ when bound to the ERBIN PDZ domain. Cells that expressed the omega-ERBIN PDZ domain and the Zif268-peptide fusion were challenged for their ability to activate the GFP reporter. Known lower-affinity interactions drove less GFP expression than high-affinity interactions (Figure 1B), but were still distinguishable from negative controls. Importantly, documented ERBIN PDZ-peptide interactions found in the mentha database³⁵ clustered in the lowest tier of GFP expression (Figure 1C). This data suggested that the MR-B2H system captures a wide range of affinities, and that many relevant interactions would be found in the lower GFP range.

Characterization of PDZ Specificity.

Having demonstrated the ability to capture a wide range of affinities, we next tested our ability to select positive PDZ-peptide interactions from large libraries. We constructed a purely *random* 7 amino acid library as well as a library comprised of ~50,000 *naturally* occurring 7 amino acid C-termini found in humans (templated from prior work).¹⁵

Combined, these libraries allow us to investigate the peptides a PDZ would be exposed to in nature in addition to all possible variations of these peptides. While the build of the random library approached 10^{10} variants, the 10^9 transformation efficiency of *E. coli* limits the practicality of sampling fractions of this library greater than 10^9 in a single experiment. Therefore, all selections reported here are based on a sampling of roughly 10^9 members of our random library. The natural library of only 50,000 members was 1000-fold oversampled with a typical sampling of 5×10^7 variants per selection.

To demonstrate the method's ability to select positive interactions from our libraries and generate results rapidly, we carried out MR-B2H selections as previously described for B1H selections^{30,36} relying only on HIS3 activation. Selections were plated on a range of 3-amino triazole (3AT) concentrations, and individual colonies were screened for surviving library members by colony PCR or pooled and barcoded for Illumina sequencing (Figure 1D). As might be expected, the pools of library members recovered from lower stringency selections were more diverse than those from higher stringency selections. We initially performed selections against 4 PDZ domains from eukaryotic proteins, and 6 PDZ domains from 4 prokaryotic proteins. Three out of four eukaryotic PDZ domains yielded specificities matching known binding preferences (with Protein scribble homologue (SCRIB) PDZ3 having too few colonies to obtain reliable information), and at least one PDZ worked for 3 out of 4 prokaryotic proteins (with the Tail-specific protease (Prc) PDZ domain failing selection) (Figure 1E, and Figure S1). Importantly, the motif from the prokaryotic Serine endoprotease DegS (DegS) PDZ alone matches closely to the motifs³⁷ from the full-length protein (expressed as a catalytically active wild-type, or catalytically impaired mutant³⁸) indicating that a domain-based screening approach can be informative of full-length protein binding as well. This entire workflow can be accomplished in as little as 3 days (Figure 1F), highlighting this method as a rapid and facile approach for characterizing SLiDs.

The B2H System Captures a Wide Range of Affinities.

To expand on the multistringency selections, we took a complementary approach to leverage GFP expression as a relative measure of interaction strength. Rather than relying on 3AT stringency to select strength of interaction in a stepwise manner, we used GFP fluorescence as a continuous measure of interaction strength. However, to clear nonfunctional members from the library we first needed to perform a weak 3AT selection on plates (typically 1 mM 3AT). We then pooled surviving members and expanded in liquid media under the same weak selective pressure until GFP fluorescence approached its maximum. The GFP populations were sorted, representing low, medium, and high GFP (Figure 2A). A negative GFP population was also recovered to provide a baseline control. Consistent with the results from the plate selections, we found a more diverse set of peptides in lower GFP populations, while a more restricted set of peptides were required for maximal GFP activation (Figure 2B). In fact, we find that, in general, motifs compiled from lower GFP bins more closely resemble the diversity of interacting partners that are found *in vivo*. For example, the low GFP results for the ERBIN PDZ domain more closely resemble results from peptide arrays³⁹ as well as confirmed *in vivo* binders curated from mentha³⁵ (Figure 2B bottom). On the other hand, motifs from the high-GFP-expressing populations resemble the high-affinity consensus targets produced by phage display. Further support of these trends relating

affinity and GFP fluorescence intensity comes from the distribution of PDZ-peptide pairs with documented affinities across GFP fluorescence intensities obtained for the same PDZ-peptide pairs in our assay. Consistent with results above, we find an enrichment of lower affinity interactions ($>50 \mu\text{M}$) in low GFP pools, and higher affinity interactions ($<10 \mu\text{M}$) in the high GFP populations (Figure 2C, Table S1). Utilizing the GFP component of the MR-B2H that accounts for GFP fluorescence intensity enhances the data by allowing a better repartition of pairwise interaction hits in relation to their binding strength but adds an additional 4–5 days to the workflow described in Figure 1F (Figure 2D).

This workflow was performed for 9 previously characterized PDZ domains from 6 eukaryotic proteins, including the ERBIN PDZ domain, Disks large homologue 4 (DLG4, also known as postsynaptic density protein-95) PDZ3, SCRIB, PDZ1, and PDZ3, Multiple PDZ domain protein (MPDZ) PDZ10, T-lymphoma invasion and metastasis-inducing protein 1 (TIAM1) PDZ1, and Membrane-associated guanylate kinase WW and PDZ domain-containing protein 1 (MAGI1) PDZ2, 3, and 4. We successfully measured the specificities of 8 out of 9 of these PDZs in good agreement with previous work, with the TIAM1 PDZ1 being the only PDZ that failed in the B2H assay. Additionally, 3 PDZ domains that had been attempted in phage display but failed to produce motifs were successfully characterized, including DLG4 PDZ1 and PDZ2, and Amyloid-beta A4 precursor protein-binding family A member 1 (APBA1) PDZ1 (Figure S2 and S3).

Predictions of Known Interactions Are Improved by Models Incorporating Lower Affinity Binders.

One goal of this project is to produce a system that captures a more representative group of peptides that interact with domains like PDZs in order to improve our capacity to predict novel interactions. Utilizing published phage display results as a comparison,¹⁵ we interrogated the data from our selections of the Natural Library. After filtering selection data to enrich for high-confidence binders (see Materials and Methods), we plotted the distribution of recovered known “annotated” binders across GFP populations (Figure 3A). Annotated binders are found in all three GFP populations. However, in most cases the lower GFP populations contain a greater proportion of these peptides, supporting the notion that there is a wealth of presumably lower-affinity interactions that may be overlooked by methods biased toward high affinity. Furthermore, the absolute recovery of known interactors by MR-B2H selections is 4–10-fold higher than phage display (Figure 3B). To control for false discovery rates and directly compare to phage display results, we then built position weight matrices (PWMs) from MR-B2H-derived peptides found in either low or high GFP populations, as well as phage display-derived peptides, and scored each C-terminus of known partners. Such PWMs are a convenient summary of the specificity data contained in the selections; positive scores correspond to good matches to peptides observed frequently in the data. With a false-discovery rate set to 0.01, we found that in most cases, the highest number of known binders passed the cutoff when compared to the PWM derived from low GFP B2H peptides (Figure 3C). Stated simply, PWMs derived from a broader set of PPI affinities and characteristics recover a higher number of known binders than PWMs based on more restrictive information. Phage-display is known to select strong as well as hydrophobic binders²⁷ but lacks the affinity range captured by MR-B2H selection. Hence,

binding profiles (*e.g.*, expressed as PWMs) derived from phage-display often fail to capture important interactions. In sum, these results demonstrate that incorporating a rich spectrum of interaction affinities into our models significantly improves our ability to predict known interactions.

The Prediction of Novel Interaction Partners Underscores the Utility of MR-B2H Data.

The agreement between our MR-B2H data and previous screens carried out by phage display^{14,15} and protein microarrays^{16,39} validates the accuracy of our technique. In addition, the improved prediction of known *in vivo* binding partners with MR-B2H data (Figure 3) demonstrates the importance of the expanded affinity profile for detecting real full-length protein interactions. However, many recovered peptides were not from known interacting partners and we next set out to determine if these represent false positive or novel true positive interactions. That said, it is important to note that the possibility of interaction does not imply biological function or purpose, but simply that if two proteins are present in the same cell they have the potential to interact. Further, it is beyond the scope of this work to thoroughly investigate the biological implications of each novel prediction of PDZ and peptide pairings. Rather, we aim to demonstrate this tool is able to accurately predict the possibility of *in vivo* interacting partners that others might follow up in future works. With this goal in mind, we turned to another hybrid assay that would express a set of full-length proteins in human cells to test for positive interactions. This approach will allow us not only to test if novel predictions are functional *in vivo*, but also to determine if selected protein–peptide interactions from our Natural Library are predictive of full-length protein interactions.

Since the PDZ domain is often found in proteins that orchestrate scaffolding interactions at the cell membrane,⁴⁰ we chose to focus our human cell validation on DLG4, a neuronal membrane protein containing 3 PDZ domains. We characterized the target specificity of each PDZ domain in DLG4 (Figures S2 and S3) and not surprisingly found that many of the predicted partners map to intermembrane proteins. To validate these predictions, we used the mammalian membrane 2-hybrid system (MaMTH)⁴¹⁻⁴³ to investigate a small set novel predicted partners (Figure 4A). This system utilizes the reconstitution of a split ubiquitin molecule that is dependent on the ability of two interacting partners to bring these halves in close proximity to one another. The reconstitution leads to cleavage of a transcription factor and its subsequent transport to the nucleus where it activates the reporter gene luciferase. In this way, luciferase activation is a measure of the interacting partners at the membrane. Using this method, we tested a set of five predicted partners, produced as full-length proteins (Four novel: Activin receptor type-2B (ACVR2B), Leucine-rich repeat and fibronectin type-III domain-containing protein 4 (LRFN4), Plasma membrane calcium-transporting ATPase 2 (ATP2B2), and Solute carrier family 2 facilitated glucose transporter member 1 (SLC2A1); One known: Frizzled-7 (FZD7)) with each of the PDZ domains alone as well as the full-length protein (Figure S4). Luciferase activity was significantly increased over the controls for all of these interactions with the full-length protein (Figure 4B) and for 4/5 interacting partners with the single PDZ domains (Figure S4). Interestingly, while still significantly above controls, the luciferase activity decreased for 4 out of 5 of the interactions with the full-length protein in comparison to any of the individual PDZ domains. In addition, the one

exception, SLC2A1, significantly increases luciferase output with the full-length protein but not a fragment that contains all 3 PDZ domains. These differences may be due to variable expression levels depending on whether a truncated or full-length version of each protein is employed. However, the results may also imply that the PDZ interaction is weak, and the doubling of reporter output with the full-length protein is due to an additional interaction in the protein. This secondary interaction may or may not require the weak PDZ interaction, but determining this interplay would require detailed follow up investigations. Regardless, these results demonstrate that the MR-B2H data was able to successfully predict 4/4 novel full-length interacting partners with DLG4.

The MR-B2H Provides Insight into the Specificities of Bacterial PDZ Domains.

The specificities of prokaryotic PDZ domains have been largely ignored outside of a handful of *E. coli* proteins. One potential complication for the prokaryotic domains is the genes that code for them often contain a protease domain. Expressing a functional protease in a hybrid assay could prove problematic. However, structures of the DegS protein reveal that the protease domain is in close proximity to the peptide bound to the PDZ domain,⁴⁴ indicating that this interaction may influence specificity. Therefore, to test the ability of the MR-B2H system to characterize these prokaryotic domains, we first focused on the DegS protein. Using the random library, we tested DegS as the PDZ domain alone, as the full-length protein including the protease domain, and a mutant version of the protein that reduces the catalytic activity of the protease (Figure 5). The DegS peptide preference is in agreement with prior investigation of this domain.³⁷ Further, we find only subtle differences between the 3 versions of the proteins, with the full-length versions enriching for an arginine at the -1 position not found with the PDZ domain alone. Therefore, the PDZ domain alone appears to provide at least an approximation of the full-length specificity. Also, our experiments demonstrated no signs of toxicity when we overexpress this *E. coli* protease in our *E. coli*-based system, indicating that investigation of other PDZ proteases are not likely to be problematic in our system. Next, we screened the PDZ domains of 3 other *E. coli* proteins, Periplasmic serine endoprotease DegP (DegP), Periplasmic pH-dependent serine endoprotease DegQ (DegQ), and Prc. Only Prc failed to enrich for specific peptides. DegP and DegQ both contain 2 PDZ domains but only one of each produced a motif (Figure 5). The three failed domains may require additional sequences to be functional or even require expression as full-length proteins, but this would make it difficult to distinguish which PDZ was responsible for binding the peptide. Regardless, for the PDZs successfully characterized, their peptide preferences are in agreement with published work^{45,46} validating the MR-B2H system's ability to characterize prokaryotic domains, in addition to their eukaryotic counterparts.

CONCLUSIONS

Here we describe a rapid, sensitive, and accurate hybrid system enabling the parallel characterization of PPIs in as little as 3 days. Relative activity of interactions can be obtained in as little as four to five additional days. We showcase the MR-B2H using the PDZ domain, one of the most prevalent SLiDs in human. We successfully apply this method to 8/9 human PDZ domains previously characterized by peptide phage display, as well as 2

domains that had failed. Additionally, we identify a novel motif for the human APBA1 PDZ1 domain. Finally, we characterize 3 PDZ domains from *E. coli* and demonstrate the successful capture of protein–ligand interactions even when these PDZ domains are expressed as full-length proteins and include a functional protease. In all cases, the specificities produced are in agreement with prior work (when available), demonstrating the accuracy of the technique. As a result, the combined speed, throughput, and accuracy of this method should make it useful for future studies of PDZ domain proteins as well as other common SLiDs. In addition, the simplicity of the assay allows for its employment by any lab skilled in basic molecular biology techniques.

The goal of any high-throughput technique is to accurately capture the true activity of a protein *in vivo*. Therefore, our goal was to create a tool that would improve our ability to predict *in vivo* binding partners, first for PDZ domains, but with the potential to expand to other protein domains. While prior methods to characterize these domains have sacrificed throughput or sensitivity,^{14-16,39} the MR-B2H system allows the screen of large libraries without sacrificing the capture of a broad range of affinities. We believed that the capture of a wide range of affinities without sacrificing scale would enable better prediction of *in vivo* binding partners. We find this to be true as our MR-B2H data is better able to predict known *in vivo* interacting partners than prior works, and our MaMTH studies demonstrate the accurate prediction of novel binding partners.

The application of this method could be beneficial in several ways. First, the simple selection of interactions at high stringency from plates can provide a consensus binding motif for a protein within 3 days. Second, application of low stringency selections, or low GFP recovery, can provide a broader understanding of the domain-peptide interaction. Scaled and applied to a large set of domains, the mechanistic fundamentals of the domain's function may be gleaned. As compared to phage display, the nature of our method sidesteps the laborious and often difficult need to generate a functional purified protein *in vitro*. This approach has been applied to many DNA-binding domains^{28-30,47,48} and the sensitivity of the MR-B2H system may enable a similar approach to understand SLiD function. Finally, the simplicity and throughput of this technique allows the seamless application of multiple libraries. The comparison of the Natural Library with a purely random library allows us to investigate what peptides a protein might interact with in a given cell, as well as how mutations in the peptide might alter the interaction. Therefore, this comprehensive approach allows us to understand how variation in peptides alter protein recognition. Further, the GFP assay allows for a rough quantification of how slight variation in the peptide may influence the relative strength of the interaction. As it has become clear that variation across human genomes is prevalent,² understanding variation in PPIs, especially those that can influence low affinity interactions, will be critical as we try to decipher whether a variant is harmful or benign.

MATERIALS AND METHODS

Library and Plasmid Construction.

The random and natural libraries were generated as previously described.⁴⁹ Plasmids were generated *via* routine restriction enzyme cloning, and will be made available at Addgene. Additional details can be found in the Supporting Information.

Selections on Plates.

Selections in the B2H assay were performed as previously described for BIH selections^{30,36,48} with cells plated on minimal media (NM) lacking histidine and containing various levels of 3AT. Surviving colonies were sequenced.

Selections in Liquid Culture.

To separate positive PDZ-peptide interactions by GFP expression, 10^9 cells were plated on NM media plates lacking histidine and containing 1 mM 3AT and grown for 24 h to enrich for the somewhat rare positive interactions (typically $<10^5$ colonies survived out of the 10^9 plated). Surviving colonies were scraped and DNA was extracted to yield “round 1” libraries. 50 ng of round 1 libraries, or the 5 alanine peptide negative control, were retransformed with 1 μ g of the same PDZ plasmid. 10^6 cells were grown in selective NM media supplemented with 1 mM 3AT for up to 36 h. Four GFP populations (negative, low, medium, and high) were sorted and collected from a time point when GFP approached its maximum. Cells from these sorted pools were plated on rich media and grown overnight. The surviving colonies were pooled, DNA barcoded, and deep-sequenced.

Next-Generation Sequencing.

PCR primers that included the Illumina adaptor sequences were used to amplify peptide coding sequences from the plate selections as well as recovered GFP populations. Reactions were gel purified and mixed in equimolar ratios before being run on a NextSeq500 with 30% PhiX DNA to increase diversity. Results were demultiplexed, and peptide counts were tallied.

MaMTH Validation Assays.

MaMTH assays were carried out as previously described.⁴¹ In short, bait and prey plasmids were transfected into HEK293 MaMTH reporter cells, expression was induced with tetracycline, and luciferase activity was measured 24 h post induction.

Bioinformatics Analysis.

For each PDZ domain a PWM was generated from the peptides recovered from each population considered, including all plate stringencies and GFP populations. For comparison, PWMs were also produced from the previously published phage display-derived peptides.^{14,15} In addition, the interacting partners of each target protein were collected from the protein-protein interaction database MENTHA.³⁵ The C-terminus of each protein annotated as a partner was scored for each PWM. After scoring, a threshold needed to be set for separating positive and negative predictions. The simplest approach is to

set a common log-likelihood threshold for all motifs. Unfortunately, log-likelihood distributions substantially vary across motifs, and the same threshold value may result in underestimating occurrences for one motif and overestimating them for another. Therefore, several score distribution-based approaches for threshold selection were proposed.⁵⁰ We selected a method based on controlling the false-positive rate, which has proven to be the least biased by the information content of the motif and therefore the most consistent between motifs. Consequently, we compared the performance of each PWM in the data set of known partners by applying a fixed FDR of 0.01.⁵¹

The accumulated hydrophobicity value was calculated as the sum of each amino acid hydrophobicity weight, multiplied by each amino acid-normalized frequency in the PWM over each position.¹⁵

All the data from the screenings is available at <http://kimweb1.cbr.utoronto.ca/library/b2h/>. The code used to analyze and preprocess the data can be found here: <https://gitlab.com/kimlab/ngskit>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

NIH T32 #GM066704 supported D.M.I. in this work. P.M.K. and C.C.V. were funded by grants of the Canadian Institute for Health Research (PJT-153279) and the Canadian Cancer Society Research Institute (#702884). MaMTH work performed by the lab of I.S. was supported by CCSRI (703889), Genome Canada (9428), Government of Ontario (RE08-009) and ORF-DIG-Stagljar. We thank current and former members of the Noyes lab for the thoughtful suggestions on the work and manuscript. We thank T.M.C. for her valued insight, support, and feedback on the manuscript. We would also like to thank the NYU Langone Genome Technology Center for their help in determining the best approach for sequencing our results.

REFERENCES

- (1). Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, and Wiuf C (2008) Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A* 105 (19), 6959–64. [PubMed: 18474861]
- (2). Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526 (7571), 68–74. [PubMed: 26432245]
- (3). The Genomes Project, C. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491 (7422), 56–65. [PubMed: 23128226]
- (4). Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, and Exome Aggregation, C. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536 (7616), 285–91. [PubMed: 27535533]

- (5). Pawson T, and Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300 (5618), 445–52. [PubMed: 12702867]
- (6). Charlotiaux B, Zhong Q, Dreze M, Cusick ME, Hill DE, and Vidal M (2011) Protein-protein interactions and networks: forward and reverse edgetics. *Methods Mol. Biol* 759, 197–213. [PubMed: 21863489]
- (7). Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ, and Davey NE (2014) Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev* 114 (13), 6733–6778. [PubMed: 24926813]
- (8). Van Roey K, Dinkel H, Weatheritt RJ, Gibson TJ, and Davey NE (2013) The switches. *ELM Resource: A Compendium of Conditional Regulatory Interaction Interfaces*. *Sci. Signaling* 6 (269), rs7.
- (9). Sundell GN, Arnold R, Ali M, Naksukpaiboon P, Orts J, Guntert P, Chi CN, and Ivarsson Y (2018) Proteome-wide analysis of phospho-regulated PDZ domain interactions. *Mol. Syst. Biol.* 14 (8), No. e8129. [PubMed: 30126976]
- (10). Sakarya O, Conaco C, Egecioglu O, Solla SA, Oakley TH, and Kosik KS (2010) Evolutionary expansion and specialization of the PDZ. domains. *Mol. Biol. Evol* 27 (5), 1058–69. [PubMed: 20026484]
- (11). Muley VY, Akhter Y, and Galande S (2019) PDZ Domains Across the Microbial World: Molecular Link to the Proteases, Stress Response, and Protein Synthesis. *Genome Biol. Evol* 11 (3), 644–659. [PubMed: 30698789]
- (12). Tramontano M, Andrejev S, Pruteanu M, Klunemann M, Kuhn M, Galardini M, Jouhten P, Zelezniak A, Zeller G, Bork P, Typas A, and Patil KR (2018) Nutritional preferences of human gut bacteria reveal their metabolic idiosyncrasies. *Nat. Microbiol* 3 (4), 514–522. [PubMed: 29556107]
- (13). Sigrist CJ, de Castro E, Cerutti L, Cucho BA, Hulo N, Bridge A, Bougueleret L, and Xenarios I (2012) New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–7. [PubMed: 23161676]
- (14). 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, and Sidhu SS (2008) A specificity map for the PDZ domain family. *PLoS Biol.* 6 (9), No. e239. [PubMed: 18828675]
- (15). Ivarsson Y, Arnold R, McLaughlin M, Nim S, Joshi R, Ray D, Liu B, Teyra J, Pawson T, Moffat J, Li SS, Sidhu SS, and Kim PM (2014) Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc. Natl. Acad. Sci. U. S. A* 111 (7), 2542–7. [PubMed: 24550280]
- (16). Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaja LA, and MacBeath G (2007) PDZ domain binding selectivity is optimized across the mouse proteome. *Science* 317 (5836), 364–9. [PubMed: 17641200]
- (17). Luck K, Charbonnier S, and Trave G (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* 586 (17), 2648–61. [PubMed: 22709956]
- (18). Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, and Vidal M (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437 (7062), 1173–8. [PubMed: 16189514]
- (19). Walhout AJM, Sordella R, Lu XW, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, and Vidal M (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287 (5450), 116–122. [PubMed: 10615043]
- (20). Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, and Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A* 98 (8), 4569–4574. [PubMed: 11283351]

- (21). Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang LY, 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 CWV, Figeys D, and Tyers M (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415 (6868), 180–183. [PubMed: 11805837]
- (22). Rolland T, Tasan M, Charleatoux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang XP, Ghamsari L, Balcha D, Beggs BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan CY, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang SL, Kiros R, Lin GN, Luck K, MacWilliams A, Menche JR, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejada AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabasi AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, and Vidal M (2014) A Proteome-Scale Map of the Human Interactome Network. *Cell* 159 (5), 1212–1226. [PubMed: 25416956]
- (23). Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wuhr M, Chick J, Zhai B, Kolipakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, and Gygi SP (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162 (2), 425–440. [PubMed: 26186194]
- (24). Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, and Wanker EE (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122 (6), 957–68. [PubMed: 16169070]
- (25). Vincentelli R, Luck K, Poirson J, Polanowska J, Abdat J, Blemont M, Turchetto J, Iv F, Ricquier K, Straub ML, Forster A, Cassonnet P, Borg JP, Jacob Y, Masson M, Nomine Y, Reboul J, Wolff N, Charbonnier S, and Trave G (2015) Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* 12 (8), 787–93. [PubMed: 26053890]
- (26). Tonikian R, Zhang Y, Boone C, and Sidhu SS (2007) Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat. Protoc* 2 (6), 1368–86. [PubMed: 17545975]
- (27). Luck K, and Trave G (2011) Phage display can select overhydrophobic sequences that may impair prediction of natural domain-peptide interactions. *Bioinformatics* 27 (7), 899–902. [PubMed: 21300698]
- (28). Oakes BL, Xia DF, Rowland EF, Xu DJ, Ankoudinova I, Borchardt JS, Zhang L, Li P, Miller JC, Rebar EJ, and Noyes MB (2016) Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing. *Nat. Commun* 7, 10194. [PubMed: 26738816]
- (29). Noyes MB (2012) Analysis of specific protein-DNA interactions by bacterial one-hybrid assay. *Methods Mol. Biol* 786, 79–95. [PubMed: 21938621]
- (30). Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH, and Wolfe SA (2008) A systematic characterization of factors that regulate *Drosophila* segmentation *via* a bacterial one-hybrid system. *Nucleic Acids Res.* 36 (8), 2547–60. [PubMed: 18332042]
- (31). Dove SL, Joung JK, and Hochschild A (1997) Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* 386 (6625), 627–30. [PubMed: 9121589]
- (32). Joung JK, Ramm EI, and Pabo CO (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A* 97 (13), 7382–7. [PubMed: 10852947]
- (33). Giesecke AV, Fang R, and Joung JK (2006) Synthetic protein-protein interaction domains created by shuffling Cys2His2 zinc-fingers. *Mol. Syst. Biol* 2, 2006–2011.

- (34). Skelton NJ, Koehler MF, Zobel K, Wong WL, Yeh S, Pisabarro MT, Yin JP, Lasky LA, and Sidhu SS (2003) Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. *J. Biol. Chem* 278 (9), 7645–54. [PubMed: 12446668]
- (35). Calderone A, Castagnoli L, and Cesareni G (2013) mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* 10 (8), 690–1. [PubMed: 23900247]
- (36). Meng X, Brodsky MH, and Wolfe SA (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol* 23 (8), 988–94. [PubMed: 16041365]
- (37). Sohn J, Grant RA, and Sauer RT (2007) Allosteric activation of DegS, a stress sensor PDZ protease. *Cell* 131 (3), 572–83. [PubMed: 17981123]
- (38). de Regt AK, Baker TA, and Sauer RT (2015) Steric clashes with bound OMP peptides activate the DegS stress-response protease. *Proc. Natl. Acad. Sci. U. S. A* 112 (11), 3326–31. [PubMed: 25733864]
- (39). Vouilleme L, Cushing PR, Volkmer R, Madden DR, and Boisguerin P (2010) Engineering peptide inhibitors to overcome PDZ binding promiscuity. *Angew. Chem., Int. Ed* 49 (51), 9912–6.
- (40). Lee HJ, and Zheng JJ (2010) PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun. Signaling* 8 (1), 8.
- (41). Petschnigg J, Groisman B, Kotlyar M, Taipale M, Zheng Y, Kurat CF, Sayad A, Sierra JR, Mattiazzi Usaj M, Snider J, Nachman A, Krykbaeva I, Tsao MS, Moffat J, Pawson T, Lindquist S, Jurisica I, and Stagljar I (2014) The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat. Methods* 11 (5), 585–92. [PubMed: 24658140]
- (42). Yao Z, Darowski K, St-Denis N, Wong V, Offensperger F, Villedieu A, Amin S, Maly R, Aoki H, Guo H, Xu Y, Iorio C, Kotlyar M, Emili A, Jurisica I, Neel BG, Babu M, Gingras AC, and Stagljar I (2017) A Global Analysis of the Receptor Tyrosine Kinase-Protein Phosphatase Interactome. *Mol. Cell* 65 (2), 347–360. [PubMed: 28065597]
- (43). Saraon P, Grozavu I, Lim SH, Snider J, Yao Z, and Stagljar I (2017) Detecting Membrane Protein-protein Interactions Using the Mammalian Membrane Two-hybrid (MaMTH) Assay. *Curr. Protoc Chem. Biol* 9 (1), 38–54. [PubMed: 28253435]
- (44). Wilken C, Kitzing K, Kurzbauer R, Ehrmann M, and Clausen T (2004) Crystal structure of the DegS stress sensor: How a PDZ. domain recognizes misfolded protein and activates a protease. *Cell* 117 (4), 483–94. [PubMed: 15137941]
- (45). Krojer T, Sawa J, Schafer E, Saibil HR, Ehrmann M, and Clausen T (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature* 453 (7197), 885–90. [PubMed: 18496527]
- (46). Kolmar H, Waller PR, and Sauer RT (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: specificity for cleavage sites and substrate conformation. *J. Bacteriol* 178 (20), 5925–9. [PubMed: 8830688]
- (47). Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, and Wolfe SA (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 133 (7), 1277–89. [PubMed: 18585360]
- (48). Persikov AV, Wetzel JL, Rowland EF, Oakes BL, Xu DJ, Singh M, and Noyes MB (2015) A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.* 43 (3), 1965–84. [PubMed: 25593323]
- (49). Persikov AV, Rowland EF, Oakes BL, Singh M, and Noyes MB (2014) Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. *Nucleic Acids Res.* 42 (3), 1497–508. [PubMed: 24214968]
- (50). Yip KY, Utz L, Sitwell S, Hu X, Sidhu SS, Turk BE, Gerstein M, and Kim PM (2011) Identification of specificity determining residues in peptide recognition domains using an information theoretic approach applied to large-scale binding maps. *BMC Biol.* 9, 53. [PubMed: 21835011]
- (51). Dabrowski M, Dojer N, Krystkowiak I, Kaminska B, and Wilczynski B (2015) Optimally choosing PWM motif databases and sequence scanning approaches based on ChIP-seq data. *BMC Bioinf.* 16, 140.

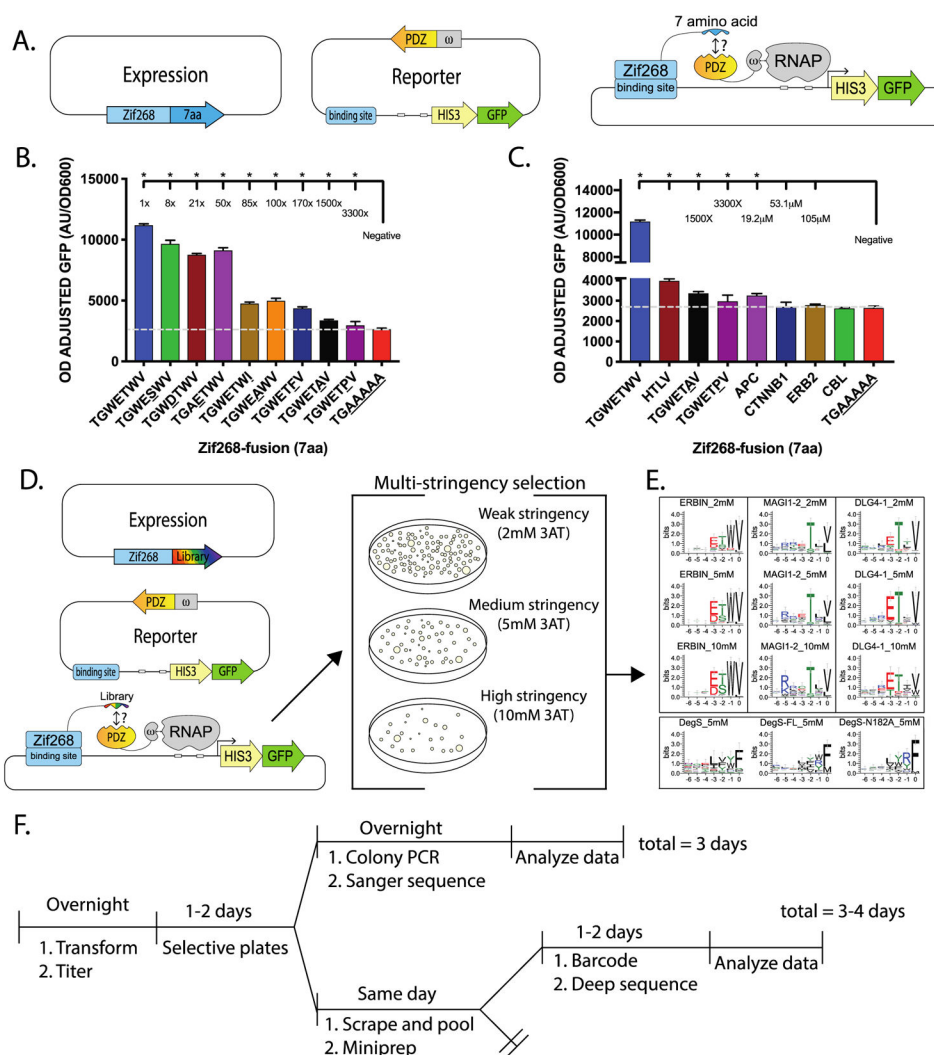


Figure 1.

A multireporter bacterial 2-hybrid (MR-B2H) tool for the selection of PDZ-peptide interactions. (A) Vectors used and selection scheme are shown. A positive interaction between a Zif268-peptide and an omega-PDZ domain recruits RNA-polymerase (RNAP) and drives activation of the reporters HIS3 and GFP. (B) GFP expression driven by known ERBIN-peptide interactions. The reduction in IC50 measurements are shown at the top for each peptide. The consensus peptide (TGWETWV) was mutated at the underlined positions to the indicated amino acid. (C) GFP expression driven by the C-termini of proteins the ERBIN PDZ domain is known to interact with. Affinities of some known interactions are noted above. (D) Scheme of a selection of positive PDZ-peptide interactions on plates with varying concentrations of 3-AT. (E) Motifs of pooled surviving colonies from plate selections at 2, 5, and 10 mM 3AT. Bottom: Selections of prokaryotic PDZ domains and full-length proteins. (F) Timeline of workflow.

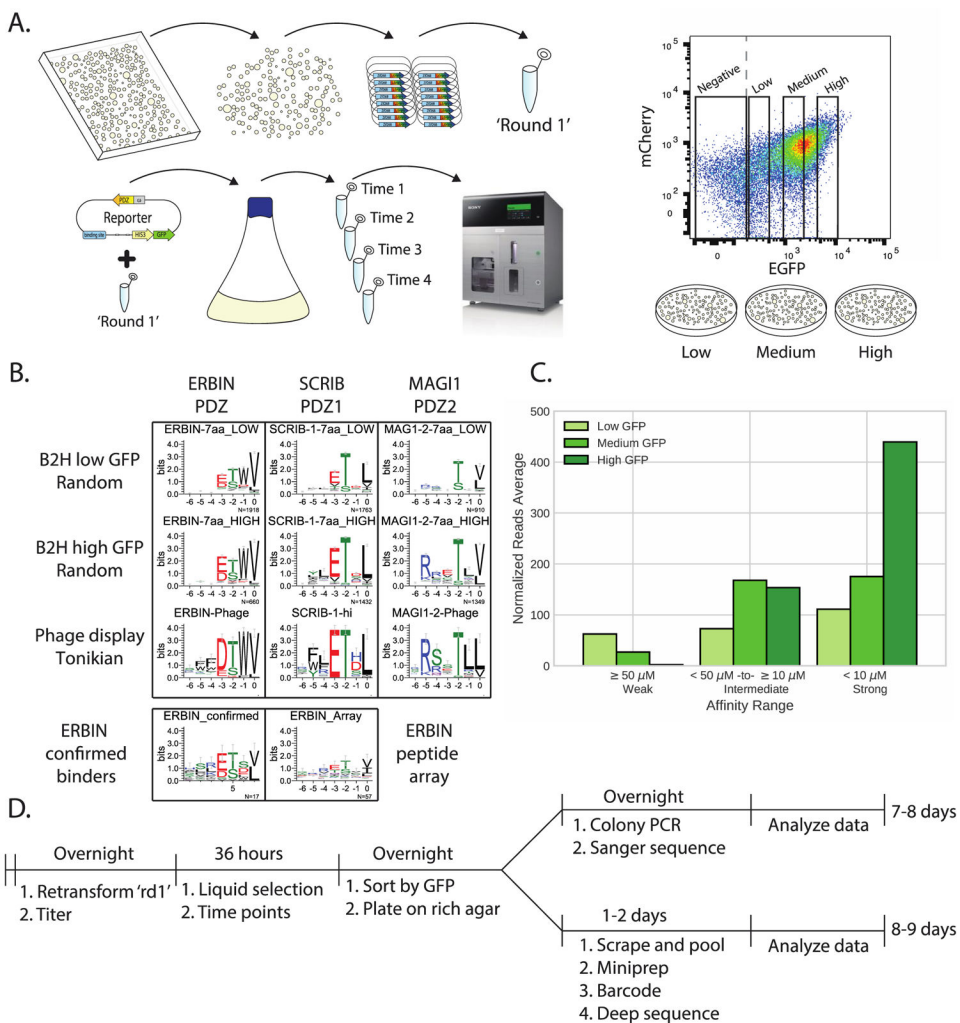


Figure 2. MR-B2H system captures a wide range of functional affinities. (A) Transformations are performed as in Figure 1D. Surviving colonies are pooled and minipreped to obtain an enriched library termed “Round1”. Retransformed DNA libraries are grown in liquid media until GFP approaches maximum fluorescence (typically 36 h). Three populations of cells with approximately 5-fold differences in mean GFP fluorescence are sorted and sequenced. (B) Motifs recovered from cell populations demonstrating either low or high GFP fluorescence are shown and compared to motifs determined by phage display and motifs of the C-termini of known *in vivo* binding partners. (C) Twenty-five proteins with known affinities for either the DLG4, ERBIN, or SCRIB PDZ domains were identified by the MR-B2H screen and the distribution of the reads for these peptides across the three populations of cells, discriminated on the basis of GFP fluorescence intensities, is plotted. Weak affinities are $>50 \mu\text{M}$, intermediate affinities = $10\text{--}50 \mu\text{M}$, and strong affinities are $<10 \mu\text{M}$. (D) A simplified timeline for experiment is depicted. This workflow picks up where Figure 1F ends.

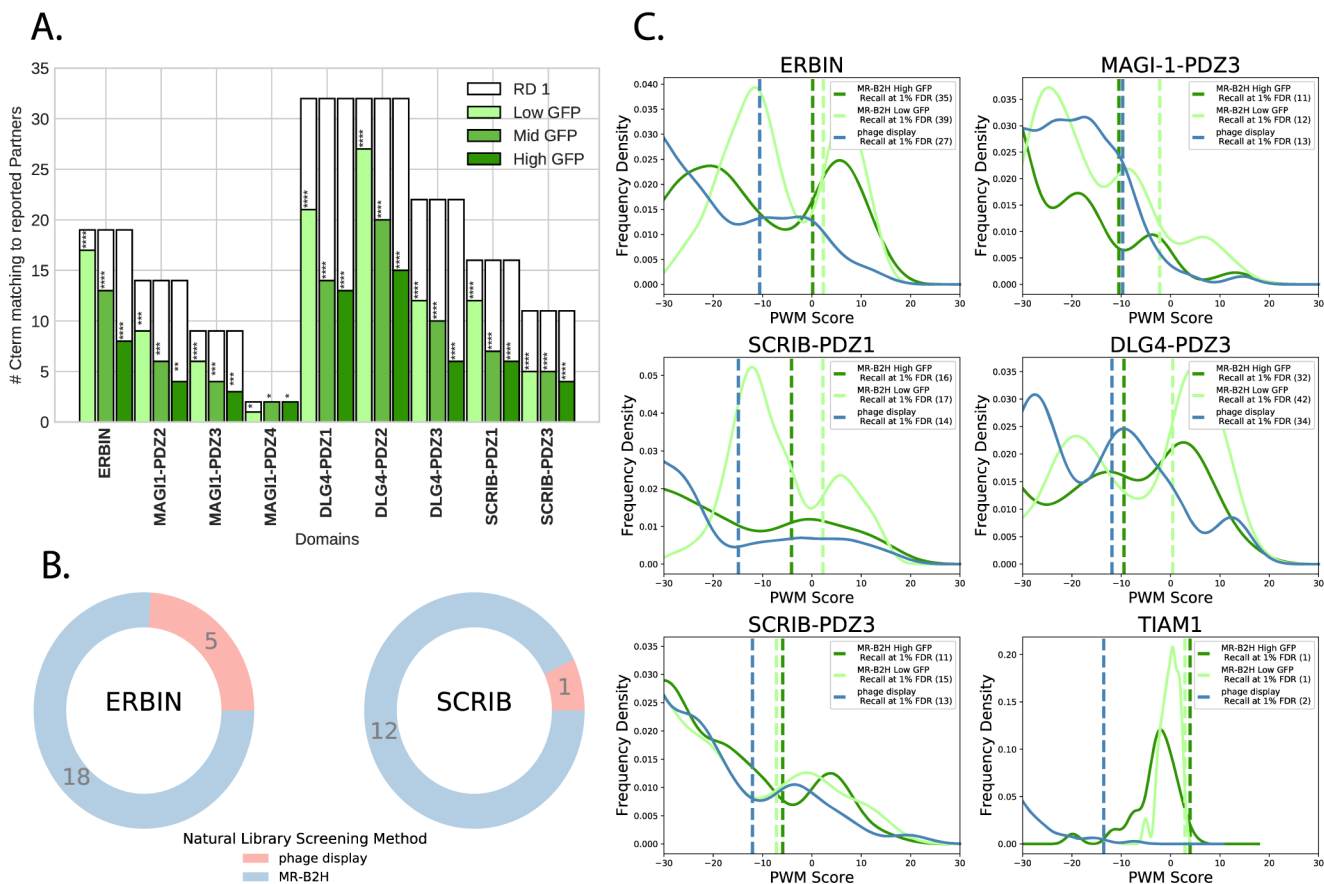


Figure 3. Capture of broad GFP outputs improve the accuracy of predicting known PDZ interacting partners. (A) Plotted counts of selected peptides matching C-termini of annotated *in vivo* partners. Empty box indicates total annotated C-termini passing the reads-threshold after round 1 selection. Green bars depict the number of annotated matches within each GFP level. The level of significance (hypergeometric test) is indicated by the number of stars on the top of the bar; one <math><0.05</math>, two <math><0.01</math>, three <math><0.001</math>, and four <math><0.0001</math>. (B) Comparison of the number of known partners recovered by either phage display or MR-B2H selection of the Natural Library for the ERBIN PDZ domain, and for SCRIB PDZ1 and PDZ3 domains. (C) Histograms of the PWM score (see Supporting Information) show distributions of C-termini of known partners for each domain. Each set was scored with 3 different PWMs; one was built from the results of phage display selections, colored in blue, and the other two were built using MR-B2H high GFP and low GFP selections, colored dark green and light green, respectively. The score threshold to satisfy a FDR of 1% was calculated in each case and plotted as vertical dashed lines, and the count of C-termini over the score threshold can be found in the legend.

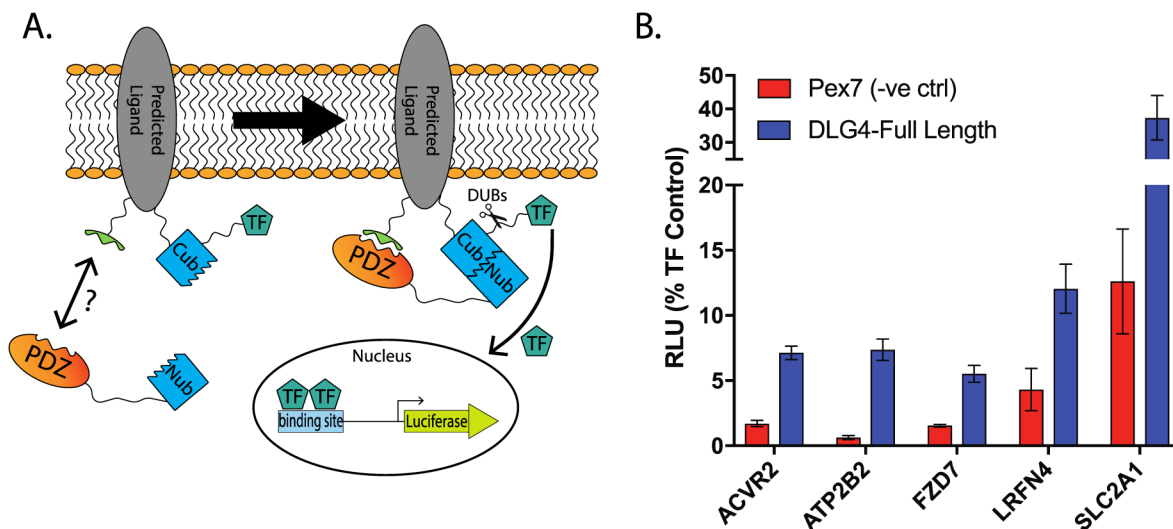


Figure 4. Specificities produced by the MR-B2H system improve the prediction of novel DLG4 interacting partners at the cell membrane. (A) Schematic of the mammalian membrane 2-hybrid (MaMTH). If the PDZ domain interacts with the C-terminus of its partner, a pseudoubiquitin molecule is reconstituted. Deubiquitinases cleave off the fused transcription factor, which translocates to the nucleus and activates expression of a luciferase reporter. (B) MaMTH luciferase expression as a fraction of a control transcription factor with a known DLG4 interaction partner (FZD7) and four MR-B2H predicted partners.

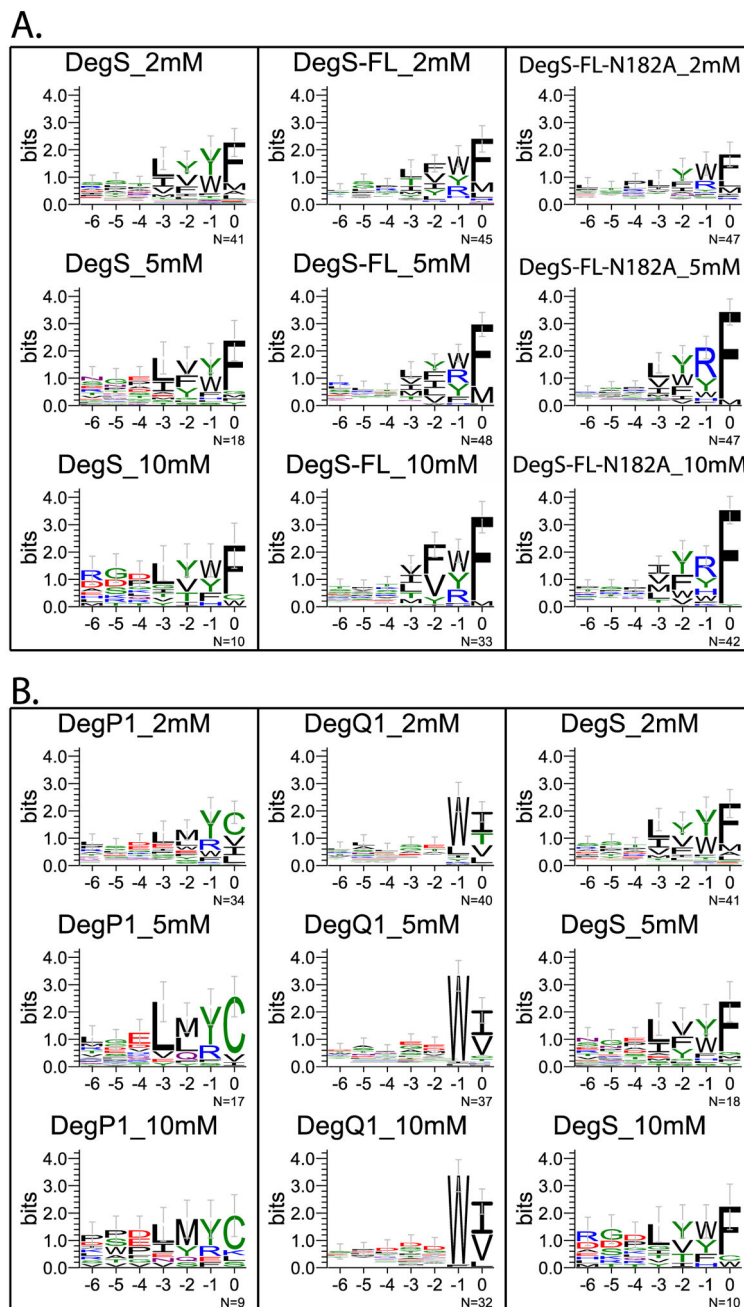


Figure 5. Peptides selected by MR-B2H to bind *E. coli* PDZ domains. Selections were done on plates at the 3AT stringency indicated above each motif. (A) DegS was tested 3 ways: as the PDZ alone (DegS), the full-length protein (DegS-FL), and a mutant version of the full-length protein (N182A) that significantly reduces the activity of the protease domain. (B) The peptides selected to the PDZ domains of DegP and DegQ are also shown in comparison to DegS.