Quantitative fragmentomics allow affinity mapping of interactomes

Gergo Gogl[1*], Boglarka Zambo[2], Camille Kostmann[1], Alexandra Cousido-Siah[1], Bastien Morlet[2], Fabien Durbesson[3], Luc Negroni[2], Pascal Eberling[2], Pau Jane[1], Yves Nomine[1], Andras Zeke[4], Søren Østergaard[5], Elodie Monsellier[1], Renaud Vincentelli[3], Gilles Trave[1]*

1 Équipe Labellisée Ligue 2015, Département de Biologie Structurale Intégrative, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Université de Strasbourg, 1 rue Laurent Fries, BP 10142, F-67404 Illkirch, France

2 Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Universite de Strasbourg, 1 rue Laurent Fries, BP 10142, F-67404 Illkirch, France

3 Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257 CNRS-Aix-Marseille Université, Marseille, France

4 Bioinformatics Research Group, Research Centre for Natural Sciences, Magyar tudosok korutja 2, 1117 Budapest, Hungary

5 Novo Nordisk A/S, Global Research Technologies, Novo Nordisk Research Park, 2760 Maaloev, Denmark

* co-corresponding authors

Gergo Gogl: goglg@igbmc.fr

Gilles Trave: traveg@igbmc.fr

Abstract

Human protein networks have been widely explored but most binding affinities remain unknown, hindering quantitative interactome-function studies. Yet interactomes rely on minimal interacting fragments displaying quantifiable affinities. Here we measured the affinities of 65,000 interactions involving PDZ domains and their target PDZ-binding motifs (PBM) within a human interactome region particularly relevant for viral infection and cancer. We calculate interactomic distances, identify hot spots for viral interference, generate binding profiles and specificity logos, and explain selected cases by crystallographic studies. Mass spectrometry experiments on cell extracts and literature surveys show that quantitative fragmentomics effectively complement protein interactomics by providing affinities and completeness of coverage, putting a full human interactome affinity survey within realistic reach. Finally, we show that interactome hijacking by the viral PBM of human papillomavirus (HPV) E6 oncoprotein deeply impacts the host cell proteome way beyond immediate E6 binders, illustrating the complex system-wide relationship between interactome and function.

Introduction

Over the past decade, proteome-wide mammalian interactomics have identified few hundreds of thousands of binary interactions (Luck *et al.*, 2017) (Luck *et al.*, 2020) (Huttlin *et al.*, 2021), (Skinnider & Foster, 2021). However data are mostly qualitative, lacking quantitative information on binding strengths, and the coverage of the human interactome remains incomplete. Remarkably, protein interactomes are built upon minimal interacting blocks, consisting of globular folded domains and short disordered linear motifs (Pawson & Nash, 2003) (Diella *et al.*, 2008) (Jadwin *et al.*, 2012) (Tompa *et al.*, 2014) (Kumar *et al.*, 2019). These fragmental interactions occur both intermolecularly and intramolecularly (Pawson & Nash, 2003); their intrinsic affinities, fine-tuned by evolution, rule cooperation, competition and specificities underlying most cellular functions (Kelil *et al.*, 2016) (Ivarsson & Jemth, 2019). They are also key targets for pathologies such as microbial infections and cancers (Davey *et al.*, 2011) (James & Roberts, 2016) (Seo & Kim, 2018) (Sámano-Sánchez & Gibson, 2020). The accurate description and modeling of complex biological systems and their pathological defects will ultimately require quantitative affinity measurements of fragmental interactomes at proteome-wide scale.

The PDZ-PBM interactome is a relevant model for such studies. The human proteome comprises 266 human PDZ (PSD95/DLG1/ZO1) domains (Luck *et al.*, 2012) and about 4000 putative PBMs, defining a network of one million potential interactions. PBMs are mostly (though not exclusively) C-terminal, with their COO⁻ implicated in binding, and are classified based on position -2, being respectively Ser/Thr, hydrophobic or acidic in classes 1, 2, and 3 (Stiffler *et al.*, 2007) (Tonikian *et al.*, 2008). PDZ-PBM interactions are rather transient and promiscuous (Vincentelli *et al.*, 2015) (Gógl *et al.*, 2019) (Gogl *et al.*, 2020) (Jané *et al.*, 2020).

The PDZ-PBM interactome is also prone to pathological perturbations such as viral infection and cancer (James & Roberts, 2016). Viral proteins bearing functional PBMs are found notably in HPV and HTLV oncoviruses as well as in HBV, WNV and coronaviruses including MERS or SARS-CoV2 (Javier & Rice, 2011) (Banks *et al.*, 2012) (James & Roberts, 2016) (Sámano-Sánchez & Gibson, 2020) (Jimenez-Guardeño *et al.*, 2014) (Caillet-Saguy *et al.*, 2021).

Here we measured with unprecedented coverage and sensitivity ~65,000 PDZ-PBM affinities in a defined region of the human interactome targeted by viral PBMs, including those of oncogenic proteins HPV E6 and HTLV1 TAX1. The data, assembled into an open-access database (https://profaff.igbmc.science), define a quantitative interactomic space where proteins are located and clustered according to their binding preferences. Using these data, PDZ-PBM specificities can be depicted by binding profiles and specificity logos, that we interpret by crystallographic studies on selected instances. We also show that amounts of prey proteins captured in cell lysates by exogenous PDZ or PBM baits correlate with the corresponding PDZ-PBM affinities. Finally, we investigated how the interactomic properties of the viral PBM of HPV E6 oncoprotein impact the host cell proteome.

Results

Large-scale affinity mapping of the PDZ-PBM interactome

To explore the PDZ-PBM interactome, we expressed a recombinant PDZome library covering all the 266 known human PDZs (Duhoo, Girault *et al.*, 2019), and we synthetized a 10-mer peptide library of 24 viral and 424 human PBMs, of which 323, 63, 51, and 11 belong to class 1, 2, 3, and to atypical or non-C-terminal subgroups, respectively (Table S1). 8 PBMs harboured post-translational modifications (phosphorylation or acetylation) which may modulate binding specificities (Gógl *et al.*, 2019) (Gogl *et al.*, 2020) (Jané *et al.*, 2020). The viral PBMs included notably 12 PBMs from oncoproteins HPV E6 (11 distinct types) and HTLV1 TAX1. The host PBM list was designed to explore the closest interactomic neighborhood of these oncoviral PBMs, while attempting to sparsely yet evenly cover the sequence diversity of human PBMs.

To quantify dissociation constants, we used the holdup method, a high-throughput comparative chromatographic retention assay we developed previously (Charbonnier *et al.*, 2006) (Vincentelli *et al.*, 2015) (Jané *et al.*, 2020). The assay measures the total and unbound concentrations of reactants at equilibrium, which can be converted into steady-state dissociation constants herein reported as pK_d values (the negative of the base 10 logarithm of the dissociation constant). To adapt the method for higher throughput, we implemented several technical improvements and rigorous benchmarks into our previous protocol (see Methods). The current protocol allows measuring up to 10,000 distinct domain-motif pairs per day, and quantifies their affinities at remarkable sensitivity, with pK_d quantification thresholds (derived from the binding detection thresholds of the source holdup data, above which we can quantify pK_d values) mostly comprised between 4 ($K_d = 100$ μM) and 3.1 ($K_d = 800$ μM). In total, we performed 79,374

single-point holdup experiments on 65,151 interactions, covering ~55% of the interactomic space defined by the 266 human PDZ-domains and the 448 human and viral PBMs (Figure 1). We quantified 18,332 unique dissociation constants. The remaining 46,825 PDZ-PBM affinities were below quantification threshold, thereby defining the "negatome" of the explored interactomic space. Negatome data (Blohm *et al.*, 2014) are often missing in interactomic studies, hindering both the benchmarking and the modeling of interaction networks (Alvarez-Ponce, 2017) (Zhang *et al.*, 2018) (Mei & Zhang, 2019). The complete affinity data (Table S1) are freely accessible on our online database ProfAff (for "Profiling Affinities") (https://profaff.igbmc.science), allowing user-friendly visualisation and analyses.

We complemented our measurements with 395 detailed competitive fluorescence polarization experiments (FP). The quantified affinities show an excellent agreement with those measured by holdup assay (Figure 2A). We also assembled, expanding upon a recent review (Amacher *et al.*, 2020), a literature compendium of affinities previously measured for mammalian PDZ-PBM pairs (Table S1). Considering unavoidable experimental discrepancies (methods, exact lengths of constructs and peptides, species of origin...) our dataset remarkably agrees with this compendium (Figure 2B).

Empirically, we have found that the quantified PDZ-PBM affinities follow a nearly perfect exponential-like distribution (Figure 2C). A handful (less than 0.5%) show pK_d >6, only 8% have a pK_d >5, and the majority has a 3< pK_d <4. Accordingly, the determined negatome remained larger than the quantified interactome (72% vs 28% of the explored space, respectively). Remarkably, the affinity histogram starts to diverge from the exponential trend at pK_d <4, precisely when entering the grey zone where some affinities start to be below quantification threshold (Figure 2C). We assume that using an ideal approach that would measure all affinities with no threshold, the probability density function of domain-motif affinities would expand the observed trend, yet only up to a certain point since the number of potential interactions in the explored space remains finite. By contrast, the literature benchmark is greatly under-represented in weak affinities and follows a normal distribution centered around 4.7 pK_d with a width of 1 pK_d (Figure 2D).

We also curated and assembled a second literature benchmark based on qualitative interactions devoid of affinity data, involving full-length PDZ- and PBM-containing proteins or their fragments (Table S1 and Figure S1). Of that qualitative benchmark, 1248 interactions found a match with at least one PDZ-PBM pair in our measured interactomic space, and 725 (58%) could be assigned to at least one quantified PDZ-PBM affinity.

We also found that only 0.5%, 2.1% and 3.4% out of the 14,839 protein pairs for which we quantified PDZ-PBM affinity constants corresponded to protein pairs recorded by the proteome-wide interactomic study BioPlex (Huttlin *et al.*, 2021) or in the interaction databases IntAct (Kerrien *et al.*, 2012) and Biogrid (Oughtred *et al.*, 2021), respectively. Our exhaustive affinity survey of a defined interactomic space thus indicates that the current coverage of the human interactome remains very sparse.

Topology of the human PDZ-PBM interactomic space

pK_d values can be considered as coordinates of multidimensional interactomic spaces. Our dataset almost exhaustively covers two such spaces, one describing PDZ-binding by PBMs and the other describing PBM-binding by PDZs, respectively comprising 133 and 448 affinity dimensions. The more two PBMs or two PDZs bind to the same targets with similar affinities, the closer they stand in these spaces. Euclidian distances between PDZs or between PBMs can be calculated from their differences in affinities (ΔpK_d = $\Delta\Delta$G/2.303RT, see Methods). These distances allow identifying closest interactomic neighbours (Figure S2A). For example, the PDZ-binding profiles of PBMs of the RhoA guanine nucleotide exchange factor NET1 and of the Ras effector protein RASSF6 are extremely similar, and this is captured by a very short Euclidian distance. Human PBMs that are closest interactomic neighbours of viral PBMs are of particular interest, as they represent their most potent rivals to bind the same host PDZs with similar affinity properties. For instance, the closest neighbors of the major oncoproteins HPV16 E6, HPV18 E6 and HTLV1 TAX1 in our explored interactome are CYSLTR2, NET1, and GRIN2C, respectively.

We clustered PBMs based on their Euclidean distances using an UPGMA (unweighted pair group method with arithmetic mean) approach (Sokal & Michener, 1958) (Figure 1 and S2B). This way, we resolved many clusters of class 1 PBMs with distinct PDZ-binding propensities (Figure S2B). We also identified a few additional clusters, implying PBMs of class 2, class 3 or mixed classes. Most interestingly, PBMs from HTLV1 TAX1 and HPV E6 were mostly clustered in a single clade (Figure S2B). This "oncoviral" clade also includes dozens of human motifs, that hence share common PDZ-binding preferences with the oncoviral PBMs. Most PBMs in this oncoviral clade share the class 1 consensus E-[TS]-x-V-COO⁻ (where x denotes for any residues), often immediately preceded by basic residue(s). These motifs showed generally high binding affinities with PDZ domains such as MAGI1_2, DLG1_2, SCRIB_1, SNTB1, or TX1BP3. The corresponding PDZ-proteins are often involved in maintenance of epithelial basolateral polarity (Etienne-Manneville, 2008). Thus, interactomic distances may be particularly relevant to decipher viral hijacking, and the host motifs within the identified oncoviral clade likely represent an interactomic hot spot for oncoviral interference.

Molecular basis of PBM recognition in the light of the quantitative interactome

For any individual PDZ or PBM measured in our quantitative interactome, binding specificities can be visualized by plotting binding profiles with pK_d values sorted in decreasing order (Figure 1, upper right panel or Figure 3A,D). The steeper the profile slope, the more specific the PBM or PDZ for particular targets within the explored interactome (Vincentelli *et al.*, 2015) (Jané *et al.*, 2020). The PBM-binding preferences of PDZ domains can also be visualized by calculating affinity-weighted sequence logos (see Methods). As compared to conventional

logos built from unranked pools of binding sequences (Schneider & Stephens, 1990), affinity-weighted logos capture determinants of recognition specificity. Binding profiles and specificity logos can be generated and analyzed by users on the ProfAff server.

To illustrate the unvaluable information provided by our affinity data and such modes of representation, we solved the crystal structures of several chosen PDZ-PBM complexes related to profiles and affinity-weighted logos of interest. The logos of PDZ domains SNTB1 and MAGI1_2 indicate preferences for x-R-E-T-x-V-COO$^-$ (Figure 3B) and R-x-E-T-x-V-COO$^-$ (Figure 3C), respectively. The oncoviral PBM of HPV35 E6 satisfies both consensus requirements hence recognizes both PDZ domains (Figure 3A). The structures of HPV35 E6 PBM bound to both SNTB1 and MAGI1_2 resolve their logo differences by revealing the distinctive contributions of Arg residues at p-4 and p-5 of HPV35 E6 in the two complexes (Figure 3B and 3C).

In another example, the profile and the logo of SYNJ2BP indicate that this PDZ-domain is highly promiscuous, with however a mild preference for E-T-x-V-COO$^-$ motifs and a moderate bias for basic residues at upstream (-4 and -5) positions (Figure 3D-F). RPS6KA1 PBM does not fully match SYNJ2BP's logo (Figure 3E) and thus binds only weakly to SYNJ2BP ($pK_d = 3.82$). Upon phosphorylation at p-3, RPS6KA1 gains a negative charge at this position, thereby getting closer to SYNJ2BP's logo. Accordingly, the p-3 phosphorylated variant RPS6KA1 gains affinity to SYNJ2BP ($pK_d = 4.48$; $\Delta pK_d = 0.66$; Figure 3D), as previously published (Gógl *et al.*, 2019). The structure of p-3 phosphorylated RPS6KA1 bound to SYNJ2BP explains this enhanced binding by revealing a network of specific phosphoryl-PDZ contacts (Figure 3E). Nonetheless, p-3 phosphorylated RPS6KA1 binds only mildly to SYNJ2BP as compared to SYNJ2BP's strongest binder in our interactome, the oncoviral PBM of HTLV1 TAX1 ($pK_d = 6.04$; Figure 3D). In the structure of HTLV1 TAX1 PBM bound to SYNJ2BP, p-3 Glu of TAX1 engages contacts reminiscent of those of p-3 phospho-Ser of RPS6KA1, confirming SYNJ2BP's preference for a negative p-3 (Figure 3F). Yet TAX1 matches better than phosphorylated RPS6KA1 to SYNJ2BP's logo thanks to a basic p-4 Arg and a p0 Val (Pro and Leu in RPS6KA1, respectively). Accordingly, in the crystal structures, p-4 Arg of TAX1 establishes favorable H-bonds with a Gln from the α2 helix of SYNJ2BP PDZ, and the Val-bound state of SYNJ2BP PDZ is more compact than its Leu-bound state, due to displacement of the carboxylate-binding GLGF loop likely driven by the larger side-chain of Leu (Figure 3G).

Specificity profiles and affinity-weighted logos also help analyzing interactions that fall out of general rules. According to its logo (Figure 3H), SNX27 normally prefers class-1 PBMs, with Thr or Ser at p-2. Yet, SNX27 also binds ($pK_d = 4.56$) to the class-3 PBM of MERS E viral protein, with Glu at p-2. The structure of SNX27 bound to the PBM of MERS E shows a p-1 Trp of MERS E interacting with a hydrophobic groove of SNX27 including its β2-β3 sheets and a Leu side-chain from its β3-α1 loop. We found out that the closest interactomic neighbor of MERS E based on Euclidian distances of our explored interactome is a class-1 PBM, ARVCF (Figure 3I). While ARVCF is a class 1 PBM with a p-2 Ser, it also possesses a Trp residue at p-1,

similarly to MERS E. The p-1 Trp thus stands out as the major PDZ-binding specificity determinant for MERS E and ARVCF PBMs, more relevant than the class-determining p-2. Indeed, several PDZ-domains showed no clear class preference. For instance, HTRA1 can strongly bind to motifs from all three PBM classes. More generally, our interactomic distance analysis revealed several clusters of PBMs showing similar PDZ-binding patterns while belonging to distinct classes (Figure S2B). While the nature of the antepenultimate p-2 residue is useful for classifying PBMs, it is not fully operative for predicting their PDZ-binding specificities.

PDZ-PBM interaction networks perturbed by oncoviral PBMs in host cell models

While HPV and HTLV1 are both oncogenic viruses, they operate in distinct tissues, respectively epithelial cells (McBride, 2017) and T-lymphocytes (Manel *et al.*, 2005) (Futsch *et al.*, 2018). Upon infection, their oncoproteins thus encounter distinct host proteomes. To investigate the influence of cellular context on oncoviral PBM-PDZ interactomes, we performed affinity purification mass spectrometry (AP-MS) from extracts of lymphoid Jurkat, keratinocyte HaCat and HeLa cells, respectively chosen as models of HTLV1 hosts, HPV hosts, and HPV18-transformed tumors.

Total cell extracts were first probed using 7 PBM baits derived from HPV E6 of 6 distinct HPV types and HTLV1 TAX1 oncoproteins. 34 PDZ-protein preys were identified, whose identities are in excellent agreement with previous reports (Thomas *et al.*, 2016) (Strickland *et al.*, 2018) (Al-Saleem *et al.*, 2019), (Figure 4A, Table S3). In particular, the oncoviral PBMs almost always captured DLG1, SCRIB, TX1BP3 and SNX27 from all cell types, all containing PDZ-domains displaying high affinities for these PBMs. PDZ-containing proteins CASK and MPP7 were also frequently pulled down, although their isolated PDZ domains do not detectably bind to E6 nor TAX1 in our PBM-PDZ interactome. Nonetheless, CASK and MPP7 contain hetero-tetramerization L27 domains (Bohl *et al.*, 2007); these L27 domains can associate with L27 domains of PDZ-containing proteins that directly and detectably bind to E6 and TAX1, such as DLG or LIN7 members.

Next, the same cell extracts were probed using 6 PDZ-domains baits taken from 4 PDZ-proteins that interact ubiquitously with the oncoviral PBMs in all assayed cell lines, namely SCRIB, DLG1, TX1BP3 and SNX27. We detected 73 enriched PBM-proteins, of which 10 belong to the closest neighborhood of E6 and TAX1 PBMs in our explored fragmentomic space (Figure 4B, Table S3). The 63 other enriched PBM-proteins mainly display class-1 motifs with sequences remarkably similar to those of E6 and TAX1 PBMs. Yet and similarly to the PBM-AP-MS experiments, the six PDZ domains also pulled-down, most likely indirectly, prey proteins devoid of C-terminal PBMs or exhibiting PBMs that did not detectably bind those PDZ domains in our PDZ-PBM interactome (see Table S3 for a comprehensive list of all the interaction partners). For example, the E-cadherin and GIT1 proteins, both repeatedly identified by our PDZ-

AP-MS assays, do not have any identifiable C-terminal PBM; they probably indirectly co-precipitate via their respective partners β-catenin (CTNNB1) and β-Pix (ARHGEF7) (Harington & Syrigos, 2000) (Valdes *et al.*, 2011), both containing PBMs that detectably bound to PDZs in our quantified interactome.

In total, we identified 133 unique interactions between the 12 oncoviral PBM baits and 34 endogenous PDZ-protein preys, and 177 between the 6 PDZ baits and 73 endogenous PBM-protein preys. 57% of the PBM bait-PDZ prey interactions showed up in at least two cell types and 31% in all three types, whereas only 25% of the PDZ bait-PBM prey interactions showed up in at least two cell types, and less than 10% in all three types. Noteworthy enough, identification of partners primarily depends upon their expression levels, prone to vary across cell types. Considering all the experiments we performed, 53% of the identified PDZ-proteins but only 39% of the identified PBM-proteins were detected in at least two distinct cell types, and 32% of PDZ-proteins versus 19% of PBM-proteins were detected in all three cell types. This suggests that the host PDZ-proteins targeted by the oncoviral PBMs are more ubiquitously expressed than their host PBM-containing target proteins, and form a potential target group for oncogenesis independently of the host cell type.

Overall, these experiments largely cross-validate the fragmentomic data while illustrating that interactomics using AP-MS from cell extracts may generate false negatives (potential preys that are too weakly expressed) as well as false positives (indirect interactions). Remarkably, comparison to the fragmentomic interactome allows efficient curation of both issues.

Capture of prey proteins by individual PDZ or PBM baits predominantly obeys the law of mass action

The enrichment of prey proteins on a given bait compared to a non-specific control is a key criterium for identifying interaction partners in AP-MS experiments (Keilhauer *et al.*, 2015). The more enriched a protein, the more likely it is to be a *bona fide* interaction partner, suggesting proportionality between fold-enrichment values and amounts of prey-bait complexes formed, despite the washing steps commonly used in AP-MS. In our experiments, the fold-enrichment values of PDZ- or PBM-containing preys captured by particular PBM or PDZ baits in particular lysates were often strongly correlated to pK_d values of the corresponding PDZ-PBM complexes, obtained from our quantitative interactome (Figure 5A). We also observed strong correlations between fold-enrichment values of PDZ- or PBM-containing preys retained by the same PBM or PDZ baits across different cell extracts (Figure 5B). The most straightforward interpretation of such enrichment-enrichment correlations across extracts is that enrichment values, in each extract, were correlated to common pK_d values as in the instance shown in Figure 5A. These observations suggest that the law of mass action, through which affinities dictate the proportions of complexes relatively to their free components, has a predominant impact on the outcome of AP-MS experiments using fixed amounts of baits, such as those we performed here. We however

noticed that both types of correlations (enrichment-affinity and enrichment-enrichment) were overall weaker for experiments using PBM baits than those using PDZ baits (see bottom panels of Figure 5 A and B). While protein preys of PDZ baits generally contain only one C-terminal PBM, preys of PBM baits may contain several PDZ domains, each displaying measurable affinity for the PBM. In such cases, we estimated affinities by assuming additivity of association constants (see Methods). This approximation may contribute to the weaker correlations observed for some experiments using PBM baits.

Based on these observations, that AP-MS applied to different extracts provide different results that can however be related to common affinity properties, we propose to distinguish "interactomes", viewed as intrinsic interaction propensities of biological systems, from "complexomes", viewed as extrinsic interactions occurring in particular contexts. Interactomes are quantifiable by binding constants and complexomes by concentrations of complexes. By generalization of the law of mass action, the quantitative interactome of an organism defined in that manner contains, in principle, the information that relates expressed proteomes to complexomes, in different cellular or sub-cellular states of that organism. Under this view, the HuRI that measures interactions between protein pairs with yeast two hybrid (Luck *et al.*, 2020) is a qualitative full-length protein interactome. Both BioPlex, an AP-MS resource (Huttlin *et al.*, 2021), and PCP-SILAM, a co-fractionation resource (Skinnider *et al.*, 2021), are qualitative protein complexomes. Finally, ProfAff (present work) is a quantitative fragmental interactome.

Proteomic perturbation upon expression of a viral PBM-containing oncoprotein

We chose HPV16 E6 as a model to investigate how the PBM of a viral oncoprotein can alter the molecular physiology of host cells. E6 is built upon a core folded region composed of two zinc-binding repeats followed by a short intrinsically unfolded C-terminus bearing its PBM (Zanier *et al.*, 2012) (Zanier *et al.*, 2013) (Suarez & Trave, 2018). HPV16 E6 has numerous partners beyond PDZ-proteins (White *et al.*, 2012) (Vande Pol & Klingelhutz, 2013). In particular, E6 forms a trimeric complex with the "LxxLL" motif of ubiquitin-ligase E6AP (Zanier *et al.*, 2013) and the core domain of tumor suppressor p53 (Martinez-Zapien *et al.*, 2016), leading to ubiquitinylation and proteasome-mediated degradation of both proteins (Scheffner *et al.*, 1993) (Martinez-Zapien *et al.*, 2016) in a PBM-independent manner (Delury *et al.*, 2013).

We measured the proteomes of HEK293T cell-lines stably expressing wild-type HPV16 E6 or HPV16 E6ΔPBM, a mutant devoid of the C-terminal PBM. As compared to a control vector, both E6 and E6ΔPBM induced a proteome-wide perturbation affecting ~10% of 2,273 human proteins detected in all three conditions, demonstrating a specific PBM-independent impact of E6 core region (Figure 6A). Among these proteins, a common group of 104 proteins (~5% of the detected proteome) showed significant changes of abundance in the same direction, with a Pearson correlation coefficient (PCC) of 0.85. This group seems to be involved in biological processes

such as viral carcinogenesis or translational control as calculated by GO-enrichment analysis (Table S4). In particular, the abundances of both E6AP and p53 strongly decreased in the presence of E6 or E6ΔPBM, with comparable efficiency (Figure 6A, D and E).

Nonetheless, we also found extensive PBM-dependent effects. Between E6- and E6ΔPBM-expressing cells, 191 proteins (8% of the observed proteome) displayed a significant change in abundance, that exceeded a two-fold ratio for 69 proteins (3% of observed proteome) (Figure 6B). 50 (72%) of those 69 markedly altered proteins varied in opposite directions in HPV16 E6 or E6ΔPBM cells compared to the control cell line (Figure 6C). For example, the concentration of the E3 ubiquitin-protein ligase HUWE1 is decreased by 3-fold by E6 yet increased by 2-fold by E6ΔPBM (Table S4). In the remaining 28% of cases, both E6 and E6ΔPBM caused a concentration variation in the same direction, with the PBM significantly boosting or softening the effectiveness of the E6 core region.

HPV E6 may target some PDZ-proteins to degradation (James & Roberts, 2016). In our data, PDZ-proteins do not particularly stand out among proteins differentially affected by the E6 PBM. Among 11 PDZ-proteins detected in our proteomic survey, no significant change was observed between E6- and E6ΔPBM-transformed cells (Figure 6D). Since the remaining ~140 known human PDZ-proteins were undetected by the MS approach, we used western blot to further investigate the amounts of five of the main E6-binding PDZ proteins (SCRIB, MAGI1, DLG1, SNX27, NHERF3) (Figure 6E and S4A). As compared to control cells, the only significant change was observed for SCRIB levels, which decreased by ~30% in full-length E6-expressing cells, while upon expression of E6ΔPBM SCRIB concentration was unchanged. We also quantified the mRNA levels of several PDZ-proteins in our stable cell lines with RT-qPCR, without observing any significant changes (Figure S4B). Therefore, the PBM-dependent decrease of SCRIB levels in cells expressing full-length E6 likely reflects a shorter protein lifetime. These variations of SCRIB levels still remain quite moderate when compared to the striking decrease in p53 and E6AP levels induced by the core region of E6 independently of its C-terminal PBM (Figure 6D, E).

Discussion

Here we systematically measured 65,000 affinities in a focused region of the human interactome, which was only sparsely covered by pre-existing full-interactome databases. The human proteome is estimated to contain 35×10^3 domain instances dispersed across 10^4 distinct families and about one million motifs, including post-translational variants (Tompa et al., 2014). The full fragmental interactome, once subdivided in its multiple domain-motif families, may thus represent hundreds of millions of potential interactions. This exceeds by 10^2 to 10^3-fold the numbers of interactions documented in Biogrid (~500,000), HuRI (~56,000), BioPlex (~120,000) and PCP-SILAM (~125,000) (Oughtred et al., 2021) (Skinnider et al., 2021) (Luck et al., 2020) (Huttlin et al., 2021). Thus, the human interactome still strongly lacks coverage due to sparse

sampling. The approach presented herein is undoubtfully suitable to fill this gap. We covered more than 6% of the 10^6 potential affinities of one of the largest domain-motif interactomes in human. This was achieved at a pace of up to 10,000 affinity measurements per day, using mainly a benchtop protocol and very limited material resources. The full quantitative human fragmental interactome is thus within realistic reach of a collective research initiative that would systematize this approach.

Exhaustive assessment of all potential interactions might be considered superfluous, on the basis that only the strongest interactions would be functionally relevant. Yet, while strong interactions may stand out by their capacity to disrupt the system when they are missing or altered, homeostatic functioning requires the collective participation of all interactions, from the few strongest ones to the numerous weaker ones. Even the negatome (here defined by the weakest interactions, for which the pK_d value was below the quantification threshold) is expected to play a critical role (Blohm *et al.*, 2014) (Zhang *et al.*, 2018) (Mei & Zhang, 2019). Hence the full interactomic space must be explored.

From a system perspective, fragmental interactomes are at least as relevant as full-length protein interactomes. Full-length proteins can exist as many diverse conformational and/or chemical proteoforms (Aebersold *et al.*, 2018), arising from RNA splicing, post-translational modifications, or binding to third-party proteins or other ligands. All those proteoforms may have distinct binding properties (Poverennaya *et al.*, 2020), but full-length interactomics only capture the binding properties of particular proteoforms, or of undefined mixtures of proteoforms. By contrast, a well-defined minimal interacting fragment pair usually displays one single structure and one single affinity constant. Once determined, these minimal block features open perspectives for bottom-up modeling of the structures and binding properties of multiple proteoforms and of their interactomic combinations, specially if we take into account recent progresses in proteome-wide structural modeling (Tunyasuvunakool *et al.*, 2021).

Quantitative fragmentomics provide access to a yet unexplored multidimensional interactomic space, where affinity-based distances constitute the ultimate way, superior to sequence-based prediction, to situate and compare molecular actors of a biological system. This turns out to be particularly useful for analyzing viral hijacking. We identified an interactomic host spot for oncoviral interference, targeted by both HPV E6 and HTLV TAX1 viral oncoproteins despite their distinctive epithelial and lymphoid tropisms. This hot spot comprises a core of ubiquitously expressed PDZ-proteins, and PBM-proteins that vary more across cell types, as supported by AP-MS experiments performed on different cell models. These experiments led us to the conceptual distinction between intrinsic interactomes and extrinsic complexomes, by showing that amounts of detected complexes vary depending on the proteomic content of cells and are governed by affinities, as already hinted in our previous work (Gogl *et al.*, 2021).

Finally, we investigated an interactome-function relationship by demonstrating that the presence of the PBM of HPV oncoprotein E6 significantly impacted 8% of the detected proteome in cells that stably expressed E6. While, as expected, E6 induced a dramatic PBM-independent

decrease of p53 and E6AP levels, it induced a moderate PBM-dependent decrease of only one detected PDZ-protein, SCRIB. This points to distinct mechanisms of E6 action over p53 and E6AP on the one hand, and PDZ proteins on the other hand. For p53 and E6AP, E6 employs a reductionist approach by plugging them to the ubiquitination system (Scheffner *et al.*, 1993). Since both p53 and E6AP are transcription factors (Catoe & Nawaz, 2011) (Li *et al.*, 2012), and E6AP a E3 ubiquitin ligase with genome-wide impact (Lopez *et al.*, 2019), their degradation expectedly induces system-wide transcriptomic and proteomic changes. With PDZ proteins, E6 rather employs a holistic approach by perturbing an intricated network of promiscuous transient interactions, which also ends up provoking system-wide proteomic changes. This mixed reductionist and holistic approach seems inspiring, if we aim to understand cellular life as well as papillomaviruses do.

Limitations of the study

In our approach, all PDZ domains are fused to the highly soluble bacterial Maltose Binding Protein (MBP) and PBMs are linked to a N-terminal biotin moiety. We have previously used MBP to facilitate and standardize the production of many recombinant proteins and domains in a folded, soluble and active form (Nominé *et al.*, 2001) (Vincentelli *et al.*, 2015) (Zanier *et al.*, 2013) (Gogl *et al.*, 2021), and also extensively used biotinylated peptides for binding assays (Vincentelli *et al.*, 2015) (Bonhoure *et al.*, 2020) (Jané *et al.*, 2020). We never found that the presence or absence of MBP or biotin tags significantly alter the structure nor the binding properties of protein-peptide pairs. However, it is statistically possible, yet unlikely that the affinities of some particular PDZ-PBM pairs addressed in our screen have been altered by the inclusion of MBP and/or biotin tags.

Author contribution

Acknowledgements

References

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst. D*. **66**, 213–221.

Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., Costello, C. E., Cravatt, B. F., Fenselau, C., Garcia, B. A., Ge, Y., Gunawardena, J., Hendrickson, R. C., Hergenrother, P. J., Huber, C. G., Ivanov, A. R., Jensen, O. N., Jewett, M. C., Kelleher, N. L., Kiessling, L. L., Krogan, N. J., Larsen, M. R., Loo, J. A., Ogorzalek Loo, R. R., Lundberg, E., Maccoss, M. J., Mallick, P., Mootha, V. K., Mrksich, M., Muir, T. W., Patrie, S. M., Pesavento, J. J., Pitteri, S. J., Rodriguez, H., Saghatelian, A., Sandoval, W., Schlüter, H., Sechi, S., Slavoff, S. A., Smith, L. M., Snyder, M. P., Thomas, P. M., Uhlén, M., Van Eyk, J. E., Vidal, M., Walt, D. R., White, F. M., Williams, E. R., Wohlschlager, T., Wysocki, V. H., Yates, N. A., Young, N. L. & Zhang, B. (2018). *Nat. Chem. Biol.* **14**, 206–214.

Al-Saleem, J., Dirksen, W. P., Martinez, M. P., Shkriabai, N., Kvaratskhelia, M., Ratner, L. & Green, P. L. (2019). *PLoS One*. **14**, 1–19.

Alvarez-Ponce, D. (2017). *Brief. Bioinform.* **18**, 1017–1020.

Amacher, J. F., Brooks, L., Hampton, T. H. & Madden, D. R. (2020). *J. Struct. Biol. X.* **4**, 100022.

Banks, L., Pim, D. & Thomas, M. (2012). *Nat. Rev. Cancer.* **12**, 877–886.

Blohm, P., Frishman, G., Smialowski, P., Goebels, F., Wachinger, B., Ruepp, A. & Frishman, D. (2014). *Nucleic Acids Res.* **42**, 396–400.

Bohl, J., Brimer, N., Lyons, C. & Vande Pol, S. B. (2007). *J. Biol. Chem.* **282**, 9392–9400.

Bonhoure, A., Forster, A., Babah, K. O., Gógl, G., Eberling, P., Kostmann, C., Volkmer, R., Tapia Mancilla, V., Travé, G. & Nominé, Y. (2020). *Anal. Biochem.* **603**, 113772.

Caillet-Saguy, C., Durbesson, F., Rezelj, V. V., Gogl, G., Tran, Q. D., Twizere, J. C., Vignuzzi, M., Vincentelli, R. & Wolff, N. (2021). *FEBS J.* 1–15.

Catoe, H. W. & Nawaz, Z. (2011). *Steroids.* **76**, 897–902.

Charbonnier, S., Zanier, K., Masson, M. & Travé, G. (2006). *Protein Expr. Purif.* **50**, 89–101.

Davey, N. E., Travé, G. & Gibson, T. J. (2011). *Trends Biochem. Sci.* **36**, 159–169.

Delury, C. P., Marsh, E. K., James, C. D., Boon, S. S., Banks, L., Knight, G. L. & Roberts, S. (2013). *J. Virol.* **87**, 9463–9472.

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N. P., Travé, G. & Gibson, T. J. (2008). *Front. Biosci.* **13**, 6580–6603.

Duhoo, Y., Girault, V., Turchetto, J., Ramond, L., Durbesson, F., Fourquet, P., Nominé, Y., Cardoso, V., Sequeira, A. F., Brás, J. L. A., Fontes, C. M. G. A., Travé, G., Wolff, N. & Vincentelli, R. (2019). *Methods Mol. Biol.* **vol 2025**, pp 439-476.

Duhoo, Y., Sequeira, A. F., Saez, N. J., Turchetto, J., Ramond, L., Peysson, F., Bras, J. L. A., Gilles, N., Darbon, H., Fontes, C. M. G. A. & Vince (2019). *Methods Mol. Biol.* **2025**, 165–190.

Ecsédi, P., Gógl, G., Hóf, H., Kiss, B., Harmat, V. & Nyitray, L. (2020). *Structure.* **28**, 943-953.e4.

Etienne-Manneville, S. (2008). *Oncogene.* **27**, 6970–6980.

Futsch, N., Mahieux, R. & Dutartre, H. (2018). *Viruses.* **10**, 1–25.

Genera, M., Quioc-Salomon, B., Nourisson, A., Colcombet-Cazenave, B., Haouz, A., Mechaly, A., Matondo, M., Duchateau, M., König, A., Windisch, M. P., Neuveut, C., Wolff, N. & Caillet-Saguy, C. (2021). *Sci. Rep.* **11**, 1–17.

Giraud, E., del Val, C. O., Caillet-Saguy, C., Zehrouni, N., Khou, C., Caillet, J., Jacob, Y., Pardigon, N. & Wolff, N. (2021). *Sci. Rep.* **11**, 1–13.

Gógl, G., Biri-Kovács, B., Durbesson, F., Jane, P., Nomine, Y., Kostmann, C., Bilics, V., Simon, M., Reményi, A., Vincentelli, R., Trave, G. & Nyitray, L. (2019). *J. Mol. Biol.* **431**, 1234–1249.

Gógl, G., Biri-Kovács, B., Póti, Á. L., Vadászi, H., Szeder, B., Bodor, A., Schlosser, G., Ács, A., Turiák, L., Buday, L., Alexa, A., Nyitray, L. & Reményi, A. (2018). *FEBS J.* **285**, 46–71.

Gogl, G., Jane, P., Caillet-Saguy, C., Kostmann, C., Bich, G., Cousido-Siah, A., Nyitray, L., Vincentelli, R., Wolff, N., Nomine, Y., Sluchanko, N. N. & Trave, G. (2020). *Structure*. **28**, 747-759.e3.

Gogl, G., Tugaeva, K., Eberling, P., Kostmann, C., Trave, G. & Sluchanko, N. N. (2021). *Nat. Commun.* **12**,.

Harington, K. J. & Syrigos, K. N. (2000). *Ann. Surg. Oncol.* **7**, 783–788.

Howarth, M., Takao, K., Hayashi, Y. & Ting, A. Y. (2005). *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7583–7588.

Huttlin, E. L., Bruckner, R. J., Navarrete-Perea, J., Cannon, J. R., Baltier, K., Gebreab, F., Gygi, M. P., Thornock, A., Zarraga, G., Tam, S., Szpyt, J., Gassaway, B. M., Panov, A., Parzen, H., Fu, S., Golbazi, A., Maenpaa, E., Stricker, K., Guha Thakurta, S., Zhang, T., Rad, R., Pan, J., Nusinow, D. P., Paulo, J. A., Schweppe, D. K., Vaites, L. P., Harper, J. W. & Gygi, S. P. (2021). *Cell*. **184**, 3022-3040.e28.

Ivarsson, Y. & Jemth, P. (2019). *Curr. Opin. Struct. Biol.* **54**, 26–33.

Jadwin, J. A., Ogiue-Ikeda, M. & MacHida, K. (2012). *FEBS Lett.* **586**, 2586–2596.

James, C. D. & Roberts, S. (2016). *Pahogens*. **5**, 1–22.

Jané, P., Chiron, L., Bich, G., Travé, G. & Nominé, Y. (2021). *Methods in Molecular Biology*, Vol. *2256*, pp. 61–74.

Jané, P., Gógl, G., Kostmann, C., Bich, G., Girault, V., Caillet-Saguy, C., Eberling, P., Vincentelli, R., Wolff, N., Travé, G. & Nominé, Y. (2020). *PLoS One*. 1–24.

Javier, R. T. & Rice, A. P. (2011). *J. Virol.* **85**, 11544–11556.

Jimenez-Guardeño, J. M., Nieto-Torres, J. L., DeDiego, M. L., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodriguez, C. & Enjuanes, L. (2014). *PLoS Pathog.* **10**,.

Kabsch, W. (2010). *Acta Cryst. D*. **66**, 125–132.

Keilhauer, E. C., Hein, M. Y. & Mann, M. (2015). *Mol. Cell. Proteomics*. **14**, 120–135.

Kelil, A., Levy, E. D. & Michnick, S. W. (2016). *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3862–E3871.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R. C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S. & Hermjakob, H. (2012). *Nucleic Acids Res.* **40**, 841–846.

Kitov, P. I. & Bundle, D. R. (2003). *J. Am. Chem. Soc.* **125**, 16271–16284.

Kolacsek, O., Krízsik, V., Schamberger, A., Erdei, Z., Apáti, Á., Várady, G., Mátés, L., Izsvák, Z., Ivics, Z., Sarkadi, B. & Orbán, T. I. (2013). *Mob. DNA.* **4**, 1–8.

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B. & Gibson, T. J. (2019). *Nucleic Acids Res.* **48**, 1–11.

Li, M., He, Y., Feng, X. & Huang, J. (2012). *Biochim. Biophys. Acta - Gene Regul. Mech.* **1819**, 684–687.

Lopez, S. J., Segal, D. J. & LaSalle, J. M. (2019). *Front. Mol. Neurosci.* **11**, 1–8.

Luck, K., Charbonnier, S. & Travé, G. (2012). *FEBS Lett.* **586**, 2648–2661.

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., Choi, D., Coté, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., Knapp, J. J., Kovács, I. A., Lemmens, I., Mee, M. W., Mellor, J. C., Pollis, C., Pons, C., Richardson, A. D., Schlabach, S., Teeking, B., Yadav, A., Babor, M., Balcha, D., Basha, O., Bowman-Colin, C., Chin, S. F., Choi, S. G., Colabella, C., Coppin, G., D'Amata, C., De Ridder, D., De Rouck, S., Duran-Frigola, M., Ennajdaoui, H., Goebels, F., Goehring, L., Gopal, A., Haddad, G., Hatchi, E., Helmy, M., Jacob, Y., Kassa, Y., Landini, S., Li, R., van Lieshout, N., MacWilliams, A., Markey, D., Paulson, J. N., Rangarajan, S., Rasla, J., Rayhan, A., Rolland, T., San-Miguel, A., Shen, Y., Sheykhkarimli, D., Sheynkman, G. M., Simonovsky, E., Taşan, M., Tejeda, A., Tropepe, V., Twizere, J. C., Wang, Y., Weatheritt, R. J., Weile, J., Xia, Y., Yang, X., Yeger-Lotem, E., Zhong, Q., Aloy, P., Bader, G. D., De Las Rivas, J., Gaudet, S., Hao, T., Rak, J., Tavernier, J., Hill, D. E., Vidal, M., Roth, F. P. & Calderwood, M. A. (2020). *Nature.* **580**, 402–408.

Luck, K., Sheynkman, G. M., Zhang, I. & Vidal, M. (2017). *Trends Biochem. Sci.* **42**, 342–354.

Manel, N., Battini, J. L., Taylor, N. & Sitbon, M. (2005). *Oncogene.* **24**, 6016–6025.

Martinez-Zapien, D., Ruiz, F. X., Poirson, J., Mitschler, A., Ramirez, J., Forster, A., Cousido-Siah, A., Masson, M., Pol, S. Vande, Podjarny, A., Travé, G. & Zanier, K. (2016). *Nature.* **529**, 541–545.

16

Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., Vandendriessche, T., Ivics, Z. & Izsvák, Z. (2009). *Nat. Genet.* **41**, 753–761.

McBride, A. A. (2017). *Philos. Trans. R. Soc. B Biol. Sci.* **372**,.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

Mei, S. & Zhang, K. (2019). *Int. J. Mol. Sci.* **20**,.

Nominé, Y., Ristriani, T., Laurent, C., Lefevre, J. F., Weiss, É. & Travé, G. (2001). *Protein Eng.* **14**, 297–305.

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., Dolma, S., Coulombe-Huntington, J., Chatr-aryamontri, A., Dolinski, K. & Tyers, M. (2021). *Protein Sci.* **30**, 187–200.

Pawson, T. & Nash, P. (2003). *Science (80-.).* **300**, 445–452.

Vande Pol, S. B. & Klingelhutz, A. J. (2013). *Virology.* **445**, 115–137.

Poverennaya, E. V., Kiseleva, O. I., Ivanov, A. S. & Ponomarenko, E. A. (2020). *Biochem.* **85**, 68–79.

Roehrl, M. H. A., Wang, J. Y. & Wagner, G. (2004). *Biochemistry.* **43**, 16056–16066.

Sámano-Sánchez, H. & Gibson, T. J. (2020). *Trends Biochem. Sci.* **45**, 526–544.

Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. (1993). *Cell.* **75**, 495–505.

Schneider, T. D. & Stephens, R. M. (1990). *Nucleic Acids Res.* **18**, 6097–6100.

Seo, M. H. & Kim, P. M. (2018). *Curr. Opin. Struct. Biol.* **50**, 162–170.

Simon, M., Gógl, G., Ecsédi, P., Póti, Á., Kardos, J. & Nyitray, L. (2020). *FEBS J.* 1–13.

Skinnider, M. A. & Foster, L. J. (2021). *Nat. Methods.* **18**, 806–815.

Skinnider, M. A., Scott, N. E., Prudova, A., Kerr, C. H., Stoynov, N., Stacey, R. G., Chan, Q. W. T., Rattray, D., Gsponer, J. & Foster, L. J. (2021). *Cell.* **184**, 4073-4089.e17.

Sokal, R. R. & Michener, C. D. (1958). A statistical method for evaluating systematic relationships.

Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A. & MacBeath, G. (2007). *Science (80-.).* **317**, 364–370.

Strickland, S. W., Brimer, N., Lyons, C. & Vande Pol, S. B. (2018). *Virology.* **516**, 127–138.

Suarez, I. & Trave, G. (2018). *Viruses*. **10**, 1–22.

Thomas, M., Myers, M. P., Massimi, P., Guarnaccia, C. & Banks, L. (2016). *PLoS Pathog.* **12**, 1–21.

Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. (2014). *Mol. Cell.* **55**, 161–169.

Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D. & Sidhu, S. S. (2008). *PLoS Biol.* **6**, 2043–2059.

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., Potapenko, A., Ballard, A. J., Romera-Paredes, B., Nikolov, S., Jain, R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli, P., Jumper, J. & Hassabis, D. (2021). *Nature*. **596**, 590–596.

Valdes, J. L., Tang, J., McDermott, M. I., Kuo, J. C., Zimmerman, S. P., Wincovitch, S. M., Waterman, C. M., Milgram, S. L. & Playford, M. P. (2011). *J. Biol. Chem.* **286**, 39403–39416.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., Forster, A., Cassonnet, P., Borg, J.-P., Jacob, Y., Masson, M., Nominé, Y., Reboul, J., Wolff, N., Charbonnier, S. & Travé, G. (2015). *Nat. Methods*. **12**, 787–793.

White, E. A., Kramer, R. E., Tan, M. J. A., Hayes, S. D., Harper, J. W. & Howley, P. M. (2012). *J. Virol.* **86**, 13174–13186.

Zámbó, B., Mózner, O., Bartos, Z., Török, G., Várady, G., Telbisz, Á., Homolya, L., Orbán, T. I. & Sarkadi, B. (2020). *Cell. Mol. Life Sci.* **77**, 365–378.

Zanier, K., Charbonnier, S., Sidi, A. O. M. hame. O., McEwen, A. G., Ferrario, M. G., Poussin-Courmontagne, P., Cura, V., Brimer, N., Babah, K. O., Ansari, T., Muller, I., Stote, R. H., Cavarelli, J., Vande Pol, S. & Travé, G. (2013). *Science (80-.).* **339**, 694–698.

Zanier, K., M'Hamed Ould Sidi, A. O., Boulade-Ladame, C., Rybin, V., Chappelle, A., Atkinson, A., Kieffer, B. & Travé, G. (2012). *Structure*. **20**, 604–617.

Zhang, L., Yu, G., Guo, M. & Wang, J. (2018). *BMC Bioinformatics*. **19**,.

Figure Legends

Figure 1,

Affinity profiling of a subsection of the PDZ-PBM interactome: principle and data overview.

The human PDZ-PBM interactome represents $\sim 10^6$ potential interactions (left panel). Focusing on a viral- and cancer-relevant region (upper part of left panel, zoomed in middle panel), we measured the 59,578 affinities of 133 human PDZs for 424 human and 24 viral PBMs (fully measured subsection, left half of middle panel) and 5,573 extra affinities involving the remaining 133 PDZs and a subset of 45 PBMs (right half of middle panel). Measured or non-measured PDZ-PBM pairs are respectively colored in heat map mode (see color scale in lower right panel) or in grey. Any horizontal or vertical cross-section of that interactome represents an individual PDZome- or PBMome- binding profile, as illustrated for NET1 PBM and SCRIB_1 PDZ (top right corner). The horizontal cross-section of the middle panel is indicated by a horizontal arrow in the corner of the PDZome binding profile and the vertical cross-section of the middle panel is indicated by a vertical arrow in the corner of the PBMome binding profile. Affinity-based Euclidian distances computed from the fully measured subsection reveal a clade of human PBMs displaying the highest interactomic similarities with HTLV1 TAX1 and HPV E6 oncoviral PBMs (middle right panel). Note the general similarity of the heatmap patterns of the oncoviral PBMs and of the identified human PBM clade, and the fine pattern differences further revealed by sub-clustering using the UGPMA approach. For more details, see Table S1.

Figure 2,

Validation and distribution of the quantified PDZ-PBM affinities .

(A) Comparison of PDZ-PBM dissociation constants quantified with holdup assay (HU) or competitive fluorescence polarization (FP). The Pearson correlation coefficient (PCC) was determined based on affinities quantifiable by both methods (n = 255). (B) Comparison of PDZ-PBM dissociation constants quantified with holdup assay or obtained from literature. PCC was determined based on interactions quantified by both methods (n = 362). (C) PDZ-PBM dissociation constants quantified from HU follow an exponential-like distribution. The grey zone indicates the range of pK_d quantification thresholds that slightly varies between different assays. The red line indicates an exponential probability distribution function, predicted from the data. (D) PDZ-PBM dissociation constants obtained from literature follow a normal distribution. The red line indicates the predicted normal distribution. Note the different affinity range in panels C and D. For more details, see Table S1.

Figure 3,

Molecular determinants behind binding specificities in the PDZ-PBM interactome.

(A) PDZome-binding profile of HPV35 E6 PBM. (B) The affinity-weighted specificity logo of SNTB1 PDZ matches well with HPV35 E6 PBM at positions p0, p-2, p-3 and p-4, which display favorable contacts in the crystal structure of the SNTB1/HPV35 E6 complex. (C) MAGI1-2 PDZ logo matches with HPV35 E6 PBM at the same positions, but also at p-5 (Arg), exposed to the acidic β2-β3 loop of MAGI1_2 in the MAGI1/HPV35 E6 complex. Accordingly, HPV35 E6 binds stronger to MAGI1-2 than to SNTB1. (D) PBMome-binding profile of SYNJ2BP PDZ. (E) RPS6KA1 PBM matches poorly to the logo of SYNJ2BP, in line with RPS6KA1 binding only weakly to SYNJ2BP. The p-3 phosphorylated RPS6KA1 PBM matches better to SYNJ2BP logo thanks to the acidic phospho-Ser at p-3, which contacts two residues of β2 and β3 strands of SYNJ2BP in the complex. Accordingly, p-3 phosphorylated RPS6KA1 shows increased binding to SYNJ2BP. (F) HTLV TAX1 PBM sequence matches strongly with SYNJ2BP logo at p0, p-1, p-2, p-3 and p-4, in line with TAX1 being the strongest SYNJ2BP binder in our interactome. In the SYNJ2BP-TAX1 complex, Glu at p-3 of TAX1 engages similar contacts as pSer of RPS6KA1_-3P, while Arg at p-4 provides additional contact to the α2 helix of SYNJ2BP. (G) Superposition of TAX1 and p-3 phosphorylated RPS6KA1 PBMs bound to SYNJ2BP. The carboxylate binding loop of SYNJ2BP is shifted in the phosphorylated RPS6KA1-bound complex, most likely related to unfavorable substitution of Val by Leu at p0 of RPS6KA1 PBM. (H) The logo of SNX27 indicates preference for class 1 PBMs (Ser/Thr residue at p-2). MERS E has a class 3 PBM (Glu at p-2) with poor match to SNX27 logo; yet MERS E binds relatively well to SNX27. In the SNX27/MERS E complex Trp at p-1 of MERS E establishes favorable hydrophobic contacts. (I) Interactomic distance profile of MERS E in the explored PBM space. The closest neighbour of MERS E is ARVCF, a class 1 PBM with Trp at p-1. For further details see Figure S2 and Table S1,2. All logos were taken from the ProfAff server.

Figure 4,

Exploring interaction networks prone to perturbation by oncoviral PBMs in three host cell models.

(A) Identification by PBM-AP-MS experiments of HPV E6 PBM-binding and HTLV1 TAX1 PBM-binding PDZ-containing proteins endogenously expressed in Hela, HaCat or Jurkat cell lines. Results of comparable studies by Strickland et al. (2018) (*) Thomas et al. (2016) (**) and Al-Saleem et al. (2019) (***) are also provided. For each PDZ-protein identified, the number of PDZ domains and the presence or absence of PDZ-oligomerization L27 domains are indicated at the bottom of the table. The 4 PDZ-proteins selected for PDZ-AP-MS experiments are highlighted in grey. (B) Identification by PDZ-AP-MS experiments of endogenously expressed PBM-containing protein partners of 6 PDZ domains isolated from the 4 PDZ proteins selected in (A). The ten proteins on top of the list contain PBMs belonging to the oncoviral-like human PBM clade revealed by our quantitative fragmental interactome. Interaction partners in (A) and (B) were found by comparing the prey quantities compared to a control resin. Binding threshold was defined at > 2-fold enrichment and < 0.01 P-value, calculated by two-tailed unpaired T-test. Blue

cells indicate detectable interaction between the immobilized bait (minimal binding fragment) and the endogenous prey (full-length protein). Note that only interaction partners containing PDZ domains or C-terminal PBMs are indicated. For a comprehensive list of all identified interaction partners, see Table S3.

Figure 5,

Amounts of complex formation between endogenous preys and bait fragments in AP-MS correlate with their corresponding affinities.

(A) Enrichment-affinity correlations. Upper left plot: fold-enrichment values of PBM-containing protein ATP2B4 captured by the 6 different PDZ baits in an HeLa lysate, plotted versus pK_d values of the corresponding ATP2B4 PBM-PDZ complexes. A strong correlation is observed, as evidenced by the high Pearson correlation coefficient (PCC). Upper right plot: fold-enrichment values of PDZ-containing protein SYNJ2BP captured by 7 different PBM baits in a Jurkat lysate, plotted versus pK_d values of the corresponding SYNJ2BP PDZ-PBM complexes. Middle histogram: enrichment/affinity PCC values, obtained as in the instances shown above, for all endogenous PBM-containing prey proteins captured from the indicated lysates by PDZ baits, for which at least 4 corresponding PDZ-PBM pK_d values were available in ProfAff. Average PCC is 0.8. Lower histogram: enrichment/affinity PCC values of all endogenous PDZ-containing prey proteins captured from the indicated lysates by PBM baits, for which at least 4 corresponding PDZ-PBM pK_d values were available in ProfAff. Here, PBM-binding affinities of proteins possessing a single PDZ domain were directly obtained from the fragmental interactome, and those of proteins possessing multiple PDZ domains were estimated using the additive model (see methods). Note that correlations are overall weaker than in the other histogram, with an average PCC of 0.5. Encapsulated legend shows bar color codes for both histograms. (B) Enrichment-enrichment correlations across different lysates. Upper left plot: fold-enrichment values of PBM-containing protein ATP2B4 captured by the 6 different PDZ baits in Jurkat versus HaCat lysates. A strong correlation is observed, as evidenced by the high PCC. Upper right plot: fold-enrichment values of PDZ-containing protein SYNJ2BP captured by 7 different PBM baits in HaCat versus HeLa lysates. Middle histogram: enrichment/enrichment PCC values, for all endogenous PBM-containing proteins that were detectably captured by PDZ baits in at least two cell lysates. Lower histogram: enrichment/enrichment PCC values, for all endogenous PBM-containing proteins that were detectably captured by PDZ baits in at least two cell lysates. Note that correlations are overall weaker than in the other histogram. Encapsulated legend shows bar color codes for both histograms. See Table S3 for more details.

Figure 6,

Proteomic impact of stable HPV16 E6 expression in HEK293T cells and the contribution of its PBM.

(A) Proteomic changes in HEK293T upon stable expression of HPV16 E6 or HPV16 E6ΔPBM compared to a control cell line, "IRES", that only contains the empty vector. Dots indicate the detected 2,273 proteins. Black dots indicate proteins whose abundances changed significantly in both E6- and E6ΔPBM-expressing cells as compared to control cells. The dotted line indicates the diagonal. (B) Proteomic difference between the cell lines expressing HPV16 E6, or HPV16 E6ΔPBM. The dotted line indicates the statistical significance threshold (P < 0.05, two-tailed unpaired T-test) and dark dots indicate proteins whose statistical difference is higher than this threshold. (C) Proteins that show statistically significant differences in concentration between E6- and E6ΔPBM-expressing cell lines often change in opposite direction compared to the control cell line. (D) Amounts of 11 PDZ proteins and p53 in the three cell lines, quantified by mass spectrometry. Relative XIC intensities (n = 3) are shown compared to the IRES control. (E) Amounts of 5 PDZ-proteins, as well as p53 and E6AP, quantified by western blot. Intensities, normalized to GAPDH (n = 3) are compared to the IRES control. Statistical significance calculated by two-tailed unpaired T-test is indicated by * for P < 0.05 and ** for P < 0.01. See Figure S3 for results of western blots and Table S4 for full proteomic results.

List of supplementary tables

Table 1, Affinity survey and benchmarking of the PDZ-PBM interactome

Table 2, Crystallographic table of the crystallized PDZ-PBM complexes

Table 3, Results of the AP-MS experiments and their analysis

Table 4, Results of proteomic characterization 293T cells stably expressing HPV16 E6, HPV16 E6ΔPBM, or the empty vector (IRES).

Supplementary Figure Legends

Figure S1,

Comparing the quantitative PDZ-PBM affinities measured in our work to qualitative interactions from the literature.

(A) Number of PDZ-PBM interactions detected by our screen, that would have been retained by other methods with different sensitivity thresholds. In our dataset, we have used, by default, the lowest possible pK_d quantification thresholds (situated in the grey zone on the plot) derived from the binding detection thresholds of the source holdup data. This way, we quantified the dissociation constants of 18,332 out of 65,157 measured interactions. The curve describes the number of these interactions that would have been retained using less sensitive methods, as a function of the minimal detection threshold $pK_{d\ min}$ of those methods. For instance, a method with $pK_{d\ min} = 5$ would only detect the ~1500 interactions displaying a $K_d < 10$ μM. (B) Coverage by our data of previously known protein-protein interactions, as a function of sensitivity threshold. For the interactomic region explored by our dataset, we plotted the percentage of qualitative protein-protein interactions (PPI) documented in large scale interactomic resources (n=1,248) which found a match to PDZ-PBM pK_d values quantified in our measurements, in the function of a variable minimal detection threshold $pK_{d\ min}$, defined as before. For proteins with multiple PDZ domains, we assumed an "additive" effect between the distinct binding sites (see Methods). If we account for all the interactions detected and quantified in our assay, our data confirm nearly 60% of the interactions (direct or indirect) reported in large scale resources. If we consider that our measurements could have been performed using less sensitive methods, this percentage further decreases, as a function of the $pK_{d\ min}$ assumed for these methods. (C) Affinity distribution histogram of all the protein-protein interactions found in large scale interactomic resources, for which we could match PDZ-PBM pK_d values quantified in our measurements. The affinities used for the distribution are taken from our dataset. For interactions involving PDZ-proteins with multiple PDZ-domains, "additive" affinities were considered. Histograms were calculated using all available affinities using a bin size of 0.1 pK_d. (D) Affinity distribution histogram of all the potential human protein-protein interactions that may exist according to our domain-motif data. The affinities used for the distribution are obtained from our PDZ-PBM affinity dataset. For interactions involving PDZ-proteins with multiple PDZ-domains, "additive" affinities were considered.

Figure S2,

Finding interactomic neighborhoods in the explored Euclidean affinity space.

(A) Closest neighbors of a few selected motifs. Eucledian distance profiles are shown for HPV16 E6, HPV18 E6, HTLV1 TAX1, NET1, and MERS E. On the right panel, a zoomed in view is shown of the closest 25 motifs. Note that the unexplored part of the PBMome may contain other

motifs with more similar binding properties to these references. (B) According to their quantified affinities with 133 common PDZ domains, PBMs were clustered based on their Euclidean distances using an UPGMA algorithm (see Methods). The identified cladogram is shown along with the sequences of motifs. Sequences are colored according to their conventional classification (PBMs of class 1, 2 and 3 in green, black and red, respectively; internal or atypical PBMs in purple). The clades enriched in oncoviral PBMs is highlighted in bold red. Sequence based frequency logos are shown for all major clusters.

Figure S3,

Validation of 293T cell lines stably expressing HPV16 E6 and HPV16 E6ΔPBM.

(A) Analysis of some PDZ protein expressions in these cells by Western-blot. (B) Top: RT-qPCR analysis of the RNA expression of different proteins in these cells (mean ± SD, n = 3). Bottom: Primers used for the RT-qPCR analyses. (C) Top: Relative copy numbers (mean ± SD, n = 5) of 293T cell lines stably expressing HPV16 E6 and HPV16 E6ΔPBM. Bottom: Primers used for copy number determination.

Material and methods

Synthesis of purified biotinylated and fluorescent peptides used for holdup, fluorescence polarization and AP-MS experiments

HPLC-purified (all >95% purity) biotinylated peptides were chemically synthesized on an ABI 443A synthesizer with standard Fmoc strategy with biotin group attached to the N-terminus via a TTDS (Trioxatridecan-succinamic acid) linker, or were commercially purchased from Genicbio (Shanghai, China) with biotin group attached to the N-terminus via an Ahx (6-aminohexanoic acid) linker. Predicted peptide masses were confirmed by mass-spectrometry. Peptide concentrations were determined based on their dry weight.

Fluorescent peptides f16E6: fluorescein-RTRRETQL; fRSK1: fluorescein-KLPSTTL and fpRSK1: fluorescein-KLPpSTTL were chemically synthesized on an ABI 443A synthesizer and HPLC purified with fluorescein coupled directly to the N-terminus. The biotinylated MERS-E peptide was FITC labeled (fMERS: biotin-Ahx-DSK(-fluorescein)PPLPPDEWV) as follows: the peptide was mixed with sub-stoichiometric FITC (Sigma-Aldrich, St. Louis, MO, USA) in a basic HEPES buffer (pH 8.2). Labeling reaction was stopped with 100 mM TRIS and the reaction mixture was buffer exchanged in order to remove fluorescent contaminants. The concentrations of fluorescent peptides and labeled fraction of fMERS were determined based on their dry weight and their fluorescence intensity. For competitive fluorescence polarization assay, only HPLC purified peptides were used.

Reagents for crude peptide library synthesis

All solvents and chemicals were used without any further purification and purchased as peptide grade when available. N-methylpyrrolidinone (NMP) was purchased from Biosolve. Diisopropylcarbodiimide and collidine was purchased from Sigma. The following L-amino acids were used: Fmoc-Ala, Fmoc-Arg(pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Asp(Otbu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Glu(Otbu)-OH, Fmoc-Glu-Otbu, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Ile-OH, Boc-Ile-OH, Fmoc-Leu-OH, Fmoc-Val-OH, Fmoc-Lys(boc)-OH, Fmoc-Ser(Otbu)-OH, Fmoc-Thr(Otbu)-OH, Fmoc-Tyr(Otbu)-OH, FmocTrp(Boc)-OH, Fmoc-Phe-OH, and they were purchased from either Protein Technologies, Novabiochem Merck or IRIS Biotech (Germany).

Biotin was purchased from Sigma. Fmoc-preloaded resin was used and purchased from Merck Millipore. Fmoc-8-amino-3,6-dioxaoctanoic acid (Fmoc-Ado-OH) was purchased from Flamma Group, Italy.

Crude peptide library synthesis used for holdup experiments

Crude peptide libraries were synthesized by parallel 96 format peptide synthesis using Intavis multipep RSi. Filter plates from NUNC (Thermo-Fischer) were used loaded with 10 mg per well of Fmoc-preloaded polystyrene resins (Merck Millipore). Fmoc-amino acids (Fmoc-AA) as 0.3 M in dimethylformamide (DMF) containing 0.3 M OxymaPure were used. Three consecutive couplings using 100 μl Fmoc-AA and 11 μl DIC (3 M in DMF) were employed with coupling times of 5, 15 and 60 min. Removal of Fmoc was done by washing twice with 120 μl of 25% piperidin in DMF for 2 and 8 min. Washing was done by adding 150 μl to each well by the 8-pin manifold 5 times. Coupling of biotin was done as a 0.3 M solution in DMSO containing 0.3 M OxymaPure and activated by DIC. Coupling time was extended to 15 min, 45 min and 360 min. TFA cleavage was done by 93% TFA containing 4% triisopropylsilane (TIPS) 1% thioanisol 3% H2O for 2 h. The total volume of 1 ml TFA added was added to each well. TFA was reduced in volume to approximately 150 μl followed by precipitation by the addition of diethyl ether. The peptides were transferred to Waters solvinert plates and washed thoroughly five times with diethyl ether. After washing with diethyl ether a small fraction of the peptide slurry was added to a microtiter plate and dried followed by UPLC-MS analysis after solubilisation in DMF/H2O (1:1). Analysis of purity and identity by MS was performed using Waters Acquity UPLC system, with Waters Acquity TUV detector 214 nm connected to a LCT Premier XE mass spectrometer from Micromass. The buffer used were 0.05% TFA in H2O (buffer A) and 0.05% TFA in acetonitrile (buffer B) using a gradient of 5-60% B for 3.5 min on a Waters column BEH C18 1.7 mm; 2.1 mm x 50 mm and a flow rate of 0.45 ml/min.

All peptides had a biotin group attached to the N-terminus via an Ado-Ado linker. Internal standards used for fluorescent holdup normalization (see below) were re-synthetized as crude peptides. No apparent differences in these internal standards were observed between HPLC purified and crude peptides in holdup experiments. Predicted peptide masses were confirmed by mass-spectrometry in all cases and the purity was found to observably reach > 90% in most cases.

Average peptide concentrations were determined based on the excess dry weight of the entire 96-well plate and we used 10-50x molar excess (based on dry weight) in holdup experiments to take into account the variability of peptide-to-peptide yields.

Preparation of PDZ library in bacterial lysates for holdup experiments (by LYSC)

The PDZome v2 library, consisting of all 266 human PDZ domains as His_6-MBP-PDZ constructs, was prepared as previously described (Duhoo, Girault *et al.*, 2019). The library was expressed in *E. coli* BL21(DE3) with an autoinduction media and was lysed as described elsewhere in details (Duhoo, Girault *et al.*, 2019). The His_6-MBP-PDZ concentrations of soluble cell lysate fractions were evaluated with a microfluidic capillary electrophoretic system (Caliper LabChip GXII, PerkinElmer, Waltham, Massachusetts) and were adjusted to 4 µM by dilutions. Lysozyme and BSA internal standards were added to the library before freezing the library in 96 well plate for the holdup experiment.

Single affinity purified protein expression and purification for holdup experiments (by SAPF)

For single affinity purification, the PDZome v2 library was expressed in the same conditions (strain, media, buffers) than for the bacterial lysate (Duhoo, Girault *et al.*, 2019) but at a bigger culture scale (24 ml per PDZ instead of 6ml per PDZ). The production (24 ml) and Ni-affinity purification (800 μl of beads/PDZ) follows strictly the protocol described in (Duhoo, Sequeira *et al.*, 2019). The elution volume of Ni-affinity purification was reduced from 4 ml (Duhoo, Sequeira *et al.*, 2019) to 2.5 ml/PDZ. 96 PDZ are produced in parallel and are purified in four blocks. At the end of the purifications of these blocks purified PDZs are desalted using Zeba Spin plates (Thermo scientific) in 50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM imidazole. The purity and concentration of each His_6-MBP-PDZ was evaluated with a microfluidic capillary electrophoretic system (Caliper LabChip GXII, PerkinElmer, Waltham, Massachusetts) and adjusted to 4 μM by dilutions before freezing the library in 96-well plate for the holdup experiment.

Double affinity purified protein expression and purification for holdup experiments (by DAPF) and fluorescence polarization assays

For double affinity purification, His_6-MBP-PDZ constructs from the PDZome v2 library (Duhoo, Girault *et al.*, 2019) were expressed in E. coli BL21(DE3) with IPTG induction (1 mM IPTG at 25 ºC for 4-5 hours) and harvested cells were lysed in a buffer containing 50 mM TRIS pH 7.5, 150-300 mM NaCl, 2 mM BME, cOmplete EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland), 1% Triton X-100, and trace amount of DNAse, RNAse, and Lysozyme. Lysates were frozen at -20 ºC before further purification steps. Lysates were sonicated and centrifuged for clarification. Expressed PDZ-domains were captured on pre-packed Ni-IDA (Protino Ni-IDA Resin, Macherey-Nagel, Duren, Germany) columns, were washed with at least 10 column volume cold wash buffer (50 mM TRIS pH 7.5, 150 mM NaCl, 2 mM BME) before elution with 250 mM imidazole. The Ni-elution was collected directly on a pre-equilibrated amylose column (amylose high flow resin, New England Biolabs, Ipswich, Massachusetts). Amylose column was washed with 5 column volume cold wash buffer before fractionated elution in a buffer containing 25 mM Hepes pH 7.5, 150 mM NaCl, 1 mM TCEP,

10% glycerol, 5 mM maltose, cOmplete EDTA-free protease inhibitor cocktail. The concentration of proteins was determined by their UV absorption at 280 nm before aliquots were snap freeze in LN2 and storage at -80 ºC.

Holdup assay

Principle of the holdup assay and previously published versions

The holdup assay is a comparative retention-based chromatographic approach devoid of washing steps, that attempts to monitor the steady-state binding equilibrium by separating the interaction partners on different, moving *vs* stationary, phases (Charbonnier *et al.*, 2006). Streptavidin (GE Healthcare, Uppsala, Sweden) resin is saturated with either biotin ("control resin"), or a biotinylated peptide. Then, a PDZ-containing solution is added to both resins and the reaction mixture is agitated for 15 min at room temperature. The whole experiment is carried out on filter plates (Millipore, Burlington, Massachusetts) and the incubation step is stopped with rapid filtration to separate the unbound PDZ fraction. Finally, the concentration (a.k.a. intensities) of His$_6$-MBP-PDZ is determined and binding intensities (BI) are calculated using eq. 1:

$$BI = \frac{I_{total} - I_{depleted}}{I_{total}} \tag{1}$$

where I_{total} is the total intensity (e.g. concentration, peak intensity, fluorescent signal, etc.) of the PDZ present in the flow-through of the biotin-saturated control resin, and $I_{depleted}$ is the intensity of PDZ in the flow-through of the peptide-saturated resin. In the holdup buffer (50 mM TRIS pH 7.5, 300 mM NaCl, 1 mM TCEP) at least a single internal standard was used (BSA/lysozyme or fluorescein/mCherry) for peak intensity normalization.

Previously, the holdup approach was automatized on liquid handling systems (Vincentelli *et al.*, 2015). In this method, a total bacterial lysate over-expressing the His$_6$-MBP-PDZ protein was used, and the His$_6$-MBP-PDZ intensities were measured using capillary electrophoresis. While the ease of use of such a complex matrix is desirable (ease of library preparation, crowded environment, etc.), the readout is both slow and requires a multi step data-curation, as described in details in (Jané *et al.*, 2020) (Jané *et al.*, 2021). To eliminate this bottleneck, we also developed an intrinsic fluorescence based holdup method, where the concentration measurement is done based on the fluorescence of Trp residues of purified His$_6$-MBP-PDZ constructs (Bonhoure *et al.*,

2020). The fluorescent readout is fast and can be done on any fluorescent plate-reader, however at a cost of accuracy. Indeed, any contaminant present in the holdup mixture, such as contaminants from bacterial source, or spontaneously cleaved His_6-MBP, will generate a background fluorescence, that will decrease the observable partial activity of each PDZ-sample. The real partial activity of each sample can be retrieved through calibration with some BI values obtained previously by capillary electrophoresis-based holdup.

New developments increase the throughput and the precision of the holdup method

Automated holdup experiments were originally designed in a way where 96 different PDZ domains were tested against 4 wells on a 384 well filter plate, 3 different/identical peptides and a single control. In a reverse holdup layout, multiple peptides are tested against the same protein. For example, 370 peptides and 14 biotin controls can be placed on a single 384 well filter plate and then tested against a single PDZ. This way, not only the number of tested peptides increase, but also the number of controls, too. In previous holdup experiments, the singlicate control for each PDZ caused an under-determined situation where the precision of the determined BI value was mainly determined by the accuracy of the single control measurement. For example, if an interaction was reportedly measured in triplicates, 3 peptide saturated wells were compared with a single control. In the reverse holdup layout, many control wells are measured on the same PDZ and therefore the total intensity can be determined at high precision and great accuracy. For example, if an interaction was reportedly measured in singlicate, each peptide saturated well was compared with 10-15 controls.

As a minor tweak, we also prepared presaturated resin stocks for the reverse layout holdup experiments, instead of saturating resins directly on the filter plate. This allowed us to achieve a higher reproducibility by minimizing the well-to-well variability of resin immobilization rates and to prepare multiple plates simultaneously. Stocks were prepared by simply up-scaling the described procedure for a single well (see below). Stocks were stored in holdup buffer and the biotin depletion step (where the nearly peptide/biotin saturated resins are incubated with large excess of biotin to deplete the remaining vacancies in the resin) was only executed after transferring the desired amounts to filter plates.

Due to the extremely slow dissociation rate of the biotin-streptavidin interaction, it is possible to recycle the holdup-treated resins in the filter plate several times (at least up to 20-25 times

from our personal experience, the holdup measurements being highly reproducible at least up to this number). Recycling was carried out in 5 steps. The plates were washed first with 10 resin volume 1 M NaCl, then with 10 resin volume 2 M Urea and finally, 3 times with 10 resin volume holdup buffer. All steps in the holdup protocol where the liquid fraction needs to be filtered can be performed either on a vacuum manifold, or by centrifugation with the exception of the last washing step before adding the PDZ-solution and the holdup reaction itself, which both need to be performed by centrifugation.

By combining the fluorescent readout, the reverse holdup layout, and the resin recycling, not only the throughput of the holdup assay increased by several orders of magnitude, but also its precision and accuracy. In addition, this optimized protocol does not require any specialized instrument and can be implemented with a simple plate reader, a centrifuge, and a multichannel pipette. Optimally, with the help of a vacuum manifold, an electric multichannel pipette (such as an E1-ClipTip pipette from Thermo Fisher, Waltham , Massachusetts), and a liquid dispensing station (such as a Flexdrop IV from Perkin Elmer, Waltham, Massachusetts), the holdup assay is capable to measure the affinities of 5-10,000 interactions a day at a fraction of the cost of the originally developed automatic setup.

Interaction measurements with holdup experiments

In the present work, we include data from 3 types of holdup measurements: holdup measurement with bacterial lysates coupled with capillary electrophoretic readout (LYSC), holdup measurement with high-throughput single affinity purified PDZ protein coupled with fluorescent readout (SAPF), and holdup measurement with double affinity purified PDZ protein coupled with fluorescent readout (DAPF).

Most LYSC data originate from previous articles (Vincentelli *et al.*, 2015) (Gógl *et al.*, 2019) (Gogl *et al.*, 2020) (Jané *et al.*, 2020) (Genera *et al.*, 2021) (Giraud *et al.*, 2021) (Caillet-Saguy *et al.*, 2021). Additional LYSC holdup experiments were carried out as previously described, using only the original layout. Briefly, 2.5 µl streptavidin resin was mixed with biotin or biotinylated peptide at 55 µM concentration in 6.5 resin volume for 15 min to achieve resin saturation, then after a single washing step (10 resin volume, holdup buffer), the resin was depleted with biotin (1 mM biotin, 5 resin volume, 15 min). Holdup experiments were carried out after 3 washing steps (10 resin volume, holdup buffer). 5 µl bacterial lysate with 4 µM His_6-MBP-PDZ supplemented

with 4 μM BSA and 0.05 mg/ml lysozyme was incubated on the filter plate for 15 min with shaking before filtration by centrifugation. Filtrates were analyzed with Caliper Labchip GXII (Perkin Elmer, Waltham, Massachusetts), and collected data was analyzed with in-house developed algorithms (Jané *et al.*, 2020). In total, our database includes 4,475 unique interactions characterized by CALIP method, originating from 9,920 measurements. We have found that the standard deviation of duplicate experiments (n = 1,893) was σ = 0.05 BI. We set our binding detection threshold based on this and defined it at 2 σ (BI = 0.1). Based on this threshold, 807 unique interactions showed detectable binding in CALIP holdup experiments.

SAPF experiments were carried out in the original layout. Resin preparation was identical as in LYSC holdup experiments. In each well 2.5 μl peptide- or biotin-saturated resin was incubated with 5 μl single affinity-purified His$_6$-MBP-PDZ at a concentration of 4 μM. His$_6$-MBP-PDZ was supplemented with 50 nM fluorescein and 100 nM mCherry. Incubation was carried out on filter plates for 15 min with shaking before filtration. Filtrates were analyzed on a PHERAstar (BMG Labtech, Offenburg, Germany) microplate reader by using 485 ± 10 nm - 528 ± 10 nm (fluorescein), 575 ± 10 nm - 620 ± 10 nm (mCherry), and 295 ± 10 nm - 350 ± 10 nm (Trp-fluorescence) band-pass filters. In total, our database includes 6,850 unique interactions characterized by SAPF method, all measured in singlicate. Due to the lack of replicates, we set the same binding detection threshold that we defined for LYSC experiments (BI ≥ 0.1). Based on this threshold, 1101 unique interactions showed detectable binding in SAPF holdup experiments.

DAPF experiments were carried in the reverse layout with presaturated resins. Resin presaturation was performed with the same volume/concentration ratio as in the resin preparation for the LYSC and SAPF experiments, but over a longer period of incubation (2-4 h). The presaturated resin slurry was supplemented with 20% ethanol for longer-term storage. After distributing 5 μl presaturated streptavidin resin on filter plates, after a single washing step (10 resin volume, holdup buffer), the resin was depleted with biotin (1 mM biotin, 5 resin volume, 10 min). After two washing steps, the resin was incubated with 10 μl double affinity-purified His$_6$-MBP-PDZ solutions, which was supplemented with 50 nM fluorescein and 100 nM mCherry. Fluorescence intensity measurements were performed as described in SAPF experiments. The reverse layout made it possible to include several peptides on the filter plate that was measured previously with Caliper-based holdup methods. In case the reverse layout was used, measured BI values were compared to LYSC standards to determine the partial activity of each PDZ sample.

Then, the determined partial activity was applied in the calculations to normalize the obtained BI values. In total, our database includes 59,179 unique interactions characterized by DAPF method, originating from 62,604 measurements. We have found that the standard deviation of duplicate experiments (n = 2,780) had an average of 0.029 BI (with a few outliers) and a median of 0.019 BI, respectively. Based on these, we used $\sigma = 0.025$ to define our binding detection threshold at 2 σ (BI = 0.05). Based on this threshold, 17,726 unique interactions showed detectable binding in DAPF holdup experiments.

Competitive fluorescence polarization (FP) assay

In direct fluorescence polarization assay, the fluorophore label can interfere with affinity determination and can cause large experimental bias. For this reason, affinities were determined with competitive FP measurements where the bound fluorescent peptide is chased out by increasing concentration of unlabeled competitor. In the fitting procedure, affinities of both peptides are considered. Each PDZ-domain was first tested with labeled peptides in direct experiments and then optimal labeled peptides were used in competitive experiments.

Double affinity-purified His_6-MBP-PDZ were used for affinity measurements by FP. Fluorescence polarization was measured with a PHERAstar microplate reader by using 485 ± 20 nm and 528 ± 20 nm band-pass filters for excitation and emission, respectively. In direct FP measurements, a dilution series of the MBP-PDZ was prepared in 96 well plates (96 well skirted PCR plate, 4ti-0740, 4titude, Wotton, UK) in a 20 mM HEPES pH 7.5 buffer that contained 150 mM NaCl, 0.5 mM TCEP, 0.01% Tween 20 and 50 nM fluorescently-labeled peptide. The volume of the dilution series was 40 µl, which was later divided into three technical replicates of 10 µl during transferring to 384 well micro-plates (low binding microplate, 384 well, E18063G5, Greiner Bio-One, Kremsmünster, Austria). In total, the polarization of the probe was measured at 8 different protein concentrations (whereas one contained no protein and corresponded to the free peptide). In competitive FP measurements, the same buffer was supplemented with the protein to achieve a complex formation of 60-80%, based on the direct titration. Then, this mixture was used for preparing a dilution series of the competitor (i.e. the studied peptides) and the measurement was carried out identically as in the direct experiment. Analysis of FP experiments were carried out using ProFit, an in-house developed, Python-based fitting program (Simon *et al.*,

2020). The dissociation constants of the direct and competitive FP experiments were obtained by fitting the measured direct data with a quadratic binding equation first and by fitting the measured competitive data with a competitive equation, using several obtained parameters from the first fit (Roehrl *et al.*, 2004). In part, competitive fluorescence polarization data was taken from (Jané *et al.*, 2020) and (Gogl *et al.*, 2020). In total, 395 unique interactions were characterized by competitive FP in our database.

Conversion of holdup binding intensities to dissociation constants

In the holdup measurements, we measure steady-state binding intensities that can be converted to steady-state dissociation constants using eq. 2:

$$K_d = \frac{([PDZ] - BI * [PDZ]) * ([PBM] - BI * [PDZ])}{BI * [PDZ]} \qquad (2)$$

where [PDZ] is the total PDZ concentration (set to 4 μM in usual cases in our assays) and [PBM] is the total peptide concentration. This later parameter is unknown, which makes a direct and accurate conversion impossible. To find the missing peptide concentration, we use orthogonal affinity measurements (Gogl *et al.*, 2020). We determine the dissociation constants by competitive FP of a set of interactions that are also measured by holdup. Then, we substitute the measured BI and K_d values into eq. 2 and calculate the corresponding [PBM] concentration. Finally, we take the median of all calculated concentrations (3-10 concentrations, depending on peptide) for a given peptide to determine an average parameter that can be used for the conversion of all measured BI values of the same peptide. We have found that in most cases, the peptide concentration lied between 10-40 μM with an average peptide concentration of 18 μM. Only ~39 peptides were characterized by FP, out of the >450 measured by holdup. For peptides that were not characterized by FP, we used this average peptide concentration for conversion.

When we perform the conversion from BI to K_d, the binding detection threshold of the holdup measurements dictates the threshold for the quantification of the K_d values in the affinity scale. As a consequence of the peptide concentration and BI detection threshold variability, PDZome-binding profiles of each PBM have a slightly different K_d quantification threshold in affinity scale. In the least sensitive experiment, the K_d quantification threshold was found to be only 30

µM and it was at 790 µM in the most sensitive experiment. On average, the quantification threshold was around $K_d = 320$ µM.

In our database, converted affinities are shown as pK_d values, where pK_d is defined as the negative 10-base logarithm of the dissociation constant. Therefore, pK_d values are closely related to the change in Gibbs free energy upon binding. Due to the different detection thresholds of the different holdup experiments, as well as the different peptide concentrations in each assay, we do not merge together the affinities determined by different holdup methods. Instead, we provide each determined affinities separately for all measured holdup assays, as well as a "composite affinity" for each interaction, which is taken from the most reliable holdup method.

The uncertainty of determined pKd values can be estimated in at least two different ways, either using the measured, or propagated uncertainty of BI values, or by measuring the pKd variability between the holdup and the FP experiments. Standard deviation of BI values is provided in our database in case more than one experiment was performed. In case of singlicate experiments, one can propagate the errors that we observed in the same kind of experiments when duplicates were measured ($\sigma = 0.05$ for LYSC/SAPF and $\sigma = 0.025$ for DAPF experiments). This standard deviation of BI values can be used to propagate the error into the determined pKd value. However, an accurate error propagation model needs to include both the variability of the BI values and the peptide concentration and the latter is difficult to assess based on current data. Alternatively, it is possible to do error estimation based on pKd values determined by FP. Note that we use the FP dataset for the determination of the peptide concentration, therefore the two datasets are globally centered. However, each peptide concentration is determined from multiple holdup-FP experiment pairs and therefore the variance of individual pairs can be still measured. The difference between the orthogonal pKd values derived from FP and holdup measurements follows a normal distribution with a standard deviation of 0.227 pKd. If we assume that neither of these methods perform systematically better than the other we can use this standard deviation as a global metrics of the uncertainty of determined pKd values in our database. For example, one can also use it to estimate the confidence of an affinity difference using the following empirical rule: >0.227 pKd difference indicates ~68.3% confidence (1 σ), >0.454 pKd difference indicates ~95.5% confidence (2 σ), and >0.681 pKd difference indicates a 99.7% confidence (3 σ). However, this does not mean that only >0.454 pKd differences should be considered as reliable in our database. One needs to assess

the uncertainty of each studied measurement in several ways (e.g. by comparing different holdup measurements, or the ways of affinity conversion) in order to make objective judgment on data quality.

Profaff database

To aid the access to our affinity measurements, we established a PHP- and MySQL-based online database that is accessible at the https://profaff.igbmc.science/ address. On the database, one can browse the affinity profiles of PDZ-domains and PBMs by searching based on gene ID, motif sequence or consensus motif. It is possible to access all additional information of each interaction, such as measured BI values, details of normalization and conversion, or even the competitive titrations of FP measurements. In addition, for each PDZ-domains where we detected binders above a certain (adjustable) binding threshold, we provide affinity weighted frequency plots, based on the sequences of all detected binders, where the weights are calculated based on eq. 3:

$$weight = 10^{pK_d - binding\,threshold} \qquad\qquad (3)$$

Finally, the ProfAff database is built so that it can include any holdup measurements done on other types of domain-motif interaction. For example, in the actual version it also includes measured affinities from some of our previous works: of 14-3-3 proteins with a few phosphopeptides (phosphorylated PBMs) (Gogl *et al.*, 2021) and of HPV E6 proteins with host LxxLL motifs (Bonhoure, A. et al, manuscript in preparation). All these sub-databases are interconnected therefore, it is possible to see whether a PBM is also targeted by 14-3-3 domains, or if a protein also contains an LxxLL site. The code of the database is uploaded to GitHub (at https://github.com/GoglG/ProfAff). We hope that the ProfAff database will serve as a proof of concept database for the community and will set a new trend for future interactomic databases.

Quantitative Reference Interactome

A non-exhaustive collection of previously published PDZ-PBM affinities used as an external benchmark was generated as follows. We started from the database published by Amacher et al

(Amacher *et al.*, 2020), which makes an inventory of different PDZ-PBM affinities published up to 2007. To complete this list, articles were collected from PubMed using the keywords "PDZ[Title] AND interaction" (April 2021). If an affinity measurement was performed, the nature of the PDZ and the PBM interacting along with their Uniprot codes, the method used for measuring the interaction, the eventual modification to the PBM peptide necessary for the measurement (i.e. fluorophore, biotin, etc), and the measured K_d along with the standard deviation of the measurement (when available), were collected. The same information was collected for the affinities reviewed in Amacher et al. The following data were excluded during collection or from the data reviewed in Amacher et al.: (i) data concerning full-length proteins or PDZ domains in tandem, as opposed to isolated PDZ domains; (ii) data concerning PBM variants (mutations or post-traductional modifications) or artificial PBM sequences (i.e. generated by phage-display); (iii) data for which the length or the sequence of the PBM was not available. The resulting database comprises 1547 K_d measurements, extracted from 132 publications, and measured between 142 different PDZ domains, some of them of different origins (human, mouse, rat, and/or vinegar fly), and 280 different PBM motifs from all classes, including 12 internal PBM. For a complete list of this reference interactome, as well as references for each interaction, see Table S1.

For interaction matching we used only affinities from human, mouse, and rat PDZ-domains. PDZ-domain boundaries were not assessed during data collection. In addition, the database contains PBMs with lengths between 4 and >100 residues, but only affinities from 8-12mer peptides were used for interaction matching. Interaction matching was performed based on corresponding 10mer PBM sequences (i.e. exact, extended, or truncated) and PDZ-domain names. In total, we have found 362 matching affinities between our measurements and this filtered benchmark dataset, where the affinities showed a PCC of 0.59 (see Figure 2B).

Estimating PPI affinities and matching with a Qualitative Reference Interactome

We used the previously generated domain-motif interactome, consisting of the composite affinities (i.e. taken from the most reliable holdup method; see above) to estimate affinities of PPIs. We assumed that a multidomain PDZ-protein can only interact with a single PBM at the same time (eq. 4):

$$PDZ + PBM \rightleftharpoons (PDZ - PBM)_{site\,1}, \text{or} (PDZ - PBM)_{site\,2}, \text{or} (PDZ - PBM)_{siten} \qquad (4)$$
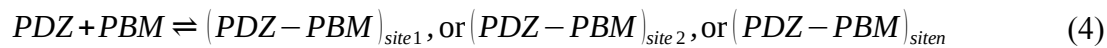
Using this approximation, as proposed by others before (Kitov & Bundle, 2003), the association constant of a multi-site interaction is the sum of the association constants of all sites (eq. 5):

$$K_{a,additive} = \frac{[PDZ - PBM]_{site\,1} + [PDZ - PBM]_{site\,2} + \ldots + [PDZ - PBM]_{siten}}{[PDZ] \times [PBM]} = \sum K_{a,site} \qquad (5)$$

This calculation does not imply neither positive, nor negative cooperation between interaction sites. In addition, we do not take into account the formation of higher order, multivalent complexes that may form, especially at higher concentrations. The contribution of unmeasurable weak interactions was also neglected in the calculations. Although for these reasons, our calculation is inherently imperfect, it readily provides a lower limit for the affinity of the "full-length complexes" since most of these neglected effects would in principle increase the global affinity. Most importantly, this simple calculation can be done without any *a priori* knowledge about the synergistic nature of each multi-domain-proteins.

To collect previously observed qualitatively reported binding events between PDZ and PBM proteins, we analyzed the Biogrid (4.4.199 version), Intact (2020_10_01 version), and Bioplex (Dec_2019 version, both 293T and HCT116 datasets) databases (Oughtred *et al.*, 2021) (Kerrien *et al.*, 2012) (Huttlin *et al.*, 2021). We performed the search within the interactomic space that we experimentally studied using the holdup method. To find instances of reported interactions, we looked for co-occurrence of the IDs of PDZ- and PBM-proteins within single entries. In case the PBM originated from a PDZ-protein, this algorithm identifies it as an observed interaction. Thus, any potential self-binding was manually checked in all databases. In total, we have identified 1,233, 629, 152 and 56 observed interactions from the Biogrid, Intact, Bioplex (293T) and Bioplex (HCT116) databases. In total, our qualitative interactome includes 1,654 unique interactions observed between the 150 PDZ-proteins and the 448 PBM-proteins. Out of these,

1248 interactions were tested in our assays and we could match quantifiable affinities to 725 interactions.

Hierarchical clustering of the human PDZ-PBM interactome

To cluster the PDZ-PBM domain-motif interactome based on their interactomic properties, we used unweighted pair-group method with arithmetic average (UPGMA) clustering based on the Euclidean distance in pK_d scale. Note that pK_d values are closely related to binding energies (eq. 6):

$$pK_d = -\log_{10} K_d = \frac{-\Delta G}{2.303 \times RT} \qquad (6)$$

Note that the Euclidean distance of one affinity in pK_d scale equals to $\Delta\Delta G/2.303*RT$, therefore the Euclidean distance of two binding profile equals to the sum of all $\Delta\Delta G/2.303*RT$ values. Clustering of both PBMs and PDZ domains were performed based on a near-complete part of the interactome consisting of composite affinities, covering 448 PBMs and 133 PDZ-domains. First, we thresholded our interactome at 3.5 pK_d, thus removing the very small number of interactions that we managed to quantified at an even weaker threshold. Then, we replaced these and all the unmeasureable weak affinities with 3.5 pK_d. Within this interactomic space, the affinities of a few unmeasured interactions (n < 5) were estimated based on a kNN (k-Nearest Neighbours) approach. First, a PBM clustering was performed by omitting the affected PDZ-domains. Based on this, the 2 nearest neighbours in the interactomic space was determined and the missing affinity was replaced by the average of their affinities with the same domain.

Crystallization

For crystallization, PDZ domains were cloned with an N-terminal TEV protease cleavable His$_6$ tag and a C-terminal ANXA2 (Annexin A2) tag (Gógl *et al.*, 2018) (Ecsédi *et al.*, 2020). His$_6$-PDZ-ANXA2 constructs were produced in E. coli BL21(DE3) with IPTG induction (1 mM IPTG at 18 ºC for overnight expression) and harvested cells were lysed in a buffer containing 50 mM TRIS pH 7.5, 150-300 mM NaCl, 2 mM BME, cOmplete EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland), 1% Triton X-100, and trace amount of DNAse, RNAse, and

Lysozyme. Lysates were frozen at -20 ºC before further purification steps. Lysates were sonicated and centrifuged for clarification. Expressed PDZ-domains were captured on pre-packed Ni-IDA (Protino Ni-IDA Resin, Macherey-Nagel, Duren, Germany) columns, were washed with at least 10 column volume cold wash buffer (50 mM TRIS pH 7.5, 150 mM NaCl, 2 mM BME) before elution with 250 mM imidazole. The Ni-elution was cleaved with TEV protease and the PDZ-ANXA2 was purified by cation exchange on a HiTrap SP HP column (GE Healthcare, Chicago, Illinois). Proteins were aliquoted at 5-7 mg/ml concentration before flash freezing in LN2. Samples were supplemented with 5-6 molar excess of selected PBM peptides and 2 mM CaCl$_2$ before crystallization.

Crystallization conditions were screened using commercially available (Qiagen, Hampton Research, Emerald Biosystems) and in-house developed kits by the sitting-drop vapor-diffusion method in 96-well MRC 2-drop plates (SWISSCI, Neuheim, Switzerland), using a Mosquito robot (TTP Labtech, Cambridge, UK). Crystals of MAGI1_2 in complex with the PBM of HPV35 E6 grew rapidly in a drop made from 5 μl of protein solution and 5 μl of reservoir solution containing 20-25% polyethylene glycol 3000, 100 mM sodium citrate buffered at pH 5.5 and 100 mM trisodium-citrate at 20°C. Crystals of SNTB1 in complex with the PBM of HPV35 E6 grew rapidly in a drop made from 2 μl of protein solution and 2 μl of reservoir solution containing 20% polyethylene glycol 3350, 200 mM sodium malonate buffered at pH 7.0 at 20°C. Crystals of SYNJ2BP in complex with RPS6KA1_-3P were grown in a drop containing 100 mM ammonium sulfate, 100 mM sodium formate, 25% PEG smear broad (Molecular Dimensions, Sheffield, UK), and 100 mM HEPES pH 7.5 at 20°C. Crystals of SYNJ2BP in complex with HTLV1 TAX1 were grown in the F10 condition (120 mM Monosaccharides, 100 mM Buffer System 3 pH 8.5, 50% Precipitant Mix 2) of the MORPHEUS screen at 20°C. Crystals of SNX27 in complex with the PBM of MERS-E grew rapidly in a drop made from 2 μl of protein solution and 2 μl of reservoir solution containing 10% polyethylene glycol 8000, 100 mM imidazole buffered at pH 8.0 at 20 °C. All crystals were flash-cooled in a cryoprotectant solution containing 25% glycerol and stored in liquid nitrogen.

X-ray diffraction data were collected at the Synchrotron Swiss Light Source (SLS) (Switzerland) on the X06DA (PXIII) beamline or at SOLEIL (France) on the PX2-A beamline. All data were processed with the program XDS (Kabsch, 2010) and the phase problem was solved by molecular replacement (McCoy *et al.*, 2007), based on the previously determined

crystal structures of MAGI1_2-ANXA2 complex (PDB ID 5N7D) and the structures of the PDZ domains of SNTB2, SYNJ2BP, and SNX27 (2VRF, 2JIK, 6SAK) using Phaser and structure refinement was carried out with PHENIX (Adams *et al.*, 2010). TLS refinement was applied during the refinement. The crystallographic parameters and the statistics of data collection and refinement are shown in Table S3. The refined model and the structure factor amplitudes have been deposited in the PDB with the accession codes 7P70, 7P71, 7P72, 7P73, and 7P74.

Cell cultures and lysate preparation for proteomics analyses

HeLa cells were grown in DMEM (1g/L glucose, Gibco) medium completed with 5% FCS and 40 µg/mL gentamicin, diluted every 3rd/4th day 1/10. HaCat cells were grown in DMEM (1g/L glucose, Gibco) medium completed with 10% FCS and 40 µg/mL gentamicin, diluted every 3rd/4th day 1/4. Jurkat cells were grown in RPMI (Gibco) medium completed with 10% FCS and 40 µg/mL gentamicin, diluted every 3rd/4th day 1/12. All cells were kept at 37°C and 5% CO2. For proteomics analyses cells were seeded on T-175 flasks. After they reached confluency, adherent cells were washed with PBS once and collected by scraping with ice-cold lysis buffer (Hepes-KOH pH 7.5 50 mM, NaCl 150 mM, Triton X-100 1%, cOmplete EDTA-free protease inhibitor cocktail 1x, EDTA 2 mM, TCEP 5 mM, glycerol 10%). Jurkat cells were collected by 1000 g x 5 min centrifugation, washed once with PBS, then collected by 1000 g x 5 min centrifugation again and lysed in ice-cold lysis buffer. Lysates were sonicated 4x20 sec with 1 sec long pulses on ice, then incubated rotating at 4°C for 30 min. Lysates were centrifuged at 12,000 rpm 4°C for 20 min and supernatant was kept for further analysis. Concentration determination was carried out by standard Bradford method (Bio-Rad Protein Assay Dye Reagent #5000006) using BSA (MP BIomedicals #160069) as control on a Bio-Rad SmartSpec 3000 spectrophotometer instrument.

Sample preparation for affinity purification mass spectrometry (AP-MS) experiments

For PDZ-AP-MS experiments, PDZ domains were cloned as His$_6$-AviTag-MBP-PDZ constructs. For co-expression with BirA, *E. coli* BL21(DE3) cells were co-transformed with His$_6$-AviTag-MBP, or His$_6$-AviTag-MBP-PDZ-coding and PET21a-BirA (Addgene #20857) plasmids (Howarth *et al.*, 2005). Simultaneously to IPTG induction, 50 µM biotin was added to the media.

After cell harvesting, the lysis buffer was also supplemented with 50 µM biotin. Otherwise, expression and double affinity purification was identical as previously described in details. Biotinylation efficiency was found to be around 60%.

For PBM-AP-MS experiments, 30 µl streptavidin resin was mixed with biotin or peptide at 50-60 µM concentration in 6-6.5 resin volume for 60 min to achieve resin saturation. For PDZ-AP-MS experiments, 30 µl streptavidin resin was mixed with biotinylated MBP or MBP-PDZ at 40-50 µM concentration in > 1000 x resin volume for 60 min to achieve resin saturation. After saturation, resins were washed a single time (10 resin volume, holdup buffer), and were depleted with biotin (1 mM biotin, 5-10 resin volume, 10 min). Finally, resins were washed 2 times. 0.5 ml 2 mg/ml cell lysate was added to the dry resin (1mg total input) and were incubated at 4 ºC for 2 h. Then, the beads were washed 3 times with 10 resin volume buffer containing: 50 mM TRIS pH 8.5, 150 mM NaCl, 1% TritonX-100, 10x cOmplete EDTA-free protease inhibitor cocktail, 2 mM EDTA, 1 mM TCEP. Then, the beads were washed 2 times with 10 resin volume buffer containing: 50 mM TRIS pH 8.5, 150 mM NaCl, 1 mM TCEP. Finally, captured protein was eluted from the resin in two steps and the eluted fractions were pooled. For each elution the beads were incubated for 30 min with 3 resin volume buffer containing: 20 mM TRIS pH8.5, 100 mM NaCl, 500 µM TCEP, 8 M Urea. Between each step, the beads were separated by mild centrifugation and the supernatant was removed by gentle pipetting.

Reagents for mass spectrometry

Acetonitrile MS grade, formic acid, trifluoroacetic acid, trichloroacetic acid, iodoacetamide, urea, tris(hydroxymethyl)aminomethane (TRIS), 2-carboxyethy-phosphine from Sigma Aldrich (St. Quentin Fallavier, France), trypsin from Promega (Charbonnieres les Bains, France), lysyl endopeptidase from Wako (Richmond, USA)

Sample digestion for mass spectrometry

The samples were precipitated with TCA 20% overnight at 4°C and centrifuged at 14,000 rpm for 10 min at 4°C. The protein pellets were washed twice with 1 mL cold acetone and air dried. The protein extracts were solubilized in urea 8 M, reduced with 5 mM TCEP for 30 min and

alkylated with 10 mM iodoacetamide for 30 min in the dark. Double digestion was performed at 37°C with 500 ng endoproteinase Lys-C for 4 h, followed by 4-fold dilution and an overnight digestion with 500 ng trypsin. Peptide mixtures were then desalted on C18 spin-column and dried on Speed-Vacuum.

LC-MS/MS Analysis

Samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose California) coupled in line with a LTQ-Orbitrap ELITE mass spectrometer via a nano-electrospray ionization source (Thermo Scientific, San Jose California). Peptide mixtures were injected in 0.1% TFA on a C18 Acclaim PepMap100 trap-column (75 µm ID x 2 cm, 3 µm, 100Å, Thermo Fisher Scientific) for 3 min at 5 µL/min with 2% ACN, 0.1% FA in H_2O and then separated on a C18 Accucore nano-column (75 µm ID x 50 cm, 2.6 µm, 150Å, Thermo Fisher Scientific) at 220 nl/min and 38°C with a 90 min linear gradient from 5% to 30% buffer B (A: 0.1% FA in H_2O / B: 99% ACN, 0.1% FA in H_2O), regeneration at 5% B. The mass spectrometer was operated in positive ionization mode, in data-dependent mode with survey scans from m/z 350-1,500 acquired in the Orbitrap at a resolution of 120,000 at m/z 400. The 20 most intense peaks from survey scans were selected for further fragmentation in the Linear Ion Trap with an isolation window of 2.0 Da and were fragmented by CID with normalized collision energy of 35%. (TOP20CID method) Unassigned and single charged states were excluded from fragmentation. The Ion Target Value for the survey scans (in the Orbitrap) and the MS2 mode (in the Linear Ion Trap) were set to 1E6 and 5E3 respectively and the maximum injection time was set to 100 ms for both scan modes. Dynamic exclusion was set to 20 s after one repeat count with mass width at ± 10 ppm.

Mass spectrometry data analysis

Proteins were identified by database searching using SequestHT (Thermo Fisher Scientific) with Proteome Discoverer 2.4 software (PD2.4, Thermo Fisher Scientific) on human FASTA database downloaded from SwissProt (reviewed, release 2020_11_27, 20304 entries, https://www.uniprot.org/). Precursor and fragment mass tolerances were set at 7 ppm and 0.6 Da

respectively, and up to 2 missed cleavages were allowed. Oxidation (M, +15.995 Da) was set as variable modification, and Carbamidomethylation (C, + 57.021 Da) as fixed modification. Peptides and proteins were filtered with a false discovery rate (FDR) at 1%. Label-free quantification was based on the extracted ion chromatography intensity of the peptides. All samples were measured in technical triplicates. The measured extracted ion chromatogram (XIC) intensities were normalized based on median intensities of the entire dataset to correct minor loading differences. For statistical tests and enrichment calculations, not detectable intensity values were treated with an imputation method, where the missing values were replaced by random values similar to the 10% of the lowest intensity values present in the entire dataset. Unpaired two tailed T-test, assuming equal variance, were performed on obtained $\log_2$ XIC intensities. The detection threshold for enriched proteins in AP-MS experiments were analyzed above 2-fold enrichment with <0.01 P value. Then, proteins with PDZ-domains or PBMs were extracted from the list of enriched proteins. In case only a specific isoform of the detected protein contains a PBM, the identified peptide fragments were manually checked to check which isoform(s) were most likely detected. All raw LC-MS/MS data (from 165 runs) have been deposited to the ProteomeXchange via the PRIDE database with identifier PXD027743.

Generation of HEK293T stable cell lines

The p10-IRES2-EGFP vector was generated previously (Zámbó *et al.*, 2020). We inserted the wild-type and C-terminal 10 amino acid-deleted (E6ΔPBM) versions of *HPV16 E6* in MCS between the *IR-DRs* by standard restriction enzyme based cloning. Sleeping Beauty transposase containing pSB100 vector (Addgene #34879) was a kindly gift of Dr. Zsuzsanna Izsvak (MDC Berlin Germany,) and Dr. Tamas I. Orban (RCNS Budapest Hungary) (Mátés *et al.*, 2009).

HEK293T cells were grown in DMEM (1 g/L glucose, Gibco) medium completed with 10% FCS and 1% penicillin-streptomycin, diluted every 3rd/4th day 1/10, and were kept at 37°C and 5% CO$_2$. Stable cells were generated as described previously (Zámbó *et al.*, 2020). Briefly, we co-transfected the cells with 2 µg p10 and 0.2 µg pSB100 (10/1 ratio) vectors on 6-well plates using jetPRIME (Polyplus #114-15) reagent as recommended by manufacturer. 3 days after transfection cells were trypsinized and sorted for EGFP positive cells on a BD FACS Aria III equipment. Two weeks after the first sort, stable cells were selected by sorting again for EGFP

44

positive cells. Relative copy number of transposon casettes integrated into the genome were determined by qPCR (Figure S3C). We do not find significant difference in the copy numbers between E6 and E6ΔPBM expressing cells, however they both contained ~3-times less copy, compared to the control cells. Despite this copy number difference, we do not observe difference in GFP expression in these cells by western blot, which indicates that the expression of GFP from the empty IRES vector is less efficient compared to the vector that also contains E6 or E6ΔPBM before the IRES sequence. Lysates were prepared for proteomics analyses and western blot as described above and below, respectively**.**

gDNA isolation and relative copy number determination:

gDNA was isolated with standard SDS lysis-isopropanol precipitation method. Relative copy numbers were determined by qPCR with SybrGreen reagent (Quiagen QuantiTect #204143) on a Roche Lightcycler 480 II equipment by using previously described method and primers (Kolacsek *et al.*, 2013) recognizing the *GFP* and the *IR-DR* left arm (part of the transposon casette) sequence. The Ct values were normalized to *RPPH1* gene using standard *ΔΔCt* method.

RT-PCR

Cells were plated on 6-well plates. When they reached confluency, RNA was isolated by using Qiagen RNeasy Mini kit (#74104). On-column DNase digestion was done by using Qiagen RNase-Free DNase Set (#79254). 2 μg total RNA per sample was transcribed to cDNA by using SuperScript IV kit (Invitrogen, #18091050) with random hexamer primers (Thermo Fisher #SO142) according to standard protocol. For qPCR, primers were designed with the help of Primer Blast (NCBI NIH, https://www.ncbi.nlm.nih.gov/tools/primer-blast/) choosing primers which recognize all common alternative transcripts of the given gene (Figure S3). qPCR was done by using QuantiTect SyberGreen (Qiagen, #204143) and LightCycler II equipment (Roche). Standard ΔΔCt method was used to determine relative quantity to GAPDH. For further analyses GraphPad Prism 7 software was used.

Western blot

Cells were directly lysed in 4x Laemmli buffer (120 mM Tris-HCl pH 7, 8% SDS, 100 mM DTT, 32% glycerol, 0.004% bromphenol blue, 1% β-mercaptoethanol). Equal amounts of samples were loaded on 8%, 10% or 12% acrylamide-gels. Transfer was done into PVDF membranes using a Trans-Blot Turbo Transfer System and Trans-Blot Turbo RTA Transfer Kit (BioRad, #1704273). After 1 h of blocking in 5% milk, membranes were incubated overnight 4°C in primary antibody in 5% milk. The following antibodies and concentrations were used: anti-GAPDH (Merck #MAB374) 1/5,000, anti-HPV16 E6 (IGBMC polyclonal, recognizing the N-terminus of the protein) 1/1,000, anti-GFP (IGBMC polyclonal) 1/5,000, anti-p53 (CST clone 7F5 #2527) 1/1,000, anti-E6AP (Sigma clone 3E5 #SAB1404508-100UG) 1/1,000, anti-SCRIB (Thermo Fisher #PA5-54821) 1/1,000, anti-MAGI1 (Santa Cruz #sc-100326) 1/1,000, anti-SAP97/DLG1 (Thermo Fisher #PA1-741) 1/1,000, anti-SNX27 (Thermo Fisher #MA5-27854) 1/500, anti- PDZK1/NHERF3 (Santa Cruz #sc-100337) 1/1,000. Membranes were washed three times with TBS-Tween and incubated at RT for 1 h in secondary antibody (Jackson ImmunoResearch, Peroxidase conjugated Affinipure goat anti-mouse(H+L) #115-035-146 and goat anti-rabbit(H+L) #111-035-003) in 5% milk (concentration 1/10,000). After washing three times with TBS-Tween, membranes were exposed to chemiluminescent HRP substrate (Immobilon, #WBKLS0100) and revealed in docking system (Amersham Imager 600, GE). Densitometry analysis was carried out on raw Tif images by using Fiji ImageJ 1.53c.
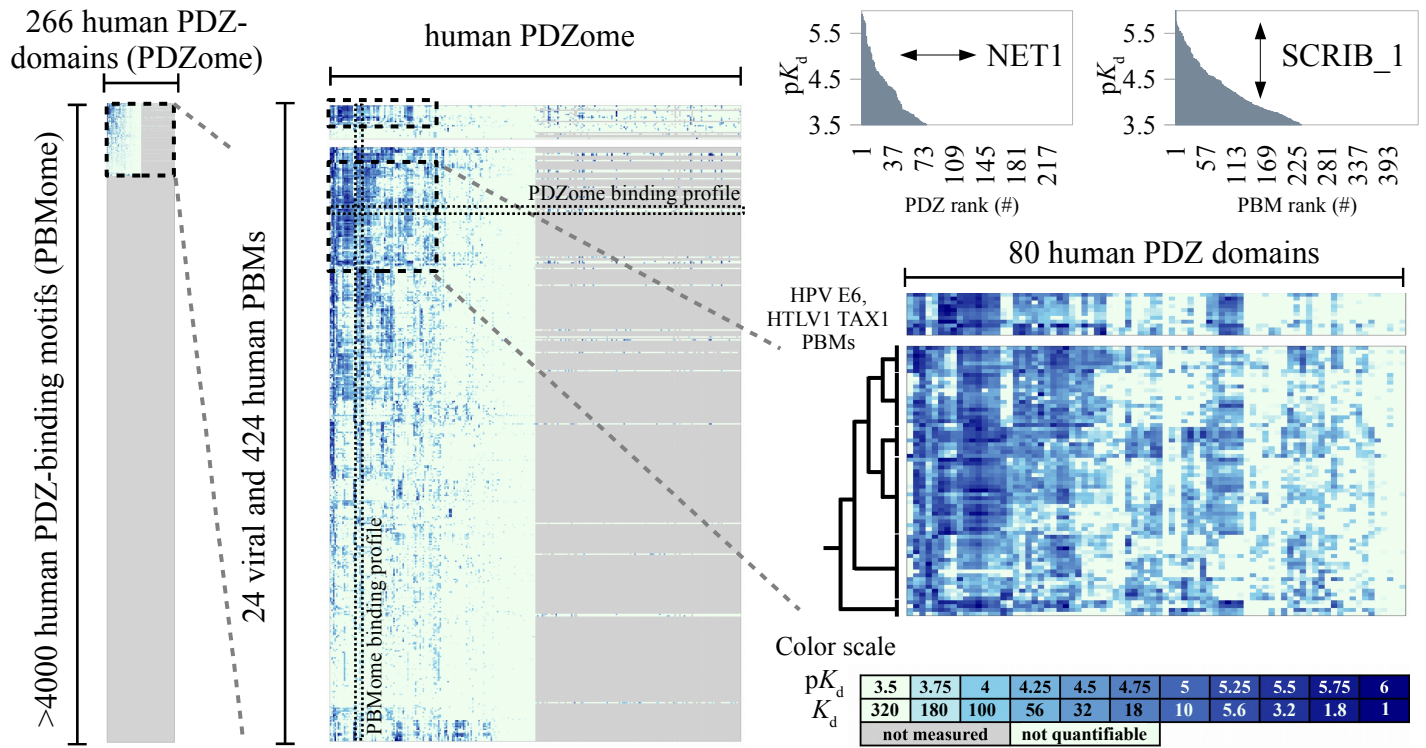
Figure 1,

Figure 2,

Figure 3,

Figure 4,

Figure 5,

A Proteome change E6 *vs.* E6ΔPBM

B Proteome difference E6 *vs.* E6ΔPBM
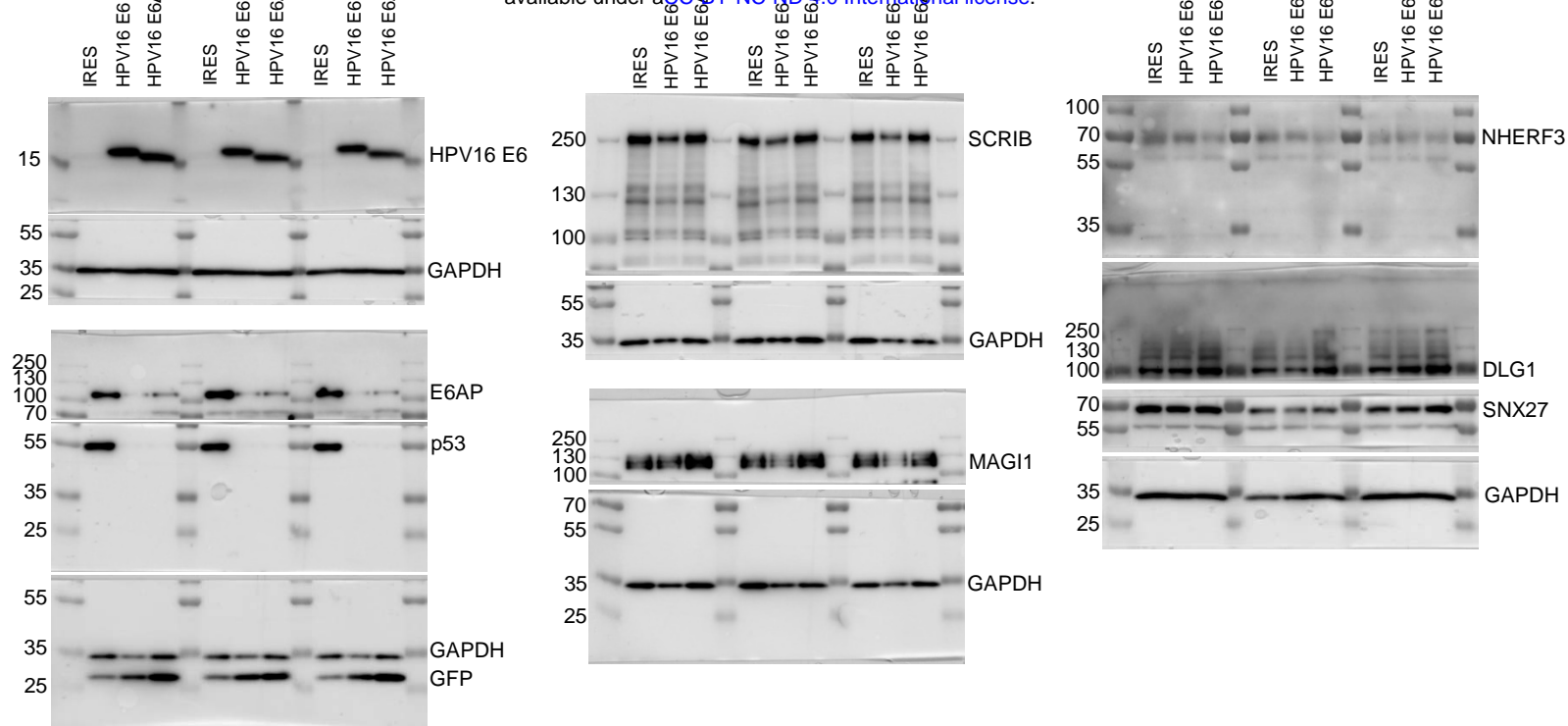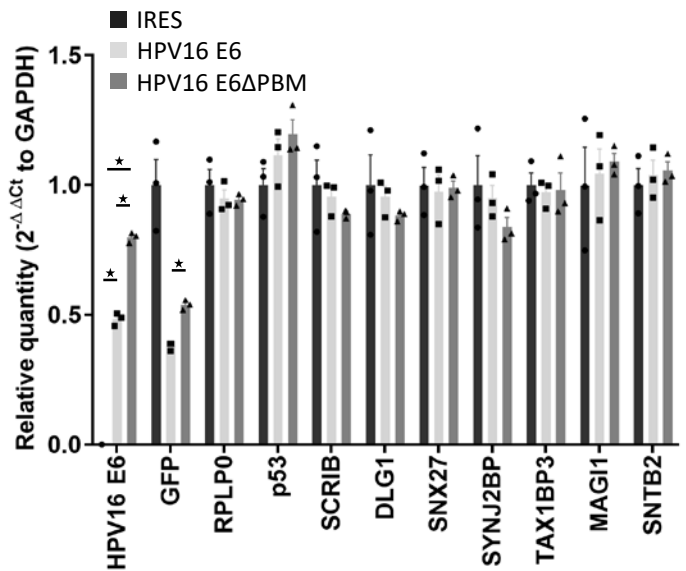
C Proteome change E6 *vs.* E6ΔPBM

D

E
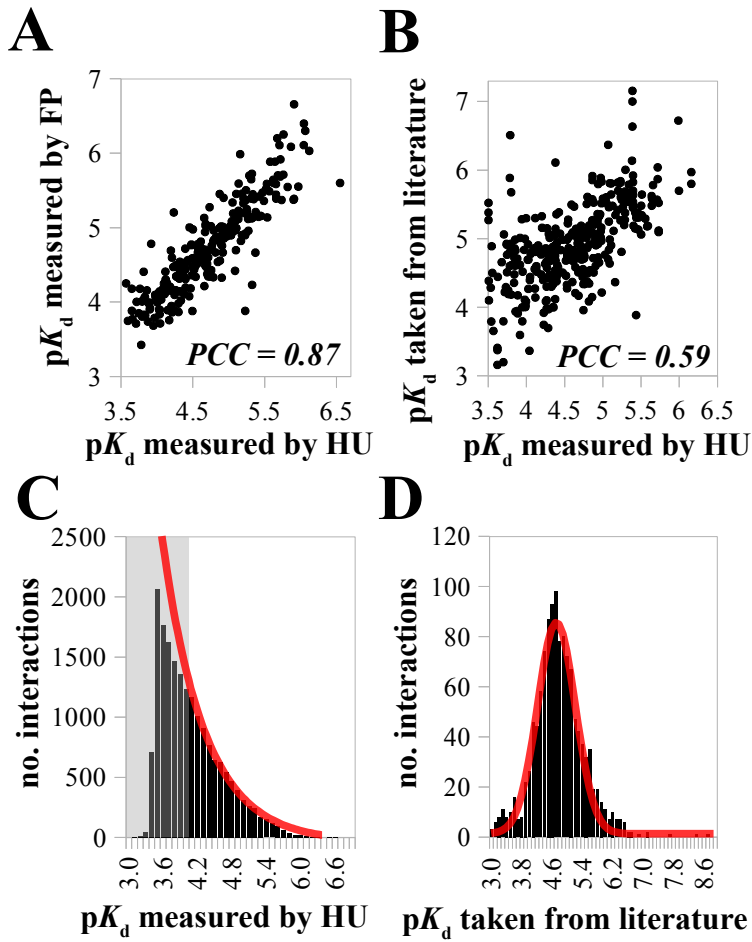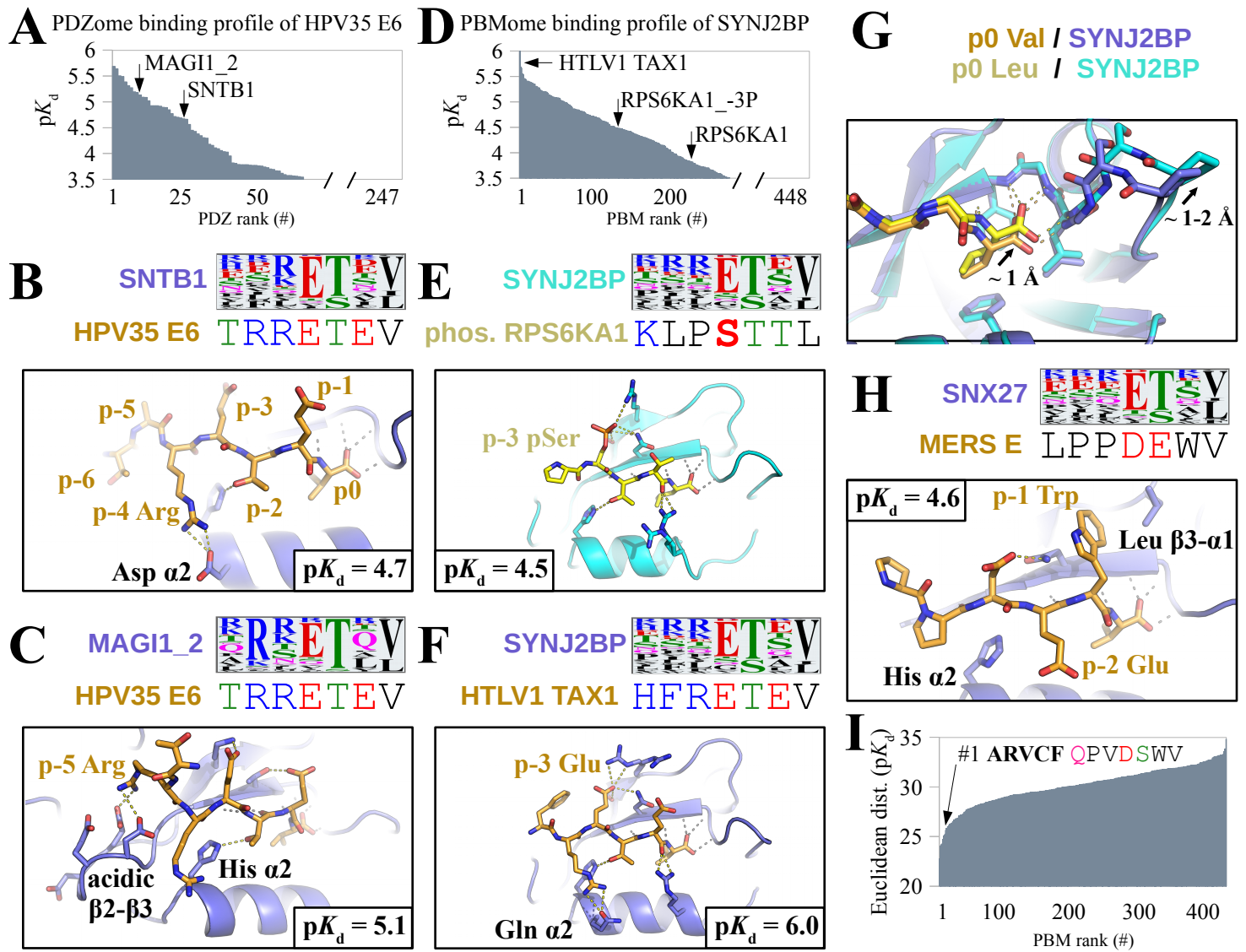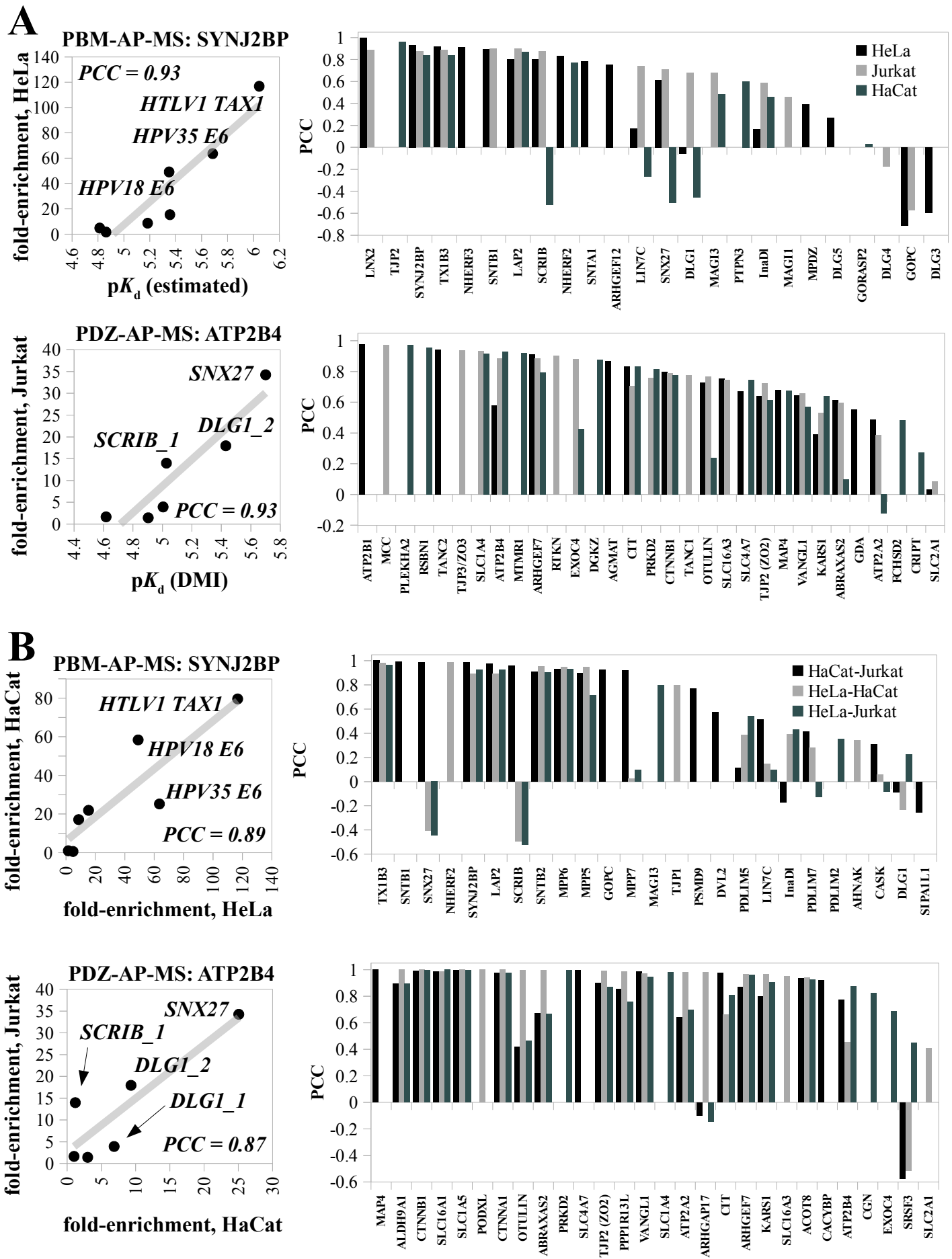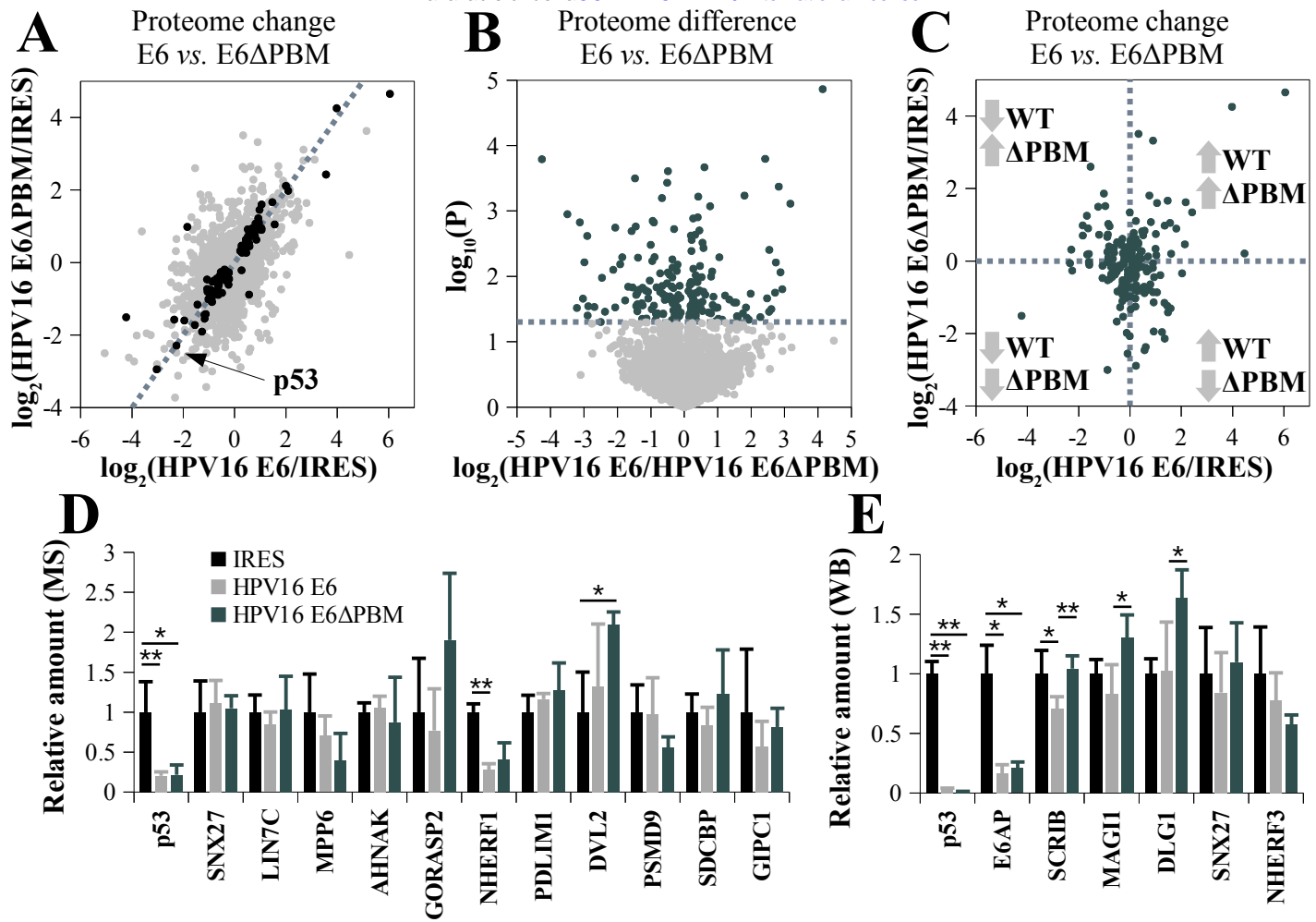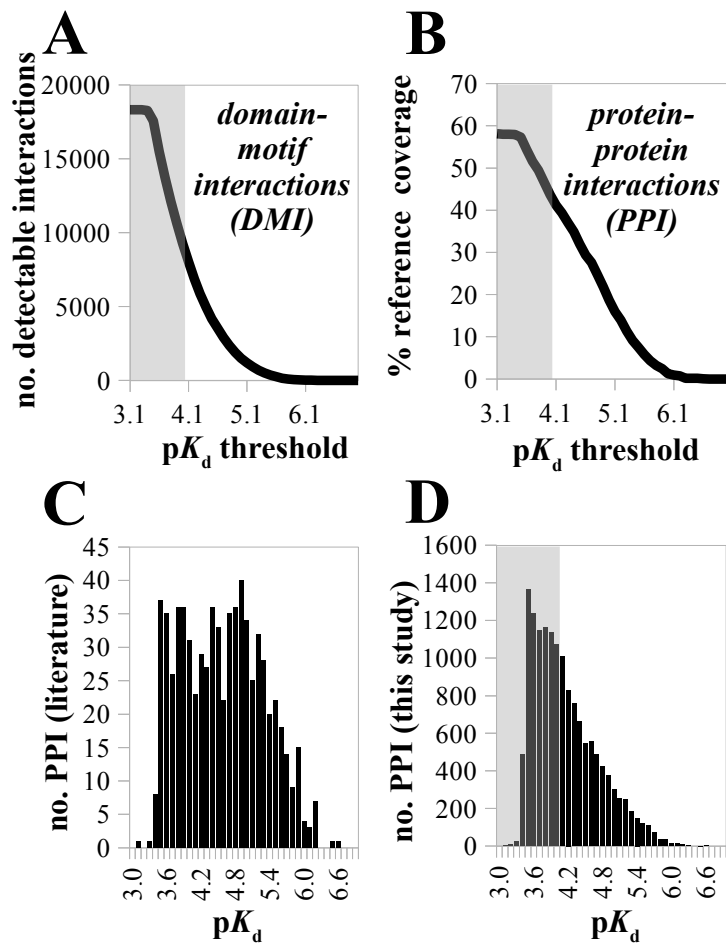
Figure 6,

Figure S1
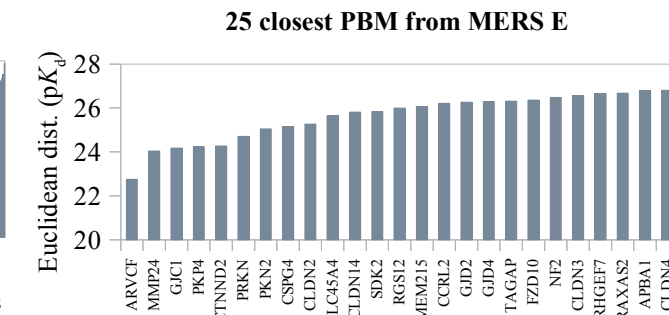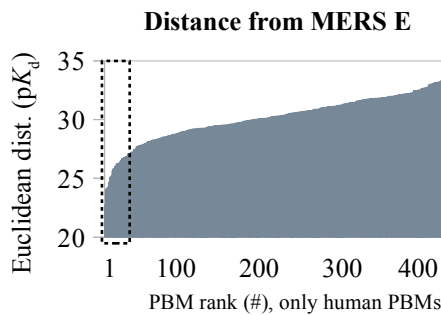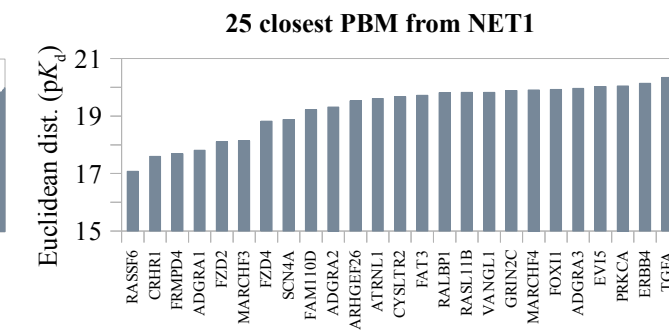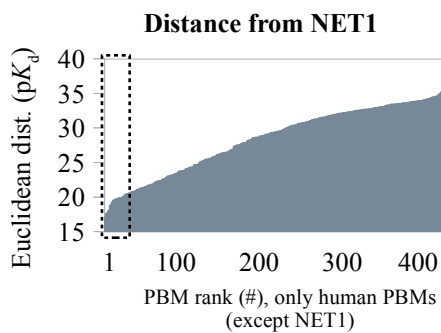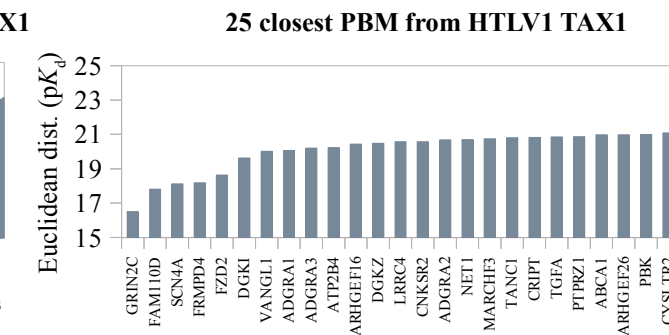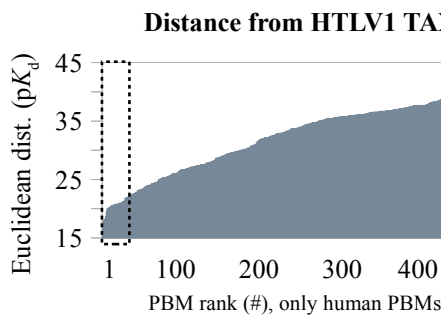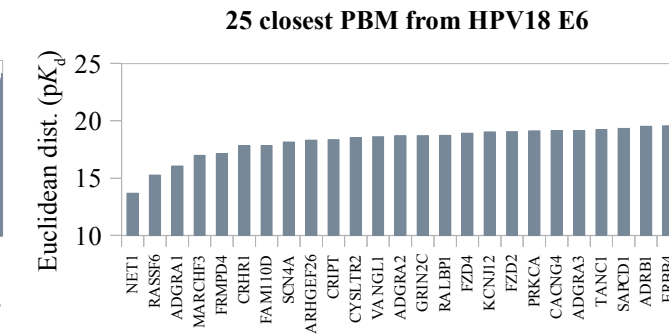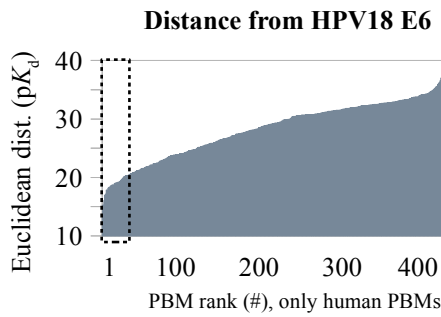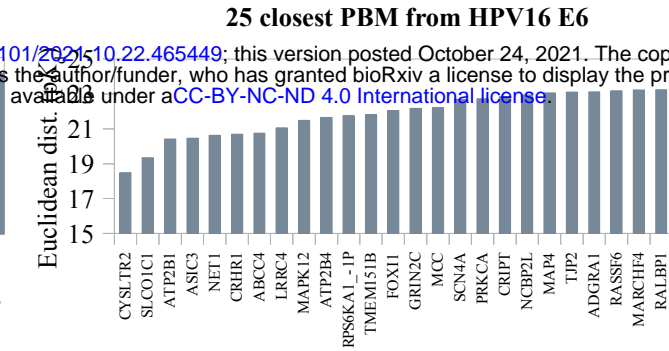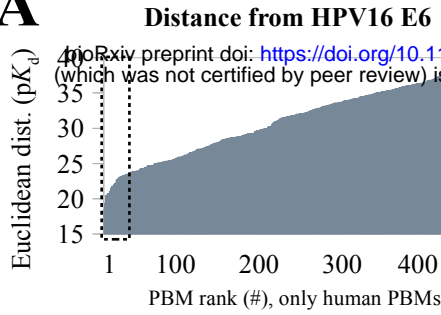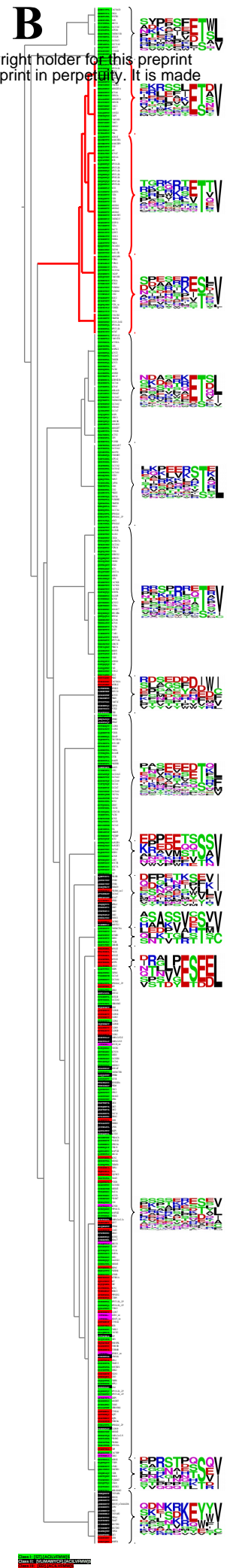
Euclidean distances were calculated in a completely explored affinity space
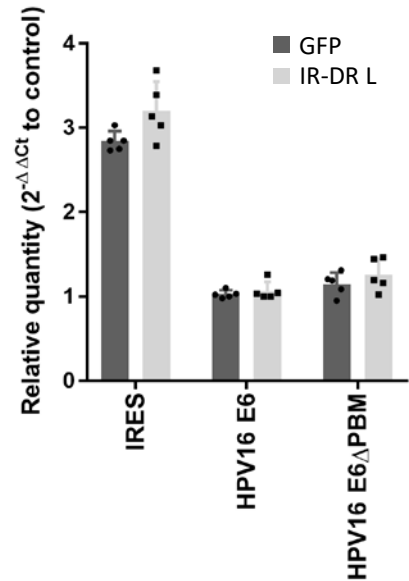*defined with 133 PDZ-domain affinity dimensions*