The FEBS Journal

FEBS PRESS science publishing by scientists

Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome

Norman E. Davey[1,*], Moon-Hyeong Seo[2,*], Vikash Kumar Yadav[3,*], Jouhyun Jeon[2], Satra Nim[2], Izabella Krystkowiak[1], Cecilia Blikstad[3], Debbie Dong[2], Natalia Markova[4], Philip M. Kim[2,5] and Ylva Ivarsson[3]

1 Conway Institute of Biomolecular and Biomedical Sciences, University College Dublin, Ireland
2 Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Canada
3 Department of Chemistry – BMC, Uppsala University, Sweden
4 Malvern Instruments Nordic AB, Solna, Sweden
5 Department of Molecular Genetics and Department of Computer Science, University of Toronto, Canada

Keywords
EVH1 domain; PDZ domain; Protein–protein interactions; short linear motifs; VHS domain

Correspondence
N. E. Davey, Conway Institute of Biomolecular and Biomedical Sciences, University College Dublin, Dublin 4, Ireland
Fax: +353 1 716 6701
Tel: +353 1 716 6700
E-mail: norman.davey@ucd.ie
and
P. M. Kim, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada
Fax: +1 416 978 8287
Tel: +1 416 946 3419
E-mail: pm.kim@utoronto.ca
and
Y. Ivarsson, Department of Chemistry – BMC, Uppsala University, Husargatan 3, 751 23 Uppsala, Sweden
Tel: +48 17 4714038
E-mail: ylva.ivarsson@kemi.uu.se

*Equal contributions

(Received 26 August 2016, revised 4 December 2016, accepted 19 December 2016)

doi:10.1111/febs.13995

The intrinsically disordered regions of eukaryotic proteomes are enriched in short linear motifs (SLiMs), which are of crucial relevance for cellular signaling and protein regulation; many mediate interactions by providing binding sites for peptide-binding domains. The vast majority of SLiMs remain to be discovered highlighting the need for experimental methods for their large-scale identification. We present a novel proteomic peptide phage display (ProP-PD) library that displays peptides representing the disordered regions of the human proteome, allowing direct large-scale interrogation of most potential binding SLiMs in the proteome. The performance of the ProP-PD library was validated through selections against SLiM-binding bait domains with distinct folds and binding preferences. The vast majority of identified binding peptides contained sequences that matched the known SLiM-binding specificities of the bait proteins. For SHANK1 PDZ, we establish a novel consensus TxF motif for its non-C-terminal ligands. The binding peptides mostly represented novel target proteins, however, several previously validated protein–protein interactions (PPIs) were also discovered. We determined the affinities between the VHS domain of GGA1 and three identified ligands to 40–130 μM through isothermal titration calorimetry, and confirmed interactions through coimmunoprecipitation using full-length proteins. Taken together, we outline a general pipeline for the design and construction of ProP-PD libraries and the analysis of ProP-PD-derived, SLiM-based PPIs. We demonstrated the methods potential to identify low affinity motif-mediated interactions for modular domains with distinct binding preferences. The approach is a highly useful complement to the current toolbox of methods for PPI discovery.

Abbreviations
AP-MS, affinity-purification coupled to mass spectrometry; co-IP, coimmunoprecipitation; EVH1, enabled/VASP homology 1; GO, gene ontology; GST, glutathione-S-transferase; GYF, glycine-tyrosine-phenylalanine; ITC, isothermal titration calorimetry; LBD, ligand-binding domain; NGS, next-generation sequencing; PBS, phosphate-buffered saline; PDZ, PSD-95, disks large, zona occludens 1; PPI, protein–protein interaction; ProP-PD, proteomic peptide phage display; SLiM, short linear motif; VHS, VPS-27, Hrs, and STAM; Y2H, yeast-two-hybrid.

Introduction

Recent years have seen a tremendous growth in the number of characterized human protein–protein interactions (PPI). A large-scale yeast-two-hybrid (Y2H) screen provided information on 14 000 potential binary PPI [1] and high-throughput affinity-purification coupled to mass spectrometry (AP-MS) of ~ 2600 bait proteins in HEK293T cells revealed more than 23 000 binary interactions or complexes [2]. In these datasets there are striking numbers of new interactions, bearing witness to the large set of unknown PPIs awaiting discovery. For example, 86% of the interactions from the AP-MS study were previously unknown. Certain categories of interactions, such as the interactions between short linear motifs (SLiMs) and SLiM-binding modules, remain particularly underrepresented [3]. SLiMs typically bury only three or four residues in the SLiM-binding pocket of their binding partner and the resulting interactions are often of low-to-medium affinity and transient. SLiM-based interactions are consequently easily lost in other high-throughput PPI discovery experiments [4] such as AP-MS. However, SLiMs are crucial for cellular signaling where transient SLiM-domain interactions are often utilized to propagate signals throughout the cell [5]. Furthermore, SLiMs encode much of the regulatory program of a protein by controlling their stability, localization, and modification state [3]. Due to their central role in cell physiology SLiM-mediated interactions drive evolution of signaling networks [6,7], are frequently deregulated in diseases such as cancers [8] and are often mimicked by pathogens to hijack host systems [9–11].

Intrinsically disordered regions cover approximately 30% of the human proteome and extensive disordered regions exist in all cellular systems [12,13]. However, outside a few well-studied proteins, few of these regions have been characterized. The predominant functional modules in these regions are SLiMs [4]. It has been estimated that the human proteome holds more than 100 000 SLiMs but, to date, only a small fraction of the expected repertoire has been discovered [14]. Consequently, there is a need for unbiased large-scale methods to identify binding SLiMs and to link them to the protein modules that recognize them. Historically, the vast majority of SLiMs have been characterized by low-throughput experimentation [15,16]. These studies generally validate novel SLiMs on a small scale, rarely discovering more than a few motif instances at a time. Although peptide arrays allow the screening of thousands of peptides in parallel, they are generally better suited for characterizing motif-binding domains with *a priori* knowledge of the binding preferences or interactors due to the limited number of peptides that are typically spotted on an array. Peptide arrays are hence typically not aimed at directly discovering novel motif instances in the proteome. Combinatorial peptide phage display with highly diverse libraries has been used to successfully determine peptide-binding preferences for a variety of modular domains, such as the PDZ domain family [17]. The identified consensus motifs can then be used to identify potential interaction partners through motif scanning. However, such predictions may be hampered due to a bias toward overly hydrophobic sequences in peptide phage selections, which may lead to tedious experimental validations [18]. There is thus a need for more efficient approaches for discovery of SLiM-based interactions.

In recent years, a variety of high-throughput methods have been developed to identify SLiM-based interactions [19]. Among the emerging methods for discovery of SLiM-based PPIs, is proteomic peptide phage display (ProP-PD) [20–22]. In ProP-PD, phage libraries are engineered to display defined regions of a target proteome. These libraries are then used in selections against bait proteins and retained phage pools are subjected to next-generation sequencing (NGS), which provides a list of ligands of potential biological relevance. In a proof-of-principle experiment, we previously designed a phage library to display all the C-terminal peptides of the human proteome. We tested the C-terminal ProP-PD library against a set of PDZ (PSD-95, disks large, zona occludens 1) domains [20,23], which are known to mainly interact with C-terminal peptides of target proteins, and successfully confirmed identified binders through affinity measurements and coimmunoprecipitations (co-IP) [24].

In this study, we expand the ProP-PD method to create a ProP-PD library displaying the intrinsically disordered regions of the human proteome. We outline a general pipeline for ProP-PD library design, bait protein screening and the analysis of ProP-PD-derived data. We validate the performance of the intrinsically disordered ProP-PD library against SLiM-binding domains from different domain families with distinct binding preferences, namely the enabled/VASP homology 1 (EVH1) domains of protein enabled homolog (ENAH) and Ena/Vasp-like protein (EVL); the glycine-tyrosine-phenylalanine (GYF) domain of PERQ amino acid-rich with GYF domain-containing protein 1 (GIGYF1); the ligand-binding domain (LBD) domains of nuclear receptors peroxisome proliferator-activated receptor gamma (PPARG) and nuclear receptor subfamily 5 group A member 2

(NR5A2); the PDZ domains of the SH3 and multiple ankyrin repeat domains protein 1 (SHANK1) and Disks large homolog 1 (DLG1); and the VHS (Vps27, Hrs, and STAM) domain of ADP-ribosylation factor-binding protein GGA1. In three cases, we included two domains of each family, to explore how the method performs when analyzing domains with similar binding preferences.

In the case of GGA1, we further validated the interactions between the recombinant VHS domain and synthetic 16 amino acid peptides, as well as its interactions with full-length proteins containing the disordered regions through co-IPs, thus demonstrating the relevance of identified interactions in the context of full-length proteins.

Results

ProP-PD library design and construction

A ProP-PD library was designed to cover the intrinsically disordered regions of the human proteome with 16-mer peptides. Peptides were tiled with an overlap of seven amino acids between peptides to optimize the display of the peptides and the coverage of intact SLiMs (Fig. 1A). The total library design consisted of 479 846 peptides derived from 18 682 proteins. Oligonucleotides encoding designed peptides flanked by annealing sites for combinatorial mutagenesis were printed on custom microarrays and obtained as an oligonucleotide library (Fig. 1B). The oligonucleotide library was used in a combinatorial mutagenesis reaction to create a library of genes encoding for the designed peptides fused N terminally to the M13 major coat protein P8 in a phagemid vector. The library was sequenced to evaluate the coverage of the designed library. Of 26 936 810 total identified reads, 3 378 746 mapping to 301 147 unique designed peptides were complete and in the correct frame. The large number of reads that did not map to any designed peptide typically represented frame-shifted or truncated versions of the design, which is largely explained by oligonucleotide library quality issues. The correctly mapped peptide sequences correspond to 62.7% coverage of the designed ProP-PD library. For future libraries, advances in oligonucleotide library synthesis quality and explicit addition of redundancy in library design can be leveraged to increase the coverage of the

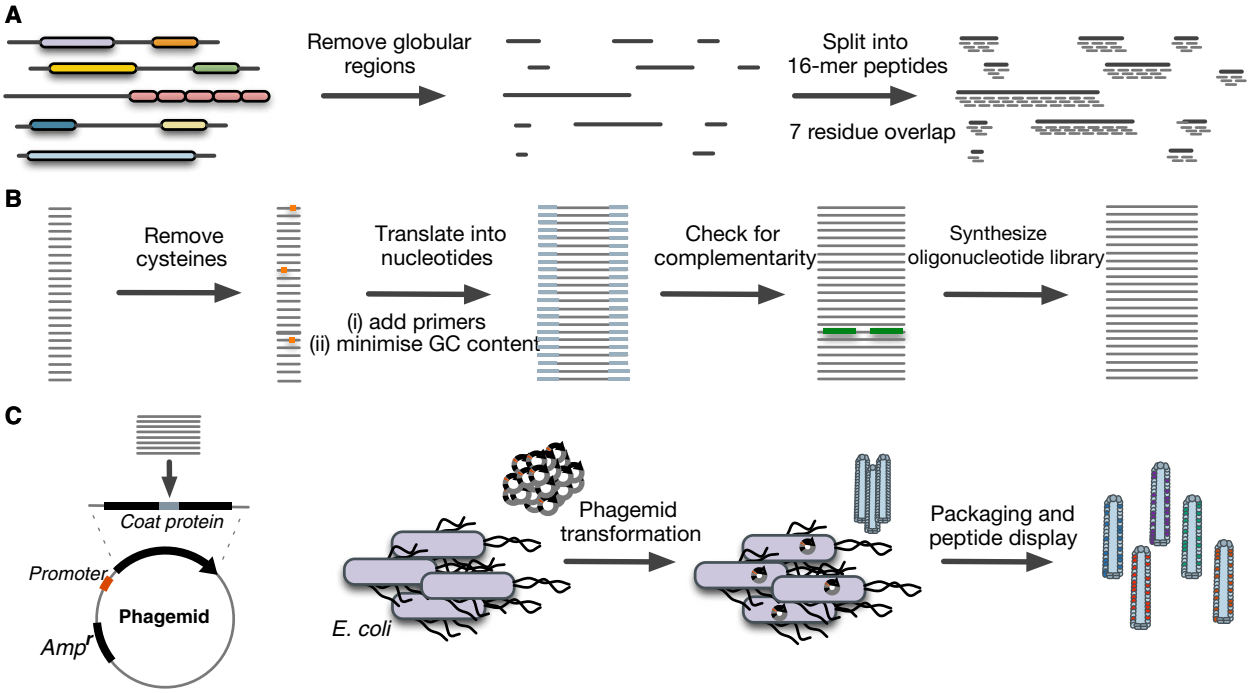

Fig. 1. Schematic overview of the design and construction of ProP-PD library of the intrinsically disordered regions of the human proteome. (A) The human proteome was scanned for disordered regions using the IUPred algorithm. The regions were split into 16 amino acid peptides, with an overlap of seven amino acids. (B) The peptides were translated into oligonucleotides, primers complementary to the template phagemid were added and the library was further optimized before being obtained from a commercial source. (C) Oligonucleotides encoding the sequences were inserted into a phagemid designed for the display of peptides fused to the N terminus of M13 P8. The library was transformed into *Escherichia coli* preinfected with M13 KO7 helper phage.

library. Although incomplete, the current ProP-PD library serves its purpose; the library quality is sufficient for identification of known and novel natural binders.

ProP-PD selection and initial data analysis

The ProP-PD library was used in phage selections (Fig. 2) against the following immobilized bait proteins: ENAH EVH1, EVL EVH1, GIGYF1 GYF, NR5A2 LBD, PPARG LBD, DLG1 PDZ2, SHANK1 PDZ, and GGA1 VHS. Enriched phage pools from the fourth round of selections were barcoded, pooled, and analyzed by NGS on the Illumina platform. We decided a cut-off of sequencing counts of peptides for each bait protein and established a consensus motif for each dataset of binding peptides using the SLiMFinder algorithm [25] (Table 1). We included ligands with sequencing counts below the threshold value for further analysis if they contained sequences that matched the consensus motif. After the processing, between 4 and 103 unique ligands were identified as binders for each bait protein (Table 1, Table S1). These unique peptides represented between 4 and 97 target proteins per bait protein; some proteins were identified as ligands based on more than one peptide and some peptides match to more than one protein.

Analysis of identified consensus motifs

In the cases where the binding preferences were previously known, the identified consensus motif matched the known specificity of the bait domains [26] (Table 1, Fig. 3A). EVH1 and GYF domains recognize proline-rich motifs [27,28]. LBD domains are known to interact with LxxLL motifs [29] and VHS domains typically bind to DxxLL motifs [30]. In all cases, the consensus motif was found in the majority of the identified peptides (Table 1). To our knowledge, there was no known consensus for internal PDZ domain-binding motifs available prior to this study for the PDZ domains of SHANK1 and DLG1, but they are known to bind the C-terminal motifs TxΦ-COOH and TxV-COOH, respectively (where Φ is a hydrophobic amino acid) [17,20,31]. We established a new TxF consensus motif as the preferred internal ligand for the SHANK1 PDZ domain, supported in three instances by overlapping peptides (Table 2). The TxF motif is similar to SHANK1 PDZ's specificity for C-terminal ligands, but the preference for internal motifs appears to be more stringent. In line with our results, a TxF containing stretch in Glutamate receptor delta2 was identified as binding site for SHANK1 PDZ in the original study reporting on SHANK1 PDZ's capability to bind internal ligands [32]. For DLG1 PDZ2 there were too few ligands to derive a motif, but we note

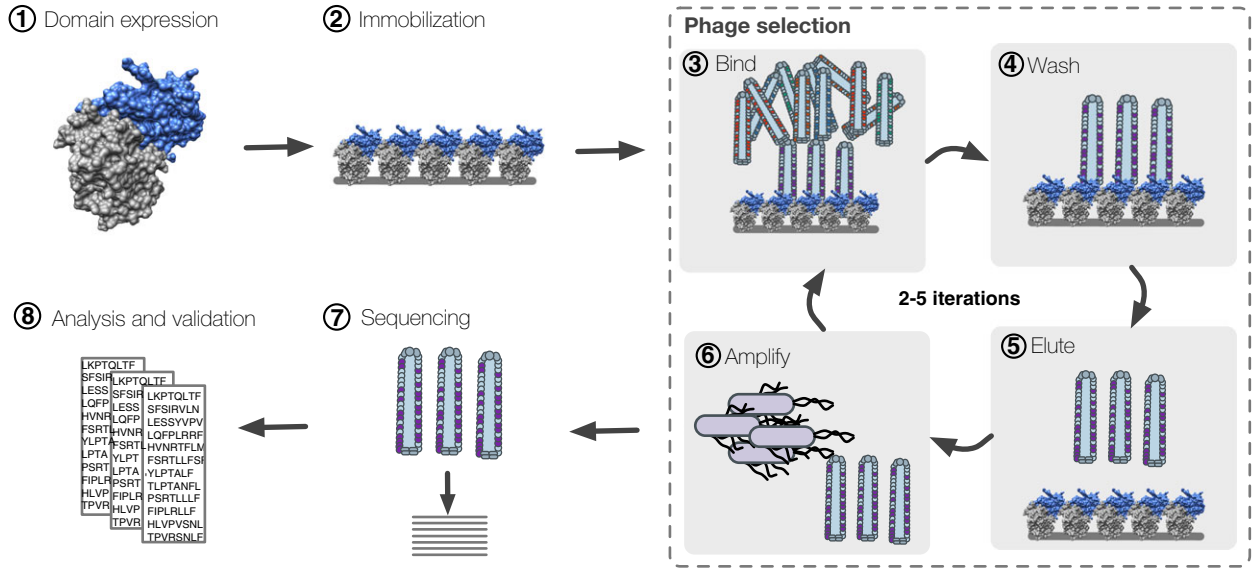

Fig. 2. Schematic overview of ProP-PD selections against GST-tagged bait proteins. GST-tagged (indicated in gray) bait proteins (in blue) are expressed and purified (1) and nonspecifically immobilized on a hydrophobic surface (2) and then used as baits for up to five repetitive rounds of phage selections (3–6). Amplicons are prepared by barcoding peptide-coding regions using enriched phage pools as PCR templates. The amplicons are analyzed by next-generation sequencing (7). The identified ligands are analyzed for consensus motifs (8) and matched against the library design, which identifies the peptide-containing host proteins. Targets are then selected for validations through biophysical affinity measurements and cell-based assays.

Table 1. Enriched motifs among the datasets of ligands obtained from ProP-PD selections against distinct bait proteins.

Bait	Expected consensus	Observed consensus	Peptides	With consensus	Consensus coverage (%)	Significance
EVL EVH1	[FYWL]Px[ALIVTFY]P [16]	[FW]Pxx[LP]	62	46	74.2	$< 1 \times 10^{-10}$
ENAH EVH1	[FYWL]Px[ALIVTFY]P [16]	[FLW]Px[AP]P	33	22	66.6	6.65×10^{-8}
GIGYF1 GYF	PPG[FILMV] [61]	[ALP]PG[FILMY]	70	48	68.6	$< 1 \times 10^{-10}$
PPARG LDB	LxxLL [62]	LxxLL	21	18	85.7	1.89×10^{-6}
NR5A2 LDB	LxxLL [62]	LxxLL	32	24	75.0	3.37×10^{-7}
SHANK1 PDZ	Unknown	TxF	103	81	78.6	$< 1 \times 10^{-10}$
DLG1 PDZ2	Unknown	None	4			
GGA1 VHS	[DE]xxL[LI] [63]	Dxx[AILM][ILMV]	54	46	85.2	$< 1 \times 10^{-10}$

that two out of four DLG1 PDZ2 peptides contain a 'TxF' motif.

Comparison between ProP-PD derived ligands of homologous domains reveals partially overlapping datasets

Many motif-binding domains are members of large families of closely related domains with overlapping specificities [3]. We included three pairs of homologous domains (the EVH1 domains of ENAH and EVL, the LBD domains of PPARG and NR5A2, and the PDZ domains of SHANK1 and DLG1) in our analysis. No overlap was found between the ligands of the PDZ domains, which might be explained by the low number of DLG1 PDZ2 peptides and subtle differences in their peptide-binding preferences [17]. The selections against the EVH1 domains identified overlapping sets of ligands as can be expected by their shared binding preferences (Fig. 3B). Similarly, the selections against the LBD domains identified partially overlapping sets of ligands. These overlaps far exceed the peptide overlap between unrelated domains. Only two identified peptides contain binding motifs for nonhomologous binding domains: One peptide was found as a ligand of both SHANK1 PDZ and EVL EVH1 and contains peptides matching the consensus binding motif for each of the domains (VVRDFPAPLPESTVFS). In addition, one ligand was shared between the ENAH EVH1 and GIGYF1 GYF domains (TPLPPPPPPPPGLPTY). Such regions with overlapping or adjacent motifs are commonly seen in proteins and are often a hallmark of regulatory switching mechanisms [33].

Comparison between ProP-PD ligands and previously known ligands

The method returned several SLiMs that have previously validated as ligands for a given bait domain, including the EVH1 domain-binding motif in Palladin (PALLD) [34]; the LBD domain-binding LxxLL motifs in Nuclear receptor-interacting protein 1 (NRIP1) [35], Transcription intermediary factor 1-alpha (TRIM24) [36] and Histone acetyltransferase p300 (EP300) [37]; and the VHS domain-binding motifs in ADP-ribosylation factor-binding protein GGA1 (GGA1) [38] and Sortilin-related receptor (SORL1) [39] (Table 3). The known NRIP1 LxxLL motif was returned in both the NR5A2 and PPARG selections. Furthermore, the internal NOS1 PDZ motif discovered in the DLG1 selection has been shown to interact with the PDZ domain of the DLG1 paralog, DLG4 [40]. However, many SLiM instances known to serve as binding sites for each bait protein were not identified in these experiments. This can at least partially be attributed to the incomplete coverage of the ProP-PD library.

We also investigated to what extent the proteins containing identified ProP-PD ligands overlap with known protein interactions by a comparison with interactors of the bait proteins (or their paralogs) listed in the HIPPIE integrated protein–protein interaction database [41]. In addition, we performed a manual literature search. We found literature support for identified ligands for the EVH1 domains of ENAH, for the LBD domains of PPARG and NR5A2 and for the VHS domain of GGA1 (Tables S1 and S4; Fig. 3C). For GIGYF1 GYF, DLG1 PDZ2, and SHANK1 PDZ we did not find any literature evidences corroborating our findings, with the exception of the DLG4–NOS1 interaction described above. In total, 4.4% (16 interactions) of the identified peptides were supported by interaction data. While these numbers may appear low, they can be compared to the recent high-throughput AP-MS study, where 86% of the interactions were novel, and might indicate the low coverage of SLiM-mediated interactions in the current human PPI databases. Indeed, the techniques upon which a large fraction of this data is based upon (namely, AP-MS and Y2H) likely have poor sensitivity toward these interactions. Hence, our results underline the potential of ProP-PD to discover SLiM-based interactions overlooked by other methods.

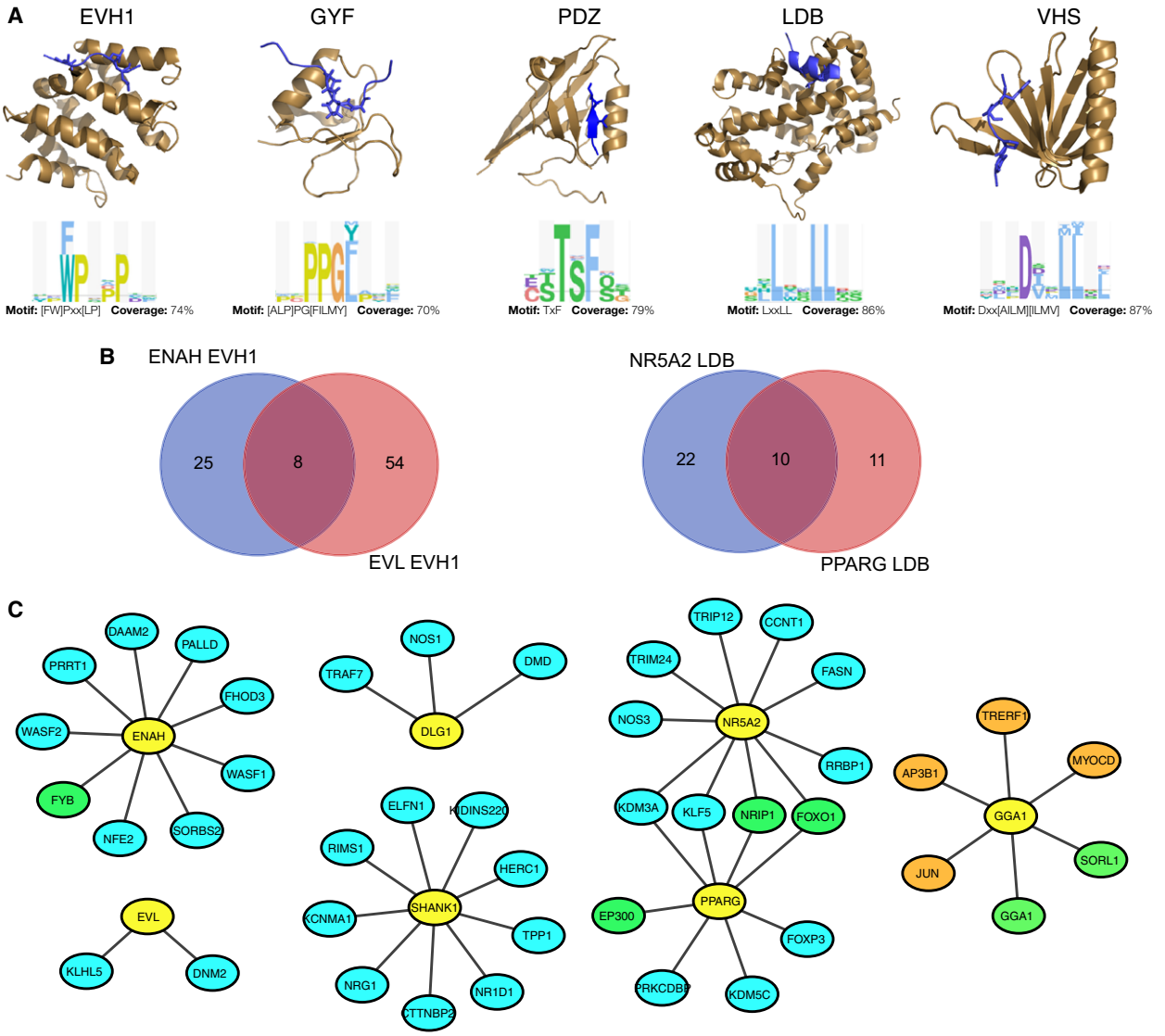

Fig. 3. Analysis of identified ligands: Representative structures of the domains used as bait proteins and their consensus motifs as established through ProP-PD (A), overlap between identified peptide ligands for domains of the same family (B), and network representations of high-confidence set of identified interactions (C). (A) Protein structures are shown in golden cartoon representation, and the bound peptides in blue cartoons. The EVH1 domain of EVL is bound to a EFPPPPT peptide (PDB: 1QC6), the GYF domains of CD2BP2 is bound to a SHRPPPPGHRV peptide (PDB: 1L2Z), the PDZ domain of SHANK1 is bound to a DETNL-cooh peptide (PDB: 3QJN), the LBD domains of PPARG is bound to a ARHKILHRLLQE peptide (PDB: 2P54), and the VHS domain of GGA1 bound to a DDISLLK peptide (PDB: 1UJJ). Key residues are indicated in the structures (sticks). The figure was made using PYMOL. The SLiMFinder defined consensus motifs are shown as relative binomial logos (Table 1). Logos show the $-\log^{10}$ of the binomial probability. The binomial probability is calculated as $\text{prob}^{aa} = \text{binomial}(k, n, p)$ where k is the observed residue count at each position for a residue, n is the number of the instances of motifs, and p is the background frequency of the residue in the intrinsically disordered regions of the human proteome. The gray line annotated as $P(0.05)$ signifies the height of a amino acid that has a prob^{aa} of 0.05. Shown are the relative binomial logos for the EVH1 domain of EVL, the GYF domain of GIGYF1, the PDZ domain of SHANK1, the LBD domain of PPARG, and the VHS domain of GGA1. (B) Venn diagram of overlapping peptide ligands identified for the EVH1 domains of ENAH and EVL (left) and the LDB domains of NR5A2 and PPARG. (C) Networks of identified high confidence ligands (Table S4), with bait proteins indicated in yellow and target proteins in blue, green, or orange. Proteins indicated in green are previously known ligands of the bait proteins. Proteins in blue share relevant GO terms with the bait proteins and/or are known to interact with bait protein paralogs. GGA1 targets indicated in orange were validated in the current study. The networks were visualized using Cytoscape.

Table 2. Overlapping target peptides identified using SHANK1 or EVL EVH1 as bait proteins. Overlapping regions are underlined, and consensus binding motifs are indicated in bold.

Bait	Peptide A	# NGS	Peptide B	# NGS	Gene	UniProt
SHANK1 PDZ	VTTSPSASST**TSF**MSS	6	T**TSF**MSSSLEDTTTAT	8	HERC1	Q15751
SHANK1 PDZ	DLESLAPWES**TDF**RGP	2	S**TDF**RGPSAVSIQAPG	2	GPR179	Q6PRD1
SHANK1 PDZ	SDVSDVSAISRTSSAS	2	SRTSSASRLSSTSFMS	34	RIMS1	Q86UR5
SHANK1 PDZ	QEYQSRSPDILET**TSF**	2	ILET**TSF**QALSPANSQ	2	SALL4	Q9UJQ4
EVL EVH1	NPLSLDSARWP**LPPLP**	8	WP**LPPLP**LSATGSNAI	344	IRS4	O14654

Significant enrichments of relevant gene ontology terms

For each bait protein, we investigated if the proteins containing the identified peptide ligands exhibit any functional enrichment through gene ontology (GO) term (biological process, cellular compartment, molecular function) and KEGG pathway enrichment analysis, using DAVID version 6.8 [42]. The analysis revealed significant enrichments (FDR < 0.01) of GO terms among the ligands of ENAH EVH1 (actin cytoskeleton, actin binding, lamellipodium, and translation activator activity), GGA1 VHS (homophilic cell adhesion via plasma membrane adhesion molecules, calcium ion binding, transcription factor binding), GIGYF1 GYF domain (chromatin binding), PPARG LDB (pathways in cancer), and DLG1 PDZ (sarcolemma) (Table S2). We further performed a GO term enrichment analysis using GORILLA [43], with similar results (Table S3). GORILLA also identified significant GO term enrichments among PPARG ligands for terms related to negative regulation of metabolic processes and hormone receptor binding. Most of the significantly enriched GO terms generated for ligands identified for a given bait protein were linked to the biological function of the bait protein.

We further investigated if the target proteins exhibit any overlap with the GO term of their bait proteins. This pairwise analysis revealed shared GO terms with the bait protein that is unlikely to occur by chance ($P < 0.01$) for 11% of the identified peptides (Tables S1 and S4). For example, numerous proteins containing disordered regions that bind SHANK1 PDZ share significant GO terms with SHANK1. Many of these GO terms reflect a shared localization or role in neurons such as 'excitatory synapse', 'positive regulation of excitatory postsynaptic potential', and 'positive regulation of dendritic spine'.

In total, 12.5% of the identified peptides contain a previously characterized SLiM for the bait protein, are contained within a protein that is a known interactor of the bait protein or share a GO term overlap with the bait protein that is unlikely to occur by chance (Table S4; Fig. 3C). Taken together, the high coverage of the consensus, the extensive overlap of peptide targets for homologous domains, the rediscovery of previously characterized SLiM-mediated and full-length protein interactions, and the significant GO term similarity between the identified ligands and bait proteins offer strong evidence that ProP-PD selections discover novel binders that are biologically relevant for the signaling network.

Table 3. Previously characterized motifs returned by the screen (references in the text).

Bait	Peptide	Gene	UniProt
EVL EVH1	PDV**FP**LP**P**PPPPLPSP [34]	PALLD	Q8WX93
PPARG LDB	DAASKHKQ**L**SE**LL**RSG [37]	EP300	Q09472
PPARG LDB & NR5A2 LDB	SPKPSVAASQ**L**AL**LL**S [35]	NRIP1	P48552
	NANYPRSI**L**TS**LL**LNS [36]	TRIM24	O15164
DLG1 PDZ2	HLE**TTF**TGDGTPKTIR[a] [40]	NOS1	P29475
GGA1 VHS	ASVSLL**D**DE**LM**SLGLS [38]	GGA1	Q9UJY5
	DAPMITGFSD**D**VP**MV**I [39]	SORL1	Q92673

[a]The NOS1 peptide was shown to bind the DLG1 paralog DLG4.

Validations of GGA1 VHS ligands through isothermal titration calorimetry and coimmunoprecipitation

To investigate to what extent interactions between the modular domains and the peptides identified through ProP-PD are relevant in the context of the full-length proteins we focused on the bait protein GGA1 and four proteins found to contain disordered regions that interact with its VHS domain. The ligands were AP-3 complex subunit beta-1 (AP3B1), transcriptional-regulating factor-1 (TRERF1), the transcription factor AP-1 (JUN), and myocardin (MYOCD). Of these ligands, AP3B1 appears as a particularly relevant ligand as it is a subunit of the nonclathrin- and clathrin-associated adaptor protein complex AP-3 that is involved in protein sorting in the trans-Golgi network. GGA1 and AP-3 have previously been shown

Table 4. Binding parameters of selected GGA1 VHS ligands as determined by ITC measurements (triplicate measurements). '# NGS' indicates the number of counts that the peptides obtained in the NGS analysis of the phage pools after the fourth round of selection.

Gene	Peptide	# NGS	n^a	K_d (μM)	ΔH_{ap} (kJ·mol^{-1})
AP3B1	KDVSLLDLDDFN	6	1	40 ± 12	−11.9 ± 2
MYOCD	MSDVTLLKIGSE	4	1	61 ± 6	−9.5 ± 0.007
TRERF1	VDTDLLLDDQDS	2	1	132 ± 9	−5.3 ± 0.07

[a]Number of binding sites fixed to 1 in the fitting of the data.

to partially colocalize [44]. Isothermal titration calorimetry (ITC) measurements using recombinant GGA1 and synthetic peptides revealed that the affinities were in the range of 40–132 μM for the interactions between GGA1 VHS domain and the interacting peptides of AP3B1, MYOCD, and TRERF1 (Table 4; Fig. 4). No affinity value was obtained for the JUN peptide due to solubility issues. The measured affinities are similar to the values obtained through ITC measurements for previously characterized GGA1 targets [45], which demonstrate that ProP-PD is a suitable approach for identifying moderate- to low-affinity interactions of potential biological relevance. There is an inverse relation between the K_d values and the sequencing counts (Table 4) and the sequencing counts thus provide an affinity ranking of identified binding peptides, consistent with previous results [20]. Interactions between GGA1 and full-length proteins were confirmed for the four targets (AP3B1, TRERF1, MYOCD, and JUN) through co-IP of HA-tagged GGA1 and FLAG-tagged target proteins (Fig. 5), demonstrating that these ligands identified through ProP-PD are relevant in the context of the full-length proteins and in a cellular system. Our results thus support that the GGA1 VHS domain interacts with internal DxxLL motifs [46].

Discussion

The SLiM-based interactions are crucial for cell function but notoriously difficult to discover by most methods developed to explore PPIs on large scale. Indeed, most of the current information on SLiM-mediated interactions has been derived from low-throughput mutagenesis studies [16]. In this study, to allow for large-scale discovery of SLiM-based interactions, we created a ProP-PD library composed of peptides representing a large proportion of the intrinsically disordered regions of the human proteome. We demonstrate that this library can be used to discover SLiMs binding modular domains with distinct SLiM-binding preferences, and we show for GGA1 that the interactions identified are of relevance in the context of full-length proteins.

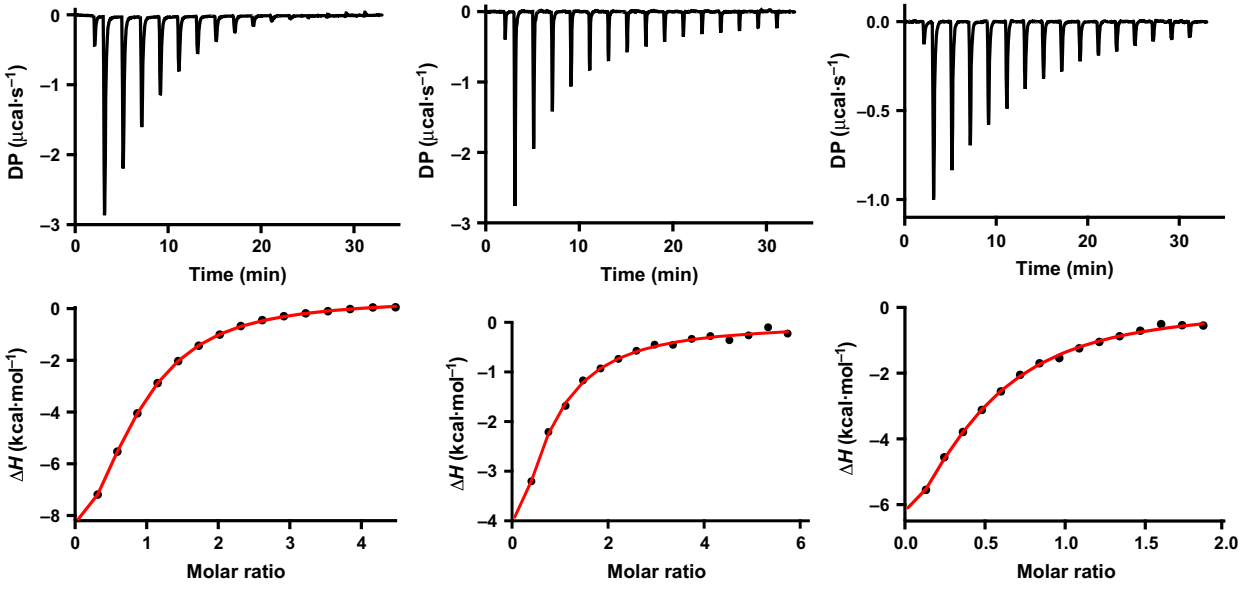

Fig. 4. ITC-based affinity measurements of the interactions between GGA1 VHS domain and dodecamer peptides of AP3B1 (left), TRERF1 (middle), and MYOCD (right). Experiments were performed at 25 °C in PBS, pH 8, using a MicroCal PEAQ-ITC (Malvern).

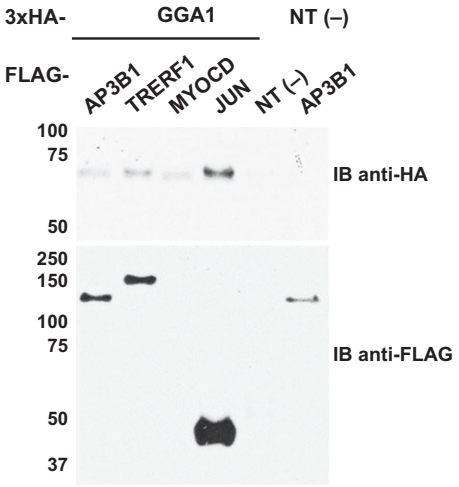

Fig. 5. Co-IPs of FLAG-tagged target proteins and 3xHA-tagged GGA1 in HEK293T cells. Marked FLAG-tagged proteins and 3xHA-tagged GGA1 were transiently cotransfected, followed by immunoprecipitation with anti-FLAG beads. Immunoblot detections were performed using anti-FLAG and anti-HA antibodies. NT, nontransfected.

ProP-PD is an unbiased method to discover accessible peptides in the human proteome, and to identify SLiMs that may represent potential biologically relevant ligands. High-throughput sequencing allows the large-scale discovery of SLiM-containing peptides. As the displayed peptides represent regions of the human proteome, the discovery of binding peptides through ProP-PD directly provides information on the potential target proteins. This limits the need for computational predictions of SLiM-containing target proteins based on consensus motifs. The peptide–bait domain pairs also provide the molecular details of the interaction that can be used to complement PPI networks obtained through standard methods. In addition, ProP-PD discovers interactions that are overlooked by methods such as AP-MS and thus allow us to expand the knowledge on PPIs into largely unexplored regions of the human interactome.

As shown here, ProP-PD returns novel and known ligands. Among the obvious limitations is the fact that SLiM-containing ligands are identified solely based on affinity, with a potential bias for interactions with slow dissociation rates. In the context of full-length proteins, additional regions of the proteins often provide additional interactions, and such multivalency contributes to the specificity and avidity of the interactions [47]. Whether such SLiM-mediated interactions within multipartite interfaces will be recognized by ProP-PD is unclear, however, the affinity range of the measured ProP-PD peptides suggests that the method will excel at characterizing low- to medium-affinity interactions. Furthermore, the method displays peptides to the bait protein without the spatiotemporal constraints such as cellular coexpression and colocalization of the bait protein and the targets resulting in interactions that cannot occur in the cell [48].

Consequently, we can consider the peptides identified by the ProP-PD method to comprise: (a) biologically relevant targets; (b) targets matching the bait domain's specificity yet lacking spatiotemporally restrictions and sequence context; and (c) experimental noise. Given the high coverage of correct motif consensus in the returned peptides the experimental noise appears to be low. The proportion of the identified peptides that are bona fide biological targets is unknown, but the high rate of interactions confirmed through co-IPs is reassuring. By integrating proteomic and ontological data, high confidence sets of biologically relevant peptides can be created. An additional limitation of the ProP-PD approach is that it does not account for post-translational modifications. This is a notable limitation, given the abundance of modifications that lead to the activation of interaction sites, as, for example, protein phosphorylation creating binding sites for phosphopeptide-binding domains [49,50]. Potentially, this limitation could be tackled by treatments of the naive phage library with desired kinases prior to selection, an approach previously taken for randomized peptide phage display [51].

Taken together, we expand our ProP-PD approach by creating a library designed to cover a large proportion of the human proteome. We thereby extended the applications of phage display, allowing the method to tackle systems biology related questions. We foresee that this method will be highly useful for the large-scale discovery of binding SLiMs and the complementary charting of PPI networks.

Materials and methods

Library design

The phage library was designed to tile the intrinsically disordered regions of the ~ 21 000 primary isoforms of reviewed human proteins from the UniProt resource. Each peptide is 16 amino acids in length and has an overlap of seven amino acids with the adjacent peptide (Fig. 1). The IUPred algorithm (cutoff = 0.3) was used to define the intrinsically disordered regions of the proteome. The resulting tiled library contained 479 846 peptides covering 4 757 112 residues in 18 684 proteins. All cysteine residues within the peptides were replaced with alanine as unpaired cysteines may compromise display on the M13 coat. The

peptides were reverse translated to oligonucleotides optimizing for *Escherichia coli* expression by replacing low abundance and high-guanine content codons with synonymous preferred codons. The peptide encoding oligonucleotides were flanked by primer annealing sites for PCR amplification and site-directed mutagenesis. Finally, codons resulting in complementary stretches within the oligonucleotide were replaced with noncomplementary codons. The designed oligonucleotide library was obtained from MYcroarray (Ann Arbor, MI, USA).

Library construction

The disorderome phage library was constructed following a published procedure [52,53] using 0.6 μg of the oligonucleotide library as primers for the oligonucleotide-directed mutagenesis. The phagemid library was converted into a phage display library by electroporation into *E. coli* SS320 cells preinfected with M13KO7 helper phage [53] with an efficiency of 2.6×10^8 transformants, thus oversampling the theoretical library size by more than 500 times. The phage-producing bacteria were grown over night in 500 mL 2YT (16 g Bacto tryptone, 10 g Bacto yeast extract, 5 g NaCl, per liter water) medium at 37 °C and then pelleted by centrifugation (10 min at 11 872 ***g***). The supernatant was transferred to a new tube and phages were precipitated by adding 1/5 volume polyethylene glycol–NaCl, (20% PEG-8000 (w/v), 2.5 M NaCl), incubating for 5 min at 4 °C and centrifuging at 20 064 ***g*** at 4 °C for 20 min. The phage pellet was resuspended in 20 mL PBT [phosphate buffered saline (PBS), 0.05% Tween-20, 0.2% BSA], insoluble debris was removed by centrifugation and the library was stored at −80 °C. The library was reamplified in *E. coli* SS320 cells in the presence of 0.3 mM IPTG. The composition of the naïve phage library was examined by Illumina sequencing.

Protein expression and purification for phage display selections

The expression constructs of DLG1 PDZ2 and SHANK1 PDZ in a pGEX vector were described previously [17]. The synthetic coding genes of the other bait proteins were generously provided by the Sidhu lab cloned in an in-house made vector (pHH0103), which carries ampicillin resistance and encodes 6-His-GST-tagged proteins. About 10 ng DNA of each domain was used to transform into chemically competent *E. coli* BL21 (DE3). Bacteria were grown over night with shaking in 440 μL 2-YT media supplemented with carbenicillin (30 $\mu g \cdot mL^{-1}$) in a 96-well format. Ten microliters of overnight cultures were used to inoculate 2 × 1.5 mL autoinducing Magic Medium (Invitrogen, Carlsbad, CA, USA) supplemented with carbenicillin (30 $\mu g \cdot mL^{-1}$) in a 96-well format deep well block. The cultures were grown at 37 °C for 6 h with 200 r.p.m. shaking. The temperature was then reduced to 20 °C and protein expression was allowed for 24 h. The bacteria were pelleted by centrifugation (15 min, 3400 ***g***) and purified in 96-well format as described by Huang and Sidhu, using a glutathione sepharose resin for purification of the PDZ domains, and a Ni-affinity resin for the other proteins [54]. Purified proteins were confirmed by SDS/PAGE analysis and the protein concentrations were estimated by Bradford assay. Freshly purified proteins were used for peptide phage selections.

Protein expression and purification for affinity measurements

Untagged GGA1 VHS domain was prepared for affinity measurements. The coding region (amino acids 7–157) of GGA1 VHS was cloned into the pETM-11 vector (EMBL, Heidelberg, Germany) using the *Nco*I and *Eco*RI restriction sites and was transformed into chemically competent *E. coli* BL21 (DE3) gold cells. The protein was expressed in 2TY-medium supplemented with kanamycin. The overnight culture was diluted 1 : 100 and grown in 37 °C until $OD_{600\ nm} \approx 0.7$. Protein expression was induced by addition of 1 mM IPTG final concentration, the temperature was decreased to 30 °C and cells were grown for 3 h and then harvested by centrifugation at 8000 ***g*** for 10 min. Protein was batch purified by Ni-nitrilotriacetic acid affinity chromatography. Bacteria pellet were resuspended in Lysis buffer consisting of PBS pH 7.2, 20 mM imidazole, 1% Triton X-100 (PerkinElmer, Waltham, MA, USA), 1 Complete mini protease inhibitors tablet (EDTA free; Roche, Basel, Switzerland), 2.5 unit DNase1, 6 μg lysozyme per 50 mL of buffer and incubated for 30 min at 4 °C under gentle agitation. Cell lysates were cleared by centrifugation at 13 000 ***g*** for 45 min. Lysates were mixed with Ni-nitrilotriacetic acid agarose pre-equilibrated with binding buffer (20 mM imidazole in PBS) and incubated 30 min. The resin was washed using 75 mM imidazole in PBS until A_{280} reached a value below 0.05. Bound proteins were subsequently eluted with 300 mM imidazole in PBS. Following Ni-nitrilotriacetic acid affinity purification, 1 mg of TEV protease was added per 50 mg protein to remove the His-tag and from the VHS domain. To remove the cleaved His-tag and the His-tagged protease, a reverse Ni-nitrilotriacetic acid chromatography was performed in the same buffer. The GGA1 VHS domain was dialyzed into PBS pH 8, 1 mM β-mercaptoethanol. All purification steps were carried out at 4 °C.

Phage selections

The selections were carried out following a published high-throughput selection protocol [54] with minor modifications. The proteins (5–10 μg in 100 μL PBS) were coated in 96-well Flat-bottom Immuno Maxisorp plates (Nunc, Roskilde, Denmark) overnight at 4 °C. In parallel, GST was

plated in a preselection plate. The Maxisorp plates were blocked with 0.5% BSA in PBS. The naïve phage library (~ 10^{12} phage particles in each well) was added to the preselection plate for 1 h, transferred to the target proteins and were allowed to bind for 2 h. Unbound phages were removed by five times washing with cold wash buffer (PBS, 0.5% Tween-20) and bound phage was eluted by direct infection into bacteria by the addition of 100 μL of log phase (A_{600} = 0.8) *E. coli* SS320 in 2YT to each well and incubation for 30 min at 37 °C with shaking. M13K07 helper phage (NEB, Ipswich, MA, USA) was added to a final concentration of 10^{10} phage·mL^{-1} to enable phage production, and the cultures were incubated for 45 min at 37 °C with shaking. Eluted phages were amplified overnight in 1.5 mL 2YT supplemented with antibiotics (carbencillin and kanamycin). Bacteria were then pelleted by centrifugation, the supernatant was heat inactivated at 65 °C for 15 min, shilled on ice and then used for the next round of selections. Five rounds of phage panning were conducted and the selections were followed by pooled phage enzyme-linked immunosorbent assays, which suggested that the selections were saturated after 4 days of selections.

Phage pools of round four were barcoded for NGS on the Illumina platform as outlined by McLaughlin and Sidhu [55]. Undiluted amplified phage pools (5 μL) were used as templates for 24 cycles 50 μL PCR reactions using unique combinations of barcoded primers for each reaction (0.5 μM each, for barcode sequences see [55]) and Phusion High Fidelity DNA polymerase (NEB) with a maximum polymerase and primer concentrations. The PCR products were confirmed by gel electrophoresis (2% agarose gel) of 1 μL PCR products. The concentrations of the PCR products were estimated using PicoGreen dye (Invitrogen) and using a two-fold dilution series (100–0.8 μg·μL^{-1}) of lambda phage double-stranded DNA (dsDNA; Invitrogen) as a standard. The PicoGreen dye was diluted 1 : 400 in TE buffer and mixed with 1 μL of dsDNA standard or PCR product in a low-fluorescence 96-well plate (Bio-Rad, Hercules, CA, USA). The plate was briefly centrifuged before reading the fluorescence in a qPCR machine (Bio-Rad; excitation 480 nm, emission 520 nm). The blank value was subtracted and the DNA concentration of the sample determined from the standard curve.

Equal amounts of each PCR products were pooled. The PCR amplicons (~ 3 μg) was sent to Cofactor Genomics (St. Louis, MO, USA) for NGS (Illumina Miseq, Toronto, Canada, paired end 150 base reads, 20% PhiX) In total, 124 890 sequencing reads are identified. The obtained sequencing reads were filtered for average Phred score of 30 (99.9% of sequencing accuracy) in peptide region (48 nucleotide variable sequence). Each read was trimmed at both 5′- and 3′- end to remove uninformative sequence (adapter, barcode and constant region). The variable sequencing reads were mapped against the library design using Bowtie [56] with maximum two mismatches between the reference sequences in the library design and the sequencing reads.

Assignment of cutoff values and data analysis

Threshold values were established for each protein individually to filter out nonspecifically retained peptides (Table S1). Consensus motifs among the ligands above the threshold were identified by motif scanning using the SLiMFinder algorithm using default settings [25] (Table 1). Ligands with counts below cutoff values and lacking consensus motifs were used to create a merged set of background 'nonspecific' peptides. Some of these nonspecific peptides occur multiple times in unrelated datasets. With continued use of the disorderome library we expect that the set of nonspecific ligands will grow and be consolidated, and that the information can be used to remove nonspecific ligands, in analogy to the frequent flyer analysis in MS studies [57]. Identified ligands were cross-referenced with motif annotations from the ELM database; motif annotations and mutagenesis data from UniProt and the published motif literature to find previously described motifs. To establish the presence of known binders for the bait protein among the ProP-PD-derived ligands, we retrieved previously identified interactors from the HIPPIE database [41]. A GO term enrichment analysis was performed using the web-based tool DAVID 6.8 [42] with the default human proteome as a background, the GO FAT and the KEGG pathway annotations [58] and medium stringency setting. The significance of the enrichments were evaluated by the Benjamini–Hochberg false discovery rate corrected *P*-values, with a cutoff value of FDR ≤ 0.01. GO terms shared by the bait protein and proteins containing the identified peptide ligands were tested for significance by calculating the likelihood of any two proteins sharing the term by chance using the equation $(N_g \times (N_g - 1))/(N \times N - 1)$ where N_g is the number of proteins with the GO term and N is the number of proteins in human protein. Probabilities were corrected for multiple testing and significance was evaluated at a cutoff value of $P \leq 0.01$. In addition, we performed a GO term enrichment analysis using GORILLA [43], using the library design minus identified ligands as a background set. Network of the high-confidence PPIs were visualized using CYTOSCAPE 3.1 [59].

Isothermal titration calorimetry

Isothermal titration calorimetry experiments were performed using a MicroCal PEAQ-ITC (Malvern Instruments, Northampton, MA, USA) at 25 °C in PBS, pH 8 with 1 mM β-mercaptoethanol. The protein concentration in the calorimeter cell ranged from 70 to 120 μM, and the concentration of ligand in the syringe ranged from 1.5 to 3.5 mM. A total of 40 μL of ligand (in 2.49 μL aliquots) was injected over the course of each titration. Data from the first (0.4 μL) injection were discarded to eliminate

diffusion-related artifacts. Data were fit to a single-site binding model in the software package provided with PEAQ–ITC (Fig. 4, Table 4). Protein active concentration along with the binding stoichiometry was established in the titrations with the tightest binder. For the rest of the measurements the stoichiometry value was fixed to 1.

Coimmunoprecipitation

The choice of constructs were based on their availability of constructs in the Openfreezer [60]. HEK293T cells were transfected with expression vectors for FLAG-tagged target protein and HA-tagged GGA1 protein. Cells were lysed 48 h after transfections with radioimmunoprecipitation assay buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40, 1× protease/phosphatase inhibitor cocktail (Cell Signaling #5872, Danvers, MA, USA)) for 30 min at 4 °C and spun down at 16 000 ***g*** for 10 min. Cell lysates were coimmunoprecipitated with anti-FLAG beads (Sigma). Protein samples were loaded on a mini PROTEAN TGX precast 4–15% SDS/PAGE gel (Bio-Rad) and transferred to PVDF (Polyvinylidene difluoride, 0.2 μm) membranes. Transferred HA-tagged GGA1 was immunoblotted with primary rabbit anti-HA antibody (Invitrogen) followed by horseradish peroxidase (HRP)-conjugated secondary antibodies (Cell Signaling #7074) and FLAG-tagged target proteins were immunoblotted with HRP-conjugated anti-FLAG antibody (Gene Script). The proteins were detected using chemiluminescence substrate (Thermo Scientific #34080, Burlington, ON, Canada).

Acknowledgements

This work was supported by a SFI Starting Investigator Research Grant (13/SIRG/2193) for NED. YI received grants from the Swedish Research Council (C0509201) and from the Carl Trygger foundation (CTS14:209). PMK acknowledges support from a Canadian Health Research Institute Operating Grant (CIHR MOP-123526) and an Natural Sciences and Engineering Research Council Discovery Grant (NSERC #386671). The phagemids used as starting template for library construction expression constructs of the peptide-binding domains were generously provided by Prof Sachdev S. Sidhu, The Donnelly Centre, University of Toronto. We thank Martha Cyert for critical reading and valuable comments on the manuscript. We thank our colleagues for critically reading the manuscript.

Conflict of interest

A part of this work (i.e., ITC measurements) was carried out in the demolab of Malvern Instruments Nordic that sells MicroCal PEAQ-ITC instruments.

Author contributions

NED, PMK, and YI conceived experiments. NED designed the phage library and annotated hits. MHS performed co-IPs. SN and DD created the ProP-PD library. YI performed phage selections and analyzed data. VY and CB subcloned constructs and performed ITC under guidance of NM. JJ analyzed NGS data. IK wrote the peptide annotation software. NED and YI wrote the manuscript. All authors commented on the text.

References

1 Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R *et al.* (2014) A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226.
2 Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K *et al.* (2015) The BioPlex network: a systematic exploration of the human interactome. *Cell* **162**, 425–440.
3 Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, Gibson TJ & Davey NE (2014) Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem Rev* **114**, 6733–6778.
4 Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H & Gibson TJ (2012) Attributes of short linear motifs. *Mol BioSyst* **8**, 268–281.
5 Pawson T & Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445–452.
6 Davey NE, Cyert MS & Moses AM (2015) Short linear motifs – ex nihilo evolution of protein regulation. *Cell Commun Signal* **13**, 43.
7 Goldman A, Roy J, Bodenmiller B, Wanka S, Landry CR, Aebersold R & Cyert MS (2014) The calcineurin signaling network evolves via conserved kinase-phosphatase modules that transcend substrate identity. *Mol Cell* **55**, 422–435.
8 Uyar B, Weatheritt RJ, Dinkel H, Davey NE & Gibson TJ (2014) Proteome-wide analysis of human disease mutations in short linear motifs: neglected players in cancer? *Mol BioSyst* **10**, 2626–2642.
9 Davey NE, Trave G & Gibson TJ (2011) How viruses hijack cell regulation. *Trends Biochem Sci* **36**, 159–169.
10 Via A, Uyar B, Brun C & Zanzoni A (2015) How pathogens use linear motifs to perturb host cell networks. *Trends Biochem Sci* **40**, 36–48.
11 Chemes LB, de Prat-Gay G & Sanchez IE (2015) Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. *Curr Opin Struct Biol* **32**, 91–101.

12 Fukuchi S, Hosoda K, Homma K, Gojobori T & Nishikawa K (2011) Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct Biol* **11**, 29.
13 Dunker AK, Silman I, Uversky VN & Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* **18**, 756–764.
14 Tompa P, Davey NE, Gibson TJ & Babu MM (2014) A million peptide motifs for the molecular biologist. *Mol Cell* **55**, 161–169.
15 Gibson TJ, Dinkel H, Van Roey K & Diella F (2015) Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun Signal* **13**, 42.
16 Dinkel H, Van Roey K, Michael S, Kumar M, Uyar B, Altenberg B, Milchevskaya V, Schneider M, Kuhn H, Behrendt A *et al.* (2016) ELM 2016–data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res* **44**, D294–D300.
17 Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y *et al.* (2008) A specificity map for the PDZ domain family. *PLoS Biol* **6**, e239.
18 Luck K & Trave G (2011) Phage display can select over-hydrophobic sequences that may impair prediction of natural domain-peptide interactions. *Bioinformatics* **27**, 899–902.
19 Blikstad C & Ivarsson Y (2015) High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Commun Signal* **13**, 38.
20 Ivarsson Y, Arnold R, McLaughlin M, Nim S, Joshi R, Ray D, Liu B, Teyra J, Pawson T, Moffat J *et al.* (2014) Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc Natl Acad Sci USA* **111**, 2542–2547.
21 Sundell GN & Ivarsson Y (2014) Interaction analysis through proteomic phage display. *BioMed Res Int* **2014**, 176172.
22 Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MA, Church GM, Kesari S, Leproust EM, Solimini NL *et al.* (2011) Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol* **29**, 535–541.
23 Garrido-Urbani S, Garg P, Ghossoub R, Arnold R, Lembo F, Sundell GN, Kim PM, Lopez M, Zimmermann P, Sidhu SS *et al.* (2016) Proteomic peptide phage display uncovers novel interactions of the PDZ1-2 supramodule of syntenin. *FEBS Lett* **590**, 3–12.
24 Ivarsson Y (2012) Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett* **586**, 2638–2647.
25 Edwards RJ, Davey NE & Shields DC (2007) SLiMFinder: a probabilistic method for identifying over-represented, convergently evolved, short linear motifs in proteins. *PLoS One* **2**, e967.
26 Dinkel H, Michael S, Weatheritt RJ, Davey NE, Van Roey K, Altenberg B, Toedt G, Uyar B, Seiler M, Budd A *et al.* (2012) ELM–the database of eukaryotic linear motifs. *Nucleic Acids Res* **40**, D242–D251.
27 Ball LJ, Jarchau T, Oschkinat H & Walter U (2002) EVH1 domains: structure, function and interactions. *FEBS Lett* **513**, 45–52.
28 Kofler MM & Freund C (2006) The GYF domain. *FEBS J* **273**, 245–256.
29 Westin S, Kurokawa R, Nolte RT, Wisely GB, McInerney EM, Rose DW, Milburn MV, Rosenfeld MG & Glass CK (1998) Interactions controlling the assembly of nuclear-receptor heterodimers and co-activators. *Nature* **395**, 199–202.
30 Shiba T, Takatsu H, Nogi T, Matsugaki N, Kawasaki M, Igarashi N, Suzuki M, Kato R, Earnest T, Nakayama K *et al.* (2002) Structural basis for recognition of acidic-cluster dileucine sequence by GGA1. *Nature* **415**, 937–941.
31 Lee JH, Park H, Park SJ, Kim HJ & Eom SH (2011) The structural flexibility of the shank1 PDZ domain is important for its binding to different ligands. *Biochem Biophys Res Commun* **407**, 207–212.
32 Uemura T, Mori H & Mishina M (2004) Direct interaction of GluRdelta2 with Shank scaffold proteins in cerebellar Purkinje cells. *Mol Cell Neurosci* **26**, 330–341.
33 Van Roey K, Gibson TJ & Davey NE (2012) Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol* **22**, 378–385.
34 Boukhelifa M, Parast MM, Bear JE, Gertler FB & Otey CA (2004) Palladin is a novel binding partner for Ena/VASP family members. *Cell Motil Cytoskelet* **58**, 17–29.
35 Heery DM, Kalkhoven E, Hoare S & Parker MG (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387**, 733–736.
36 Le Douarin B, Nielsen AL, Garnier JM, Ichinose H, Jeanmougin F, Losson R & Chambon P (1996) A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J* **15**, 6701–6715.
37 Heery DM, Hoare S, Hussain S, Parker MG & Sheppard H (2001) Core LXXLL motif sequences in CREB-binding protein, SRC1, and RIP140 define affinity and selectivity for steroid and retinoid receptors. *J Biol Chem* **276**, 6695–6702.
38 Cramer JF, Gustafsen C, Behrens MA, Oliveira CL, Pedersen JS, Madsen P, Petersen CM & Thirup SS (2010) GGA autoinhibition revisited. *Traffic* **11**, 259–273.
39 Jacobsen L, Madsen P, Nielsen MS, Geraerts WP, Gliemann J, Smit AB & Petersen CM (2002) The

sorLA cytoplasmic domain interacts with GGA1 and -2 and defines minimum requirements for GGA binding. *FEBS Lett* **511**, 155–158.
40 Christopherson KS, Hillier BJ, Lim WA & Bredt DS (1999) PSD-95 assembles a ternary complex with the N-methyl-D-aspartic acid receptor and a bivalent neuronal NO synthase PDZ domain. *J Biol Chem* **274**, 27467–27473.
41 Schaefer MH, Fontaine JF, Vinayagam A, Porras P, Wanker EE & Andrade-Navarro MA (2012) HIPPIE: integrating protein interaction networks with experiment based quality scores. *PLoS One* **7**, e31826.
42 da Huang W, Sherman BT & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57.
43 Eden E, Navon R, Steinfeld I, Lipson D & Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.
44 Puertollano R, van der Wel NN, Greene LE, Eisenberg E, Peters PJ & Bonifacino JS (2003) Morphology and dynamics of clathrin/GGA1-coated carriers budding from the trans-Golgi network. *Mol Biol Cell* **14**, 1545–1557.
45 Shiba T, Kametaka S, Kawasaki M, Shibata M, Waguri S, Uchiyama Y & Wakatsuki S (2004) Insights into the phosphoregulation of beta-secretase sorting signal by the VHS domain of GGA1. *Traffic* **5**, 437–448.
46 Doray B, Misra S, Qian Y, Brett TJ & Kornfeld S (2012) Do GGA adaptors bind internal DXXLL motifs? *Traffic* **13**, 1315–1325.
47 Van Roey K & Davey NE (2015) Motif co-regulation and co-operativity are common mechanisms in transcriptional, post-transcriptional and post-translational regulation. *Cell Commun Signal* **13**, 45.
48 Scott JD & Pawson T (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224.
49 Yaffe MB (2002) Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol* **3**, 177–186.
50 Reinhardt HC & Yaffe MB (2013) Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response. *Nat Rev Mol Cell Biol* **14**, 563–580.
51 Dente L, Vetriani C, Zucconi A, Pelicci G, Lanfrancone L, Pelicci PG & Cesareni G (1997) Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J Mol Biol* **269**, 694–703.
52 Rajan S & Sidhu SS (2012) Simplified synthetic antibody libraries. *Methods Enzymol* **502**, 3–23.
53 Sidhu SS (2000) Phage display in pharmaceutical biotechnology. *Curr Opin Biotechnol* **11**, 610–616.
54 Huang H & Sidhu SS (2011) Studying binding specificities of peptide recognition modules by high-throughput phage display selections. *Methods Mol Biol* **781**, 87–97.
55 McLaughlin ME & Sidhu SS (2013) Engineering and analysis of Peptide-recognition domain specificities by phage display and deep sequencing. *Methods Enzymol* **523**, 327–349.
56 Langmead B, Trapnell C, Pop M & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.
57 Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY *et al.* (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* **10**, 730–736.
58 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29.
59 Lopes CT, Franz M, Kazi F, Donaldson SL, Morris Q & Bader GD (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26**, 2347–2348.
60 Olhovsky M, Williton K, Dai AY, Pasculescu A, Lee JP, Goudreault M, Wells CD, Park JG, Gingras AC, Linding R *et al.* (2011) OpenFreezer: a reagent information management software system. *Nat Methods* **8**, 612–613.
61 Kofler M, Motzny K & Freund C (2005) GYF domain proteomics reveals interaction sites in known and novel target proteins. *Mol Cell Proteomics* **4**, 1797–1811.
62 Plevin MJ, Mills MM & Ikura M (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci* **30**, 66–69.
63 Bonifacino JS (2004) The GGA proteins: adaptors on the move. *Nat Rev Mol Cell Biol* **5**, 23–32.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Annotated ligands obtained from NGS analysis of enriched phage pools.

Table S2. DAVID 6.8 GO term enrichment analysis of ProP-PD-derived ligands.

Table S3. GORILLA GO term enrichment analysis of ProP-PD-derived ligands.

Table S4. Annotated set of high-confidence ligands.