Article

Functional Classification and Interaction Selectivity Landscape of the Human SH3 Domain Superfamily

Neda S. Jasemi Kazemein , Mehrnaz Mehrabipour , Eva Magdalena Estirado , Luc Brunsveld ,
Radovan Dvorsky , Mohammad R. Ahmadian [*]

Article

Functional Classification and Interaction Selectivity Landscape of the Human SH3 Domain Superfamily

Neda S. Kazemein Jasemi [1,§]**, Mehrnaz Mehrabipour** [1,§]**, Eva Magdalena Estirado** [2]**, Luc Brunsveld** [2]**, Radovan Dvorsky** [1] **and Mohammad R. Ahmadian** [1,*]

[1] Institute of Biochemistry and Molecular Biology II, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; mehrnaz.mehrabipour@hhu.de (M.M.); neda.jasemi@hhu.de (N.S.K.J.)

[2] Laboratory of Chemical Biology, Department of Biomedical Engineering and Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, PO Box 513, 5600MB Eindhoven, The Netherlands.

[*] Correspondence: reza.ahmadian@hhu.de (M.R.A.); Tel.: +49-2118112384 (M.R.A.)

[§] These authors contributed equally to this work.

Abstract: SRC homology 3 (SH3) domains are critical interaction modules that orchestrate the assembly of protein complexes involved in diverse biological processes. They facilitate transient protein-protein interactions by selectively interacting with proline-rich motifs (PRMs). A database search revealed 298 SH3 domains in 221 human proteins. Multiple sequence alignment of human SH3 domains is useful for phylogenetic analysis and determination of their selectivity towards PRM-containing peptides (PRPs). However, a more precise functional classification of SH3 domains is achieved by constructing a phylogenetic tree only from PRM-binding residues and using existing SH3-PRP structures and biochemical data to determine the specificity within each of the 10 families for particular PRPs. In addition, the C-terminal proline-rich domain of the RAS activator SOS1 covers 13 of the 14 recognized proline-rich consensus sequence motifs, encompassing differential PRP pattern selectivity among all SH3 families. To evaluate binding capabilities and affinities, we conducted fluorescence dot blot and polarization experiments using 25 representative SH3 domains and various PRPs derived from SOS1. Our analysis has identified 45 interacting pairs with binding affinities ranging from 0.2 to 125 micromolar out of 300 tested and potential new interactors of SOS1. Furthermore, it establishes a framework to bridge the gap between SH3 and PRP interactions and provides predictive insights into potential interactions of SH3 domains with PRMs based on sequence specifications. This novel framework has the potential to enhance the understanding of protein networks mediated by SH3-PRM interactions and be utilized as a general approach for other domain-peptide interactions.

Keywords: ARHGAP12; GRB2; NCK1; proline-rich motifs; protein-protein interaction; SH3 domain; signal transduction; SOS1; SRC homology 3; WRCH1/RHOU

1. Introduction

Protein-protein interactions are fundamental to the intricate machinery that controls virtually all biological processes. [1]. Among the diverse array of protein domains that facilitate these interactions, the SRC homology 3 (SH3) domains stand out as central modular units. These compact domains, consisting of approximately 60 amino acids with similar sequences that adopt a compact β-barrel fold made of five β-strands [2], are found predominantly in various signaling proteins and various protein families [3,4]. The selective interactions of the SH3 domain with proline-rich motifs (PRMs) are fundamental for the assembly and orchestration of multiprotein complexes [5]. It is rational that the SH3 domain-containing proteins (SH3DCPs) are involved in a wide variety of biological processes [4], subsequently leading to a substantial influence on a spectrum of diseases, such as cancers [6], neurological disorders [7], kidney and urinary disorders [8,9], muscle and myopathy disorders [10,11], immune disorders [12,13], genetic and developmental disorders [4,14,15].

To date, a series of seven types of PRM binding modules have been reported, including SH3, WW (two highly conserved tryptophan amino acids), EVH1 (Ena/VASP homology domain 1), GYF (glycine-tyrosine-phenylalanine), Profilin, CAP-Gly (cytoskeleton-associated protein-glycine-rich), and UEV (ubiquitin E2 variant) [16,17]. Proline-rich target peptides possess core PRMs with unique properties that influence interaction selectivity. The diversity of PRMs results from the inclusion of one or more proline residues in various combinations within the peptide sequences. The proline side chains and carbonyl groups are exposed at regular intervals, allowing intermolecular hydrogen bonding with PRM-binding domains [16]. Interactions involving PRMs have a low entropic cost of binding due to the restricted rotational freedom of proline residues along the peptide backbone. This restricted flexibility contributes to a higher overall binding energy for complexes involving PRM-containing peptides (or PRPs). This property increases the affinity of PRM interactions and influences ligand recognition [16]. In addition, PRMs can interact with their binding partners in two distinct orientations, influenced by the arrangement of non-proline residues either located N- or C-terminal to the core motif, often involving positively charged counterparts (R or K) [16,18].

In the *context of* SH3 domain interaction with PRMs, a specific contact recognition occurs, where a positively charged PRM residue binds to conserved negatively charged residues in the variable loop region of the SH3 domain, resulting in moderate selectivity and affinity [5,16]. Comparative analysis of SH3 domain binding sites reveals remarkable variability and flexibility in the loop regions, contributing to the specificity and affinity of PRM binding [6,19]. The preference of the SH3 domain for specific PRMs is usually moderate, with affinities typically in the low micromolar range [20–22]. Furthermore, SH3 domains exhibit a broad spectrum of both conventional (PRM-based) and nonconventional selectivity, effectively recognizing a diverse array of protein interactors in a differentiated manner [4,5].

SH3DCPs play a pivotal role in biological processes by facilitating diverse protein-protein interactions through SH3-PRMs, relying on their selectivity and affinity. The intricate nature of protein assembly orchestrated by SH3 domains raises significant questions regarding the underlying selectivity framework governing complex networks of SH3-PRMs interactions [23]. Despite sharing a 25% sequence homology, accurate prediction of the selective PRM recognition by SH3 domains remains a formidable challenge [21,24,25]. In this study, we analyzed the phylogenetic and structure-function relationships of all 298 human SH3 domains, specifically focusing on the sequences of their PRM binding sites. We then performed classification based on their PRM-binding selectivity, organizing them into 10 distinct families. In addition, the distinctive recognition pattern of PRMs within SOS1, a well-established PRM-containing protein, caught our attention. This pattern of selectivity across all SH3 families led us to use SOS1 as a comprehensive model to elucidate the recognition mechanisms of established SH3 domains within the human proteome in our research. The binding capabilities and affinities of 25 representative SH3 domains toward 10 SOS1-derived PRPs and two reference peptides were carefully evaluated. Our investigation using fluorescence dot blot and polarization techniques revealed a significant finding: out of 300 SH3-peptide combinations, only 45 exhibited binding affinities ranging from 0.2 to 125 micromolar. This study pioneers the understanding of the selectivity and affinity of SH3 protein modules for specific PRMs, encompassing a wide range of proteins. This framework lays the foundation for a predictive matrix that enables the anticipation of SH3-PRM-mediated protein-protein interactions within complex cell signaling networks.

2. Materials and Methods

2.1. Bioinformatics, databases, and structural analysis

The Sequences of the SH3DCPs were obtained from the UniProt database by combining full-text searches and sequence homology searches performed with the HMMER software package. Isolated sequences of the SH3 domains were then extracted from the previously obtained proteins, again using the subprograms of HMMER. Alignment of the SH3 domain sequences was then performed using BioEdit 7.2.5 software, and the resulting phylogenetic tree was constructed using MEGA 10.2.6 software.

Available structures containing SH3 domains were retrieved from the Protein Data Bank (PDB) website using the BLAST program. To further analyze the SH3 domain structure and define its binding residues for interaction with PRMs, Python scripts were used to identify all residues in SH3 structures within 4.0 Å of the PRM bound to it. The information thus obtained was then projected onto the global sequence alignment of all SH3 domains, and only homologous residues potentially contacting PRMs were selected to define the alignment of PRM-binding residues of SH3 domains. Finally, PRMs were collected from published articles available on the NCBI website. In addition, BLAST analysis of SOS1 PRMs was performed on the NCBI platform.

2.2. Constructs, peptides, and proteins

The constructs and peptides employed in our study are listed in Tables S1 and S2, respectively. All fluorescein-labeled PRPs are synthesized and used under the conditions described previously [22]. p3XFLAG-CMV ARHGAP12^{wt} [26] was used to generate a SH3 domain deletion (ARHGAP12$^{\Delta SH3}$). NCK1^{wt}, NCK1$^{\Delta SH3-3}$, and NCK1^{Set-1} (N205D, D206T, D226Q, and P227D) were ordered in pcDNA3.0-Flag vector from BioCat GmbH. HA-SOS1 in the pCGN vector was ordered from addgene (#32920). All SH3 domains of proteins listed in Table S1 within the pGEX4-T1 vector were initially purified as GST-tagged fusion proteins. The purification process involved affinity chromatography on a glutathione Sepharose column [22]. For subsequent polarization analysis, a portion of these GST fusion proteins underwent cleavage of the GST tag using a proteolytic enzyme. Following this cleavage step, the proteins were subjected to further purification and separation by employing size exclusion. All purified proteins underwent analysis via SDS-PAGE and were subsequently stored at -80°C for further use.

2.3. Pull-down and fluorescence dot blot analysis

Pull-down of 10 μM FITC-labeled peptides (Table S2) with 5 μM purified GST-SH3 domains (Table S1) was performed using 10 μl glutathione Sepharose beads (GE Healthcare, UK) in a buffer containing 30 mM Tris-HCl at pH 7.5, 3 mM dithiothreitol, and 5 mM MgCl2 for 1 h at 4°C. Purified GST was used as a negative control. After three washes, bound proteins were eluted by incubation in the same buffer containing 20 mM reduced glutathione for 15 min at 4°C, and the beads were separated by centrifugation. Bound FITC-labeled peptides were detected by dot blot analysis using 1 μl of eluent at an emission wavelength of 600 nm and an Odyssey Fc imaging system (LI-COR Biosciences). Detected signals were quantified densitometrically using LI-COR Biosciences Image Studio version 5.2 imaging software.

2.4. Fluorescence polarization

The interaction between fluorescein-labeled proline-rich peptides (0.2 μM) and increasing concentrations of SH3 domains (ranging from 0 to 200 μM) was evaluated in a buffer, containing 30 mM Tris/HCl (pH 7.5), 5 mM MgCl$_2$, and 3 mM DTT at 25°C using a Fluoromax 4 fluorimeter in polarization mode. Excitation was performed at 470 nm and emission was measured at 560 nm. Dissociation constants (K$_d$) were determined by fitting the concentration-dependent binding curve to a quadratic ligand binding equation.

2.5. Trandfection and immunoprecipitation analysis

CHO-K1 cells were cultured in Dulbecco's modified Eagle's serum (DMEM, Gibco) supplemented with 10% FBS (Gibco) and 1% penicillin/streptomycin (Genaxxon). Cells were co-transfected with HA-tagged SOS1 full-length (FL) and FLAG-tagged NCK1^{wt}, NCK1$^{\Delta SH3-3}$, and NCK1^{Set-1} or FLAG-tagged ARHGAP12^{wt} ARHGAP12$^{\Delta SH3}$, and ARHGAP12^{Set-1} using a Turbofect reagent (Thermo Fisher Scientific). To perform co-immunoprecipitation assays, the CHO-K1 cells were lysed on ice for 5 min employing a buffer, containing 50 mM Tris/HCl (pH 7.4), 100 mM NaCl, 2 mM MgCl$_2$, 10% glycerol, 20 mM ß-glycerophosphate, 1 mM Na$_3$VO$_4$, 1% IGPAL (ThermoFisher Scientific), and 1x protease inhibitor cocktail (Roche). Lysates were cleared by centrifugation (20.000 x g at 4°C for 5 min). Protein concentrations were determined using the Bradford assay. Lysates were then incubated overnight at 4°C with either anti-Flag M2 agarose beads (Sigma Aldrich) or Protein A-Sepharose beads with anti-Flag antibody and anti-IgG as control. The beads were then washed three times with a wash buffer containing 50 mM Tris-HCl and 150 mM NaCl with 1 mM EDTA. Proteins bound to the beads were eluted with 2.5x Laemmli loading buffer and subjected to SDS-PAGE for further analysis. Primary antibodies used in Western blot analysis include anti-GST (own antibody), anti-Flag (1:1000 WB and 1:50 CO-IP, #F742; and #F3165, both from Sigma), anti-IgG (1:50; # sc-2025, Santa Cruz), anti-SOS1 (1:1000; #sc-256, Santa

Cruz), and anti-Vinculin (1:1000; #V9131, Sigma). The secondary antibodies used were purchased from LI-COR (anti-mouse 700 nm: IRDye #926-32213; anti-rabbit 800 nm: IRDye #926-6807).

3. Results

3.1. Sequence-structure-function classification of human SH3 domains

In our previous study, we performed a comprehensive survey in which we identified 298 SH3 domains within 221 SH3DCPs spanning a range of 13- to 720-kilodalton proteins [4]. This analysis included a phylogenetic assessment of human SH3DCPs based on their multidomain architecture, providing a convenient functional classification within different physiological pathways. However, this approach did not address the intrinsic PRM selectivity of the SH3 domain itself. Therefore, we set out to comparatively study the sequence-structure-function relationships of human SH3 domains with a focus on three key aspects: the amino acid sequence, three-dimensional (3D) structure, and spatial arrangement of PRM-binding sites combining bioinformatics with experimental and structural biology.

As a first step, we focused on elucidating the critical aspects of the PRM binding properties of the SH3 domain. To achieve this, we obtained primary sequences covering a collection of 298 human SH3 domains from the UniProt database. These sequences were then aligned and used to construct a phylogenetic tree of the SH3 domain superfamily using MEGA software (version 10.2.6). The resulting phylogenetic tree, designated tree #1 (Figure S1), depicted the evolutionary relationships among human SH3 domains. Examination of this tree revealed a remarkable conservation of key regions essential for the 3D structure of SH3 domains (Figure S2). This finding underscored a robust and consistent sequence-structure relationship spanning specific parts of the SH3 domain responsible for PRP interactions in SH3DCP families [4]. However, a more complex scenario emerged when delving into the comprehensive analysis of SH3-PRP interactions. Despite meticulous exploration of SH3-PRP structures available in the protein database (Table S3) and a comprehensive review of published biochemical data on SH3-PRP interactions (Table S4), no discernible structure-function relationship was revealed from tree #1. Strikingly, PRPs exhibit clustering patterns that are inconsistent with established SH3 domain families. Instead, they were distributed among distantly related families (Figure S1).

A different approach to characterize the SH3-PRM interaction was to perform a second phylogenetic analysis focusing only on the active site regions of the SH3 domain, specifically the residues involved in PRM binding. Unlike the methodology for tree #1, which considered complete SH3 domain sequences, the second phylogenetic tree (tree #2) was constructed exclusively from PRM-binding residues potentially involved in PRM interactions. The PRM-binding residues within the SH3 domains were inferred from the structures of the SH3 domains in complex with their specific PRPs (Table S3). It is noteworthy that tree #2 shows 10 different families of SH3 domains that can be very well assigned to their respective PRPs (Figures 1, S3, and S4). This strategic approach now allowed us to explore the intricate associations between individual SH3 domains and specific PRMs, as discussed in the following sections.
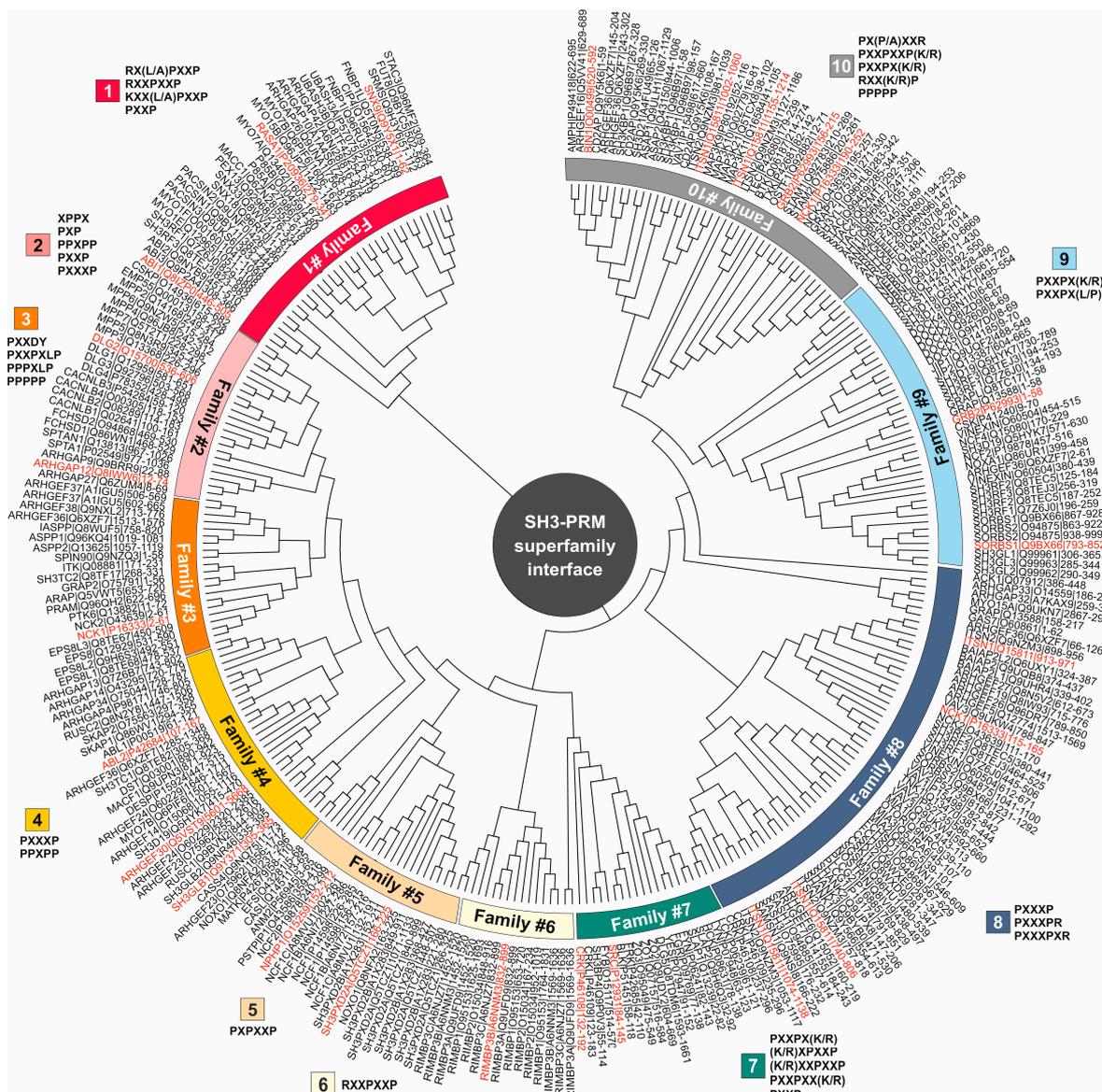
Figure 1. Evolutionary relationships among PRM-binding residues within SH3 domains. The phylogenetic tree (tree #2) was constructed by collecting PRM-binding residues from 298 human SH3 domains, using published SH3-PRM structures (Table S3) and biochemical data (Table S4), and utilizing the MEGA7 software. SH3 domains were systematically classified into ten distinct families based on their interaction properties with specific PRMs as indicated by the color codes. SH3 domains highlighted in red were selected as representatives for further analysis in this study. The protein name is accompanied by the corresponding Uniprot ID and residue span for the SH3 domain.

3.2. PRM selectivities of different SH3 families

Classification of PRMs into distinct families based on published structural and biochemical data (Tables S3 and S4) reveals specific sequence patterns that guide protein-protein interactions (Figure S5). Family 1 includes motifs such as RX(L/A)PXXP, RXXPXXP, KXX(L/A)PXXP, and PXXP, suggesting a diverse yet structured arrangement for interaction. Family 2, characterized by a PPXPPXP consensus, shows patterns such as XPPX, PXP, PPXPP, PXXP, and PXXXP, indicating a diverse but consistent motif profile. Family 3 sequences, including PXXDY, PXXPXLP, PPPXLP, and PPPPP, show specificity around proline and other residues such as D and Y. Family 4 follows a PXXXPPXPP consensus with specific motifs such as PXXXP and PPXPP. Family 5 has a specific PXPXXP motif. Notably, family 6 lacks structural and biochemical data. However, RIMBP1/2 (RIM-binding protein 1 and 2) can recognize a potential consensus of RXXPXXP and can likely bind to

motifs such as RQLPQL/VP, RLLPPTP, and RQLPQTP found in RIM1/2 (RAB3-interacting molecule 1 and 2). RIMBP1/2 has been shown to bind and couple RIM1/2 to voltage-gated Ca²⁺ channels [27]. Family 7 shows patterns such as PXXPX(K/R), (K/R)XPXXP, (K/R)XXPXXP, PXXPXX(K/R), and PXXP. Family 8 motifs, including PXXP, PXXXPR, and PXXXPXR, highlight a selective array of proline-rich sequences. Family 9, characterized by PXXPX(K/R) and PXXPX(L/P), shows specificity for proline and amino acid residues such as K/R and L/P. Family 10 presents PX(P/A)XXR, PXXPXXP(K/R), PXXPX(K/R), RXX(K/R)P, and PPPPP motifs, illustrating a specific yet versatile proline-rich arrangement. These results highlight the complex yet diverse nature of PRMs across families controlling specific protein interactions and functions.

3.3. Affinity and selectivity of the SH3 family proteins for SOS1 PRP

The intriguing recognition pattern of PRMs observed in SOS1, a particularly PRM-rich protein, caught our attention. SOS1 shows co-occurrence of 13 out of a total of 14 PRMs (including sequences such as PPPP, XPPX, PXP, PXPXP, PPXPP, PXXP, PXXPX[KR], [KR]XXPXXP, PXXPXXP, PXXXP, PXXXPXXXP, PXXXPR, PXXXXP) as shown in Table S5. This distinct pattern, showing selectivity across all SH3 families, motivated us to use SOS1 as a comprehensive model to uncover the recognition mechanisms of established SH3 proteins in the human proteome. Therefore, 25 SH3 domains from different SH3 families were selected (Figure 1), cloned, purified as GST fusion proteins, and used for PRP binding analysis (Table S1). In addition, we selected 10 different PRPs from the proline-rich domain (RPD) of the SOS1 protein (Table S5). This collection was designed to cover the full spectrum of PRM types (P1-P10; Table S2). Two reference peptides were included as a control: RP1, a well-studied SOS1 derivative encompassing part of P3, and RP2, a peptide derived from the N-terminal extension of the RHO GTPase WRCH1/RHOU (Table S2). The 12 PRPs were labeled with FITC to assess their binding capacities with purified GST fusion proteins of the 25 SH3 domains using fluorescence dot blot and polarization analysis.

The binding of 12 FITC-labeled PRPs to 25 GST-SH3 domains was qualitatively analyzed by combining GST pull-down and dot blot assays. GST protein was used as a negative control. Different binding strengths were observed among the proteins tested, particularly with the P2, P3, P4, P7, and P9 peptides (Figure 2A). The strongest interactions (dot intensity >80) were between ABI1 and P2, ITSN1-1 and RP2, and NCK1-3 and P9 and RP1. In contrast, no PRP binding (dot intensity 0) was detected for NCK1-1, NPHP1, RASA1, SH3GLB-1, SNX9, ITSNS1-2, ITSNS1-3, and ITSNS1-4. These data provide valuable insight into the varying degrees of interaction across the panel of PRPs and SH3 domains tested (Figure 2A).

Fluorescence polarization measurements were performed to determine the binding affinities of SH3-PRP interaction pairs from the dot blot analysis. SH3 proteins were titrated at increasing concentrations against fluorescent PRPs, which were kept at a constant concentration of 0.2 μM. GRB2-2^W193K, which is defective in the binding of PRPs such as RP1, was used as a negative control as previously described [22]. Interestingly, none of the PRPs we examined showed any binding for seven SH3 domains: ITSN1-2/-3/-4, NPHP1, RASA1, and SH3GLB1 (Table S6). The resulting data (Figure S6) allowed the evaluation of equilibrium dissociation constants (K_d) for 45 interactions between the SH3 domains and the PRPs (Figure 2B; Table S6). In particular, the results confirmed that the peptides P2, P3, P4, P7, and P9 were associated with approximately 17 SH3 domains. The K_d values determined were categorized into four affinity levels (Figure 2B; Table S6): high (0.1 to 1.0 μM; green), intermediate (1.1 to 5 μM; blue), low (5.1 to 25 μM; red), and very low (26 to 125 μM; black). While previously reported SH3-PRM interactions exhibited micromolar affinities, our results revealed interactions with nanomolar affinities in some cases. The most notable and novel pairs of interaction were ARHGAP12/P7, NCK1-3/P9, and NCK1-2/RP2 with affinities in the nanomolar range of 0.2, 09, and 1.0 μM respectively.
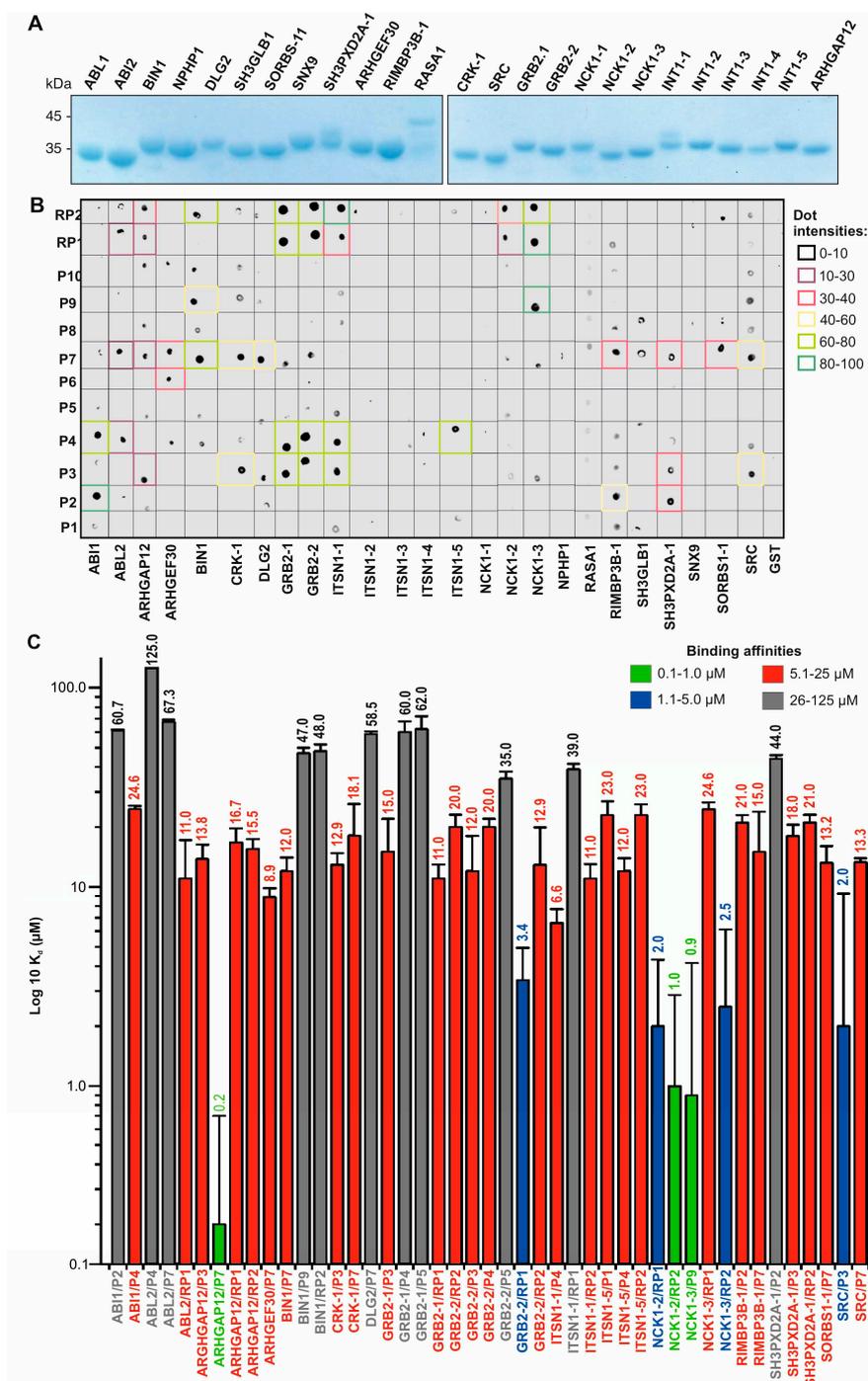
Figure 2. Evaluation of the binding selectivity of SH3 domain representatives with different PRP types. (A) Coomassie brilliant blue stained SDS gels show purified SH3 domains as GST fusion proteins. (B) Fluorescence dot blots revealed variable binding strengths of 12 fluorescent PRPs with 25 GST-SH3 domains. Dot intensities are categorized into five groups ranging from 0 (black) to 100 (dark green). (C) Bar graphs show the evaluated dissociation constants (K_d) for the selected SH3-PRP interactions determined by fluorescence polarization (Table S6; Figure S6). The color codes indicate the K_d values, classified into high affinity (green), intermediate affinity (blue), low affinity (red), and very low affinity (black).

3.4. Non-conserved residues define the selectivity and affinity of SH3-PRM interactions

To better understand the role of residues of SH3 domains in selective binding to PRPs, we have generated a multiple sequence alignment (Figure S7). It highlights the conserved and variable residues that are likely to be critical for the selectivity and affinity of the SH3-PRM interactions. The

importance of variable residues was investigated by specifically selecting ARHGAP12 and NCK1-3 for mutational analysis due to their high binding affinities of 0.2 and 0.9 µM for P7 and P9, respectively. In contrast, ABL2 and BIN1 were selected for their very low affinity for P7 and P9, respectively. Two different sets of mutations were generated by substituting a combination of amino acids from ARHGAP12 and NCK1-3 for ABL2 and BIN1, and *vice versa* (Figure 3A; Table S1).
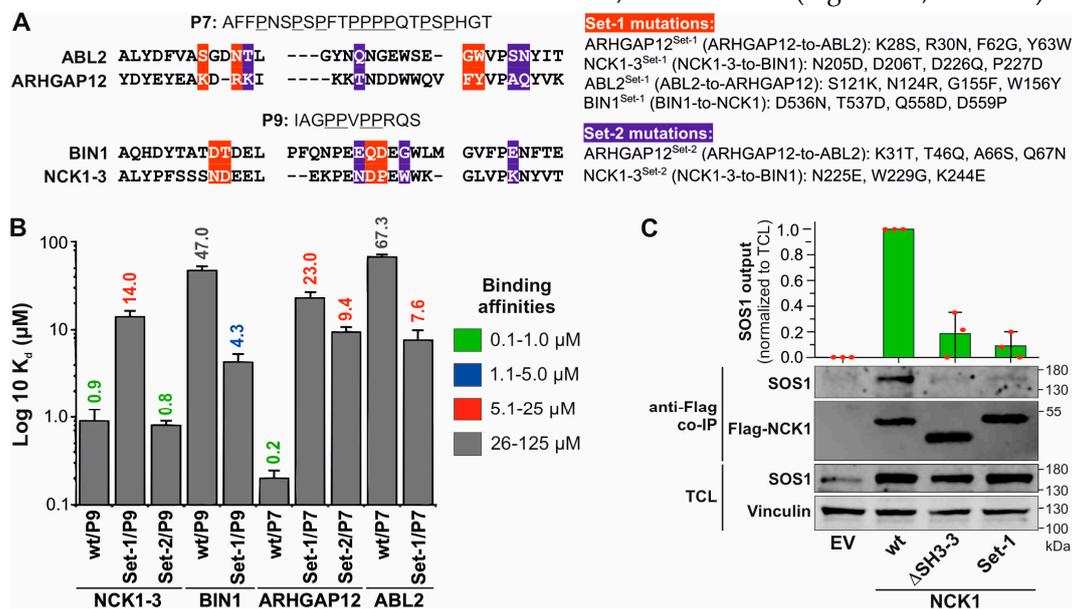
Figure 3. Variable residues determine the specificity and affinity of SH3-PRP interactions. (A) A sequence alignment of the PRM-binding residues of ABL2, ARHGAP12, BIN1, and NCK1-3 is extracted from the alignment shown in Figure S7. Residues in red (Set-1) and blue (Set-2) are variable residues (left panel) and are the subject of mutation analysis (right panel). (B) K_d values determined by fluorescence polarization partially revealed shifts in the binding affinities of the investigated SH3 domain mutants for P7 and P9, respectively, with decreased affinity for ARHGAP12^{Set1} and ARHGAP12^{Set2}, and NCK1^{Set-1}, and increased affinity for ABL2^{Set1} and BIN1^{Set-1}. NCK1^{Set-2} had the same K_d value as NCK1^{wt}. (C) SOS1 co-immunoprecipitated (Co-IP) with NCK1^{wt}, but not with NCK1-3$^{\Delta SH3-3}$ and NCK1^{Set-1} overexpressed in CHO-K1 cells.

Comparative fluorescence polarization measurements between wt and mutant SH3 domains (Figure S8) revealed that variable residues determine selectivity and affinity. The determined K_d values of 14 and 23 µM showed a drastic reduction of the binding affinity of the Set-1 mutants of ARHGAP12 and NCK1 by 115-fold and 15.6-fold for P7 and P9, respectively (Figure 3B). This demonstrates the importance of selected variable residues for the PRP interactions, especially because ABL2^{Set-1} and BIN1^{Set-1} showed a 9- and an 11-fold increase in binding to P7 and P9, respectively (Figure 3B). In light of this result, we decided to investigate another set of variable residues (Set-2; Figure 3A). The binding affinity for ARHGAP12^{Set-2} was reduced 47-fold, indicating the critical role of these residues in determining the selectivity and affinity of ARHGAP12 for P7 (Figure 3B). However, NCK1-3^{Set-2} did not differ from the NCK1-3^{wt} in terms of P9 binding, suggesting that the Set-2 residues are not critical for NCK1-3/P9 interaction (Figure 3B).

To investigate a potential SOS1 binding of ARHGAP12 and NCK1 and to assess the relevance of variable residues in the SH3-PRMs interaction in cells, CHO-K1 cells were co-transfected with wt and mutants of NCK1 and ARHGAP12 together with HA-SOS1, containing P7 and P9 at its C-terminal PRD (Figure 4). Co-immunoprecipitation (Co-IP) with anti-Flag beads was performed to investigate the possible interaction of ARHGAP12 and NCK1 with HA-SOS1 (Figures S9 and S10). As shown in Figure 3C, all three NCK1 proteins, wt, ΔSH3-3, and Set-1, were immunoprecipitated, but SOS1 was only co-immunoprecipitated with NCK1^{wt} but not with NCK1$^{\Delta SH3-3}$ and NCK1^{Set-1} (Figure 3C). No ARHGA12-SOS1 interaction was observed in similar experiments (Figure S10). Taken together, our data not only highlight the essential role of these flanking residues in determining the selectivity and

affinity of SH3 domains for their cognate PRPs but also provide unprecedented insight into a potential SOS1-NCK1 interaction in cells.

3.5. SH3-PRP relationships beyond SOS1

We used the position-specific iterated BLAST (PSI-BLAST) algorithm to perform an analysis of the SOS1 PRPs in the human proteome using the NCBI (National Center for Biotechnology Information) database as a reference. Our goal was to identify homologous peptides from proteins other than SOS1 as potential interaction partners for the SH3 domains investigated in this study. In our analysis, we considered alignments with percent identity ranging from 98% to 100% and with E values up to 10, which, although less stringent, may still indicate potential similarities between the SOS1 PRPs and other protein sequences. More than 30 proteins were found with partial sequence identities with P2 to P9 (Table S7; Figures 4 and S11), and 0 with P1 and P10 (Refer to Table S1). Some of these proteins contain multiple PRM repeats (Figure S11), for example, IQSEC2, paxillin, DLGAP1, PI3KAP1, and WRCH1/RHOU within P3 (Motif: PVPPPVP), SSTR5 and SLX4 within P7 (Motif: PPPPQTP), DCAF1 and MAGED4 within P8 (Motif: HLPSPP) and SOS2 and HCG2013210 within P9 (Motif: PPVPPRQ). This suggests that the SH3 interactions with the PRPs characterized in this study go beyond SOS1 as a binding partner, although the binding specificities of the listed proteins (Figure 4) remain to be investigated. However, several studies confirm the interactions of the identified proteins with SH3DCPs, including zinc finger proteins with p130Cas [37], MACF1 with Spectrin [38], DLGAP with DLG [39], WRCH1/RHOU with GRB2 [42,43], Paxillin with SRC [44,45], and SSTR5 with Homer, Dynamin, IRSp53, and Cortactin [46].
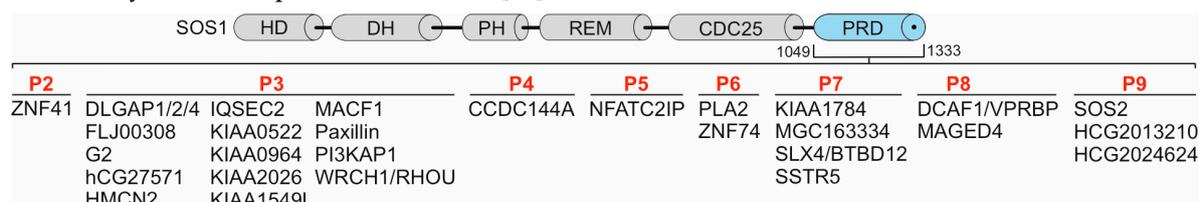
Figure 4. Proteins containing PRMs homologous to P2-P9 derived from the SOS1 PRD. BLAST searches with each SOS1 PRP identified several human proteins (see Table S7) with a high degree of sequence similarity to P2-P9. CDC25, cell division cycle 25; DH, DBL homology domain; HD, histone-like domain; PH, pleckstrin domain; REM, RAS exchange motif; PRD, proline-rich domain.

4. Discussion

SH3 domains are critical in multiple signaling pathways, interacting with diverse proteins involved in apoptosis, proteasomal degradation, endocytosis, and with SRC family protein tyrosine kinases, influencing downstream processes including proliferation, cell survival, growth, actin reorganization, and cell migration [4]. The broad influence of SH3 domains on cellular functions raises fundamental questions about the specificity of their interaction networks. Previous research has also highlighted the importance of proline amino acids in forming the polyproline type II helix (PPII) conformation, which provides a binding pocket for SH3 domain residues, particularly from the RT and n-SRC loops [17]. Despite the discovery of 14 PRM consensus sequences in the human proteome (Table S5), the specificity pattern of the interaction of SH3 domains with PRMs is still unclear. Understanding the molecular basis of SH3 domain-PRM interactions is crucial to gain insight into how these interactions regulate signaling pathways.

In recent years, several studies have been devoted to elucidating the diverse nature of SH3 domain interactions. Cesareni and colleagues used a novel chip technology to perform high-throughput qualitative analyses, revealing a variety of human SH3 domains that fall into two categories: those characterized by classical proline-rich core motifs accompanied by positively charged amino acids, and atypical ones lacking the core motif [20]. Nevertheless, the diversity of PRM selectivity patterns is evident among all human SH3 domains. Furthermore, a comprehensive analysis of SH3 domain interactions in the evolution of four yeast species, Saccharomyces cerevisiae,

Ashbya gossypii, Candida albicans, and *Schizosaccharomyces*, revealed that nearly 75 percent of SH3 families identified within the phylogenetic tree have a conserved SH3 specificity profile over 400 million years of evolution [47]. Utilizing the evolutionary relationships of peptide recognition domains in eukaryotes, we identified common structural features and ancestry that allowed us to group SH3 domains into similar binding preference families. This comprehensive investigation aimed to clarify the specificity profiles of SH3 PRMs within the human proteome through categorization based on the phylogenetic tree of SH3DCPs.

To provide an accurate specificity map of SH3 domains, we performed deep phylogenetic analyses coupled with computational analysis of related structural data. Initial evolutionary analysis of the sequence-structure-function of full-length SH3 domains was unsuccessful due to the presence of SH3 regions that do not interact with PRMs. Within these, each SH3 domain exhibited variation in binding specificity to PRMs, making the characterization of SH3DCPs infeasible. Instead of relying on full-length sequences, we focused on binding residues that directly interact with PRMs, as revealed by sequence alignments coupled with analysis of published SH3 domain structures in complex with PRMs. This refined phylogenetic approach led to the identification of ten distinct families based on both structural and biochemical assessments of SH3-PRM interactions and their distribution within the phylogenetic tree. This approach facilitates the assessment of cross-reactivity among SH3 domain recognition sites for PRMs, a phenomenon also observed in previous studies examining the yeast SH3 domain peptide library [48] and including SH3 domains that recognize multiple PRMs. The findings of this study highlight the fact that each SH3 domain family interacts with different but distinct sets of PRMs.

Considering the significant involvement of SOS1 in interactions with SH3DCPs such as GRB2 [49], ITSN1 [50], NCK1 [51], and ABI1 [52], along with its comprehensive coverage of all known PRMs in the human proteome (Tables S2 and S5), we decided to use SOS1 as a model for in-depth exploration of SH3 domain specificity in the realm of polyproline interactions. We performed *in vitro* studies with 25 representative SH3 domains selected from the phylogenetic tree. We performed low-throughput analyses, including pull-down assays, dot blotting, and fluorescence polarization, to investigate SH3-SOS1-PRP interactions. These investigations revealed novel interactions with nanomolar affinities, which were subsequently confirmed by mutational studies.

The general concept of protein association is essential for describing protein-protein interactions in complexes, especially those with weak affinities in the micromolar range or transient interactions such as the SH3 domain with PRMs [53]. Mayer and Saksela noted that the limited selectivity of SH3 domains for PRMs implies that SH3-mediated interactions may be highly dependent on external environmental factors [54]. In certain scenarios, the presence of additional surfaces on either the SH3 domain or the ligand it recognizes, along with the presence of either multiple SH3 domains or different domains within the same protein, or even the co-localization of two partners within a multi-protein complex, can cooperatively enhance SH3-PRM specificity to a significant degree [55]. This suggests that the low-affinity region results in our study may be compensated by these scenarios, ultimately increasing the affinity and specificity of SH3-PRM interactions. It has also been reported that the moderate affinities of SH3-mediated interactions imply that the interactions have a high dynamic remodeling potential (rapid off-rates), depending on the subcellular localization and accessible binding partners [54]. This observation is consistent with our polarization data, which showed high K_d values for many low-affinity interactions.

However, in some cases, a negative effect of other domains on SH3-PRM binding was observed. Notably, the NCK SH3.1-3 constructs exhibited more extensive protein precipitation than the full-length protein, possibly indicating a detrimental effect of the SH2 domain on specific SH3-mediated interactions. In addition, the range of proteins precipitated with the SH3.1-3 construct is not simply the cumulative sum of proteins precipitated with individual SH3 domains [56]. Of note, the novel high nanomolar affinity of NCK1-2 and NCK1-3 in our measurements with PRPs could potentially explain this newfound high affinity when considering only the SH3 domain rather than the full-length protein. Furthermore, Zarrinpar's results show that isolated SH3 domains contain enough information to encode interaction specificity between yeast SH3 domains [57]. This may also be true

for certain high-affinity human SH3-PRM interactions, where SH3 alone may determine their high affinity and specificity.

A closer examination of the PRMs revealed that the positioning of the proline residues plays a critical role in the recognition of the SH3 domain, providing the structural basis for defining interaction specificity. The current results indicate that residues -2, -1, +1, and +2 are critical for the recognition of SH3 PRMs. In addition, adjacent positively charged residues contribute additional features that help stabilize the transient interaction [58]. For example, structural studies of the PI3K SH3 domain in association with the peptide RKLPPRPSK provided evidence for the role of adjacent non-proline residues such as Arg-1, Leu-3, and Arg-6 in contributing to the SH3 domain interaction [59]. Arginine residues at positions R+5, R+6, and R+7 are thought to play an important role in enhancing the affinity of GRB2-SH3 domains for SOS1-PRM by contributing to the overall free energy of the interaction [60]. While the majority of SH3 domains interact with PRMs, there have been documented cases where the SH3 domain deviates from the typical classical proline-rich interaction pattern [4]. For example, the RASA1 (or p120RASGAP) SH3 domain specifically interacts with the catalytic arginine finger of the RHOGAP domain of DLC1, thereby competitively and potently inhibiting its RHOGAP activity [61]. Interestingly, none of the PRPs we examined showed any binding for seven SH3 domains: ITSN1-2/-3/-4, NPHP1, RASA1, and SH3GLB1 (Table S6). While our study comprehensively highlights the major SH3 domain-PRM interactions in the human genome, the specificity and mechanism of the PRM-independent interaction of SH3 domains remain to be elucidated in further studies.

In our study, SOS1 is used as a PRP model for biophysical and bioinformatic analysis of the SH3-PRM interaction landscape because the SOS1 PRD contains 13 out of 14 different classified proline-rich consensus sequence motifs (Table S5). This alternative model reveals a spectrum of interactions between different SOS1 PRPs and a number of SH3DCPs, including ABI1, ABL2, ARHGAP12, ARHGEF30 (Obscurin/OBSCN), BIN1, CRK-1, DLG2, GRB2, ITSN1, NCK1, SRC, SH3PXD2A-1, and SORBS1-1. Among them, ABI1 [28], ITSN1 [29], SRC [30], NCK1 [31], GRB2 [32], CRK [33], SH3PXD2A (TKS5) [34] have been previously established as SOS1 binding partners in cells. ABL1, but not ABL2, has also been shown to interact with SOS1 [35]. Importantly, the precise binding sites for most of these proteins have yet to be investigated. Our study not only elucidates the binding sites of these established SOS1 partners, but also uncovers novel interactions, including ABL2, BIN1, DLG2, SORBS1, ARHGEF30, and ARHGAP12. In particular, a high affinity (<0.5 μM) interaction was observed between ARHGAP12 and P7 of SOS1, demonstrating the interplay between small GTPase regulators, GAPs, and GEFs. However, immunoprecipitation experiments with overexpressed ARHGAP12 in CHO-K1 cells did not confirm an interaction with SOS1 (Figure S10). The reliability of this result may depend on the expression level and affinity of other interaction partners, like accessory proteins, that could potentially bind more strongly, possibly in a multivalent manner. However, it is important to note that this result does not definitively rule out the existence of this interaction in cells, especially considering examples of reported GEF-GAP interactions. A study using immunoprecipitation and mass spectrometry unravels the intricate protein interaction networks involving the synaptic proteins SYNGAP1 (RASGAP), KALIRIN (RHOGEF), and AGAP2 (ARFGAP) in both the postsynaptic density (PSD) and non-PSD fractions of adult mouse cortex. This investigation sheds light on their role in the organization of GAP and GEF protein families and their associations with proteins associated with intellectual disability and psychiatric disorders [36]. In conclusion, to confirm the significance and broader implications of these novel findings, additional studies within the cellular context are warranted.

Predicting the potential interaction of SH3 domains with PRMs by considering their sequence specificities, as we did in this study, is a promising approach in the field of molecular biology and protein-protein interactions. In pursuing this goal, we are faced with an interesting challenge: the identification of binding affinities between SH3 domains and peptides containing PRMs. The sequence specificity of these interactions is paramount, as SH3 domains exhibit diverse binding preferences that depend on the PRMs present in the peptides. Using bioinformatics and phylogenetic trees, complemented by biophysical measurements, our goal is to decipher the intricate matrix that

governs these interactions. This predictive approach may have far-reaching implications for the understanding of cellular signaling pathways and the development of novel therapeutic interventions.

References

1. Mabonga, L. and A.P. Kappo, *Protein-protein interaction modulators: advances, successes and remaining challenges.* Biophysical Reviews, 2019. **11**(4): p. 559-581.
2. Musacchio, A., et al., *Crystal structure of a Src-homology 3 (SH3) domain.* Nature, 1992. **359**(6398): p. 851-855.
3. Stahl, M.L., et al., *Sequence similarity of phospholipase C with the non-catalytic region of src.* Nature, 1988. **332**(6161): p. 269-272.
4. Mehrabipour, M., et al., *A Systematic Compilation of Human SH3 Domains: A Versatile Superfamily in Cellular Signaling.* Cells, 2023. **12**(16): p. 2054.
5. Teyra, J., et al., *Comprehensive analysis of the human SH3 domain family reveals a wide variety of non-canonical specificities.* Structure, 2017. **25**(10): p. 1598-1610. e3.
6. Berndt, S., V.V. Gurevich, and T. Iverson, *Crystal structure of the SH3 domain of human Lyn non-receptor tyrosine kinase.* Plos one, 2019. **14**(4): p. e0215140.
7. Bircher, J.E. and A.J. Koleske, *Trio family proteins as regulators of cell migration and morphogenesis in development and disease – mechanisms and cellular contexts.* Journal of Cell Science, 2021. **134**(3): p. jcs248393.
8. Hildebrandt, F., et al., *A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type 1.* Nat Genet, 1997. **17**(2): p. 149-53.
9. Saunier, S., et al., *Characterization of the NPHP1 locus: mutational mechanism involved in deletions in familial juvenile nephronophthisis.* Am J Hum Genet, 2000. **66**(3): p. 778-89.
10. Wu, T., Z. Shi, and T. Baumgart, *Mutations in BIN1 associated with centronuclear myopathy disrupt membrane remodeling by affecting protein density and oligomerization.* PloS one, 2014. **9**(4): p. e93060.
11. Hohendahl, A., A. Roux, and V. Galli, *Structural insights into the centronuclear myopathy-associated functions of BIN1 and dynamin 2.* J Struct Biol, 2016. **196**(1): p. 37-47.
12. Manso, J.A., et al., *PSTPIP1-LYP phosphatase interaction: structural basis and implications for autoinflammatory disorders.* Cell Mol Life Sci, 2022. **79**(2): p. 131.
13. Starnes, T.W., et al., *The F-BAR protein PSTPIP1 controls extracellular matrix degradation and filopodia formation in macrophages.* Blood, 2014. **123**(17): p. 2703-14.

14. Qin, Y., et al., *A recurrent SHANK1 mutation implicated in autism spectrum disorder causes autistic-like core behaviors in mice via downregulation of mGluR1-IP3R1-calcium signaling.* Mol Psychiatry, 2022. **27**(7): p. 2985-2998.

15. Leblond, C.S., et al., *Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders.* PLoS genetics, 2012. **8**(2): p. e1002521.

16. Ball, L.J., et al., *Recognition of Proline-Rich Motifs by Protein–Protein-Interaction Domains.* Angewandte Chemie International Edition, 2005. **44**(19): p. 2852-2869.

17. Li, S.S.-C., *Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction.* Biochemical Journal, 2005. **390**(3): p. 641-653.

18. Cesareni, G. and M. Gimona, *Modular Protein Domains.* 2005.

19. Cámara-Artigas, A., et al., *3D domain swapping in a chimeric c-Src SH3 domain takes place through two hinge loops.* Journal of structural biology, 2014. **186**(1): p. 195-203.

20. Carducci, M., et al., *The protein interaction network mediated by human SH3 domains.* Biotechnology advances, 2012. **30**(1): p. 4-15.

21. Panni, S., L. Dente, and G. Cesareni, *In vitro evolution of recognition specificity mediated by SH3 domains reveals target recognition rules.* Journal of Biological Chemistry, 2002. **277**(24): p. 21666-21674.

22. Kazemein Jasemi, N.S., et al., *The intramolecular allostery of GRB2 governing its interaction with SOS1 is modulated by phosphotyrosine ligands.* Biochemical Journal, 2021. **478**(14): p. 2793-2809.

23. Mayer, B.J., *SH3 domains: complexity in moderation.* Journal of cell science, 2001. **114**(7): p. 1253-1263.

24. Kay, B.K., *SH3 domains come of age.* FEBS letters, 2012. **586**(17): p. 2606-2608.

25. Landgraf, C., et al., *Protein interaction networks by proteome peptide scanning.* PLoS Biol, 2004. **2**(1): p. e14.

26. Bae, D.J., et al., *ArhGAP12 plays dual roles in Stabilin-2 mediated efferocytosis: Regulates Rac1 basal activity and spatiotemporally turns off the Rac1 to orchestrate phagosome maturation.* Biochim Biophys Acta Mol Cell Res, 2019. **1866**(10): p. 1595-1607.

27. Hibino, H., et al., *RIM binding proteins (RBPs) couple Rab3-interacting molecules (RIMs) to voltage-gated Ca(2+) channels.* Neuron, 2002. **34**(3): p. 411-23.

28. Fan, P.-D. and S.P. Goff, *Abl interactor 1 binds to sos and inhibits epidermal growth factor-and v-Abl-induced activation of extracellular signal-regulated kinases.* Molecular and cellular biology, 2000. **20**(20): p. 7591-7601.

29. Tong, X.K., et al., *The endocytic protein intersectin is a major binding partner for the Ras exchange factor mSos1 in rat brain.* Embo j, 2000. **19**(6): p. 1263-71.

30. Qian, X., et al., *The Sos1 and Sos2 Ras-specific exchange factors: differences in placental expression and signaling properties.* Embo j, 2000. **19**(4): p. 642-54.

31. Hu, Q., D. Milfay, and L.T. Williams, *Binding of NCK to SOS and activation of ras-dependent gene expression.* Mol Cell Biol, 1995. **15**(3): p. 1169-74.

32. Chardin, P., et al., *Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2.* Science, 1993. **260**(5112): p. 1338-1343.

33. Birge, R.B., et al., *Crk and CrkL adaptor proteins: networks for physiological and pathological signaling.* Cell Communication and Signaling, 2009. **7**(1): p. 13.

34. Rufer, A.C., et al., *Isoform-selective interaction of the adaptor protein Tks5/FISH with Sos1 and dynamins.* J Mol Biol, 2009. **390**(5): p. 939-50.

35. Sini, P., et al., *Abl-dependent tyrosine phosphorylation of Sos-1 mediates growth-factor-induced Rac activation.* Nature Cell Biology, 2004. **6**(3): p. 268-274.

36. Wilkinson, B., J. Li, and M.P. Coba, *Synaptic GAP and GEF complexes cluster proteins essential for GTP signaling.* Scientific Reports, 2017. **7**(1): p. 5272.

37. Nakamoto, T., et al., *CIZ, a zinc finger protein that interacts with p130(cas) and activates the expression of matrix metalloproteinases.* Mol Cell Biol, 2000. **20**(5): p. 1649-58.

38. Salem, D. and R.J. Fecek, *Role of microtubule actin crosslinking factor 1 (MACF1) in bipolar disorder pathophysiology and potential in lithium therapeutic mechanism.* Translational Psychiatry, 2023. **13**(1): p. 221.

39. Rasmussen, A.H., H.B. Rasmussen, and A. Silahtaroglu, *The DLGAP family: neuronal expression, function and role in brain disorders.* Molecular Brain, 2017. **10**(1): p. 43.

40. Maruoka, M., et al., *Identification of B cell adaptor for PI3-kinase (BCAP) as an Abl interactor 1-regulated substrate of Abl kinases.* FEBS Lett, 2005. **579**(14): p. 2986-90.

41. Kim, E., et al., *GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules.* J Cell Biol, 1997. **136**(3): p. 669-78.

42. Bustelo, X.R., V. Sauzeau, and I.M. Berenjeno, *GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo.* Bioessays, 2007. **29**(4): p. 356-70.

43. Risse, S.L., et al., *SH3-mediated targeting of Wrch1/RhoU by multiple adaptor proteins.* Biol Chem, 2013. **394**(3): p. 421-32.

44. López-Colomé, A.M., et al., *Paxillin: a crossroad in pathological cell migration.* Journal of Hematology & Oncology, 2017. **10**(1): p. 50.

45. Webb, D.J., et al., *Paxillin phosphorylation sites mapped by mass spectrometry.* Journal of Cell Science, 2005. **118**(21): p. 4925-4929.

46. Kreienkamp, H.-J., et al., *Physiology of somatostatin receptors: from genetics to molecular analysis.* Somatostatin, 2004: p. 185-202.

47. Verschueren, E., et al., *Evolution of the SH3 domain specificity Landscape in Yeasts.* PloS one, 2015. **10**(6): p. e0129229.

48. Cesareni, G., et al., *Can we infer peptide recognition specificity mediated by SH3 domains?* FEBS letters, 2002. **513**(1): p. 38-44.

49. Chook, Y.M., et al., *The Grb2-mSos1 complex binds phosphopeptides with higher affinity than Grb2.* Journal of Biological Chemistry, 1996. **271**(48): p. 30472-30478.

50. Herrero-Garcia, E. and J.P. O'Bryan, *Intersectin scaffold proteins and their role in cell signaling and endocytosis.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2017. **1864**(1): p. 23-30.

51. Alfaidi, M., M.L. Scott, and A.W. Orr, *Sinner or Saint?: Nck adaptor proteins in vascular biology.* Frontiers in Cell and Developmental Biology, 2021. **9**: p. 688388.

52. Yu, X., et al., *Inhibitory short peptides targeting EPS8/ABI1/SOS1 tri-complex suppress invasion and metastasis of ovarian cancer cells.* BMC Cancer, 2019. **19**(1): p. 878.

53. Hahn, S. and D. Kim, *Transient protein-protein interaction of the SH3-peptide complex via closely located multiple binding sites.* PLoS One, 2012. **7**(3): p. e32804.

54. Mayer, B.J. and K. Saksela, *SH3 domains.* Modular protein domains, 2005: p. 37-58.

55. Dionne, U., et al., *SRC homology 3 domains: Multifaceted binding modules.* Trends in Biochemical Sciences, 2022.

56. Lettau, M., et al., *The adapter protein Nck: role of individual SH3 and SH2 binding modules for protein interactions in T lymphocytes.* Protein Sci, 2010. **19**(4): p. 658-69.

57. Zarrinpar, A., S.-H. Park, and W.A. Lim, *Optimization of specificity in a cellular protein interaction network by negative selection.* Nature, 2003. **426**(6967): p. 676-680.

58. Ahmad, M., W. Gu, and V. Helms, *Mechanism of fast peptide recognition by SH3 domains.* Angewandte Chemie International Edition, 2008. **47**(40): p. 7626-7630.

59. Yu, H., et al., *Structural basis for the binding of proline-rich peptides to SH3 domains.* Cell, 1994. **76**(5): p. 933-945.

60. McDonald, C.B., et al., *SH3 domains of Grb2 adaptor bind to PXψPXR motifs within the Sos1 nucleotide exchange factor in a discriminate manner.* Biochemistry, 2009. **48**(19): p. 4074-4085.

61. Jaiswal, M., et al., *Functional cross-talk between ras and rho pathways: a Ras-specific GTPase-activating protein (p120RasGAP) competitively inhibits the RhoGAP activity of deleted in liver cancer (DLC) tumor suppressor by masking the catalytic arginine finger.* Journal of Biological Chemistry, 2014. **289**(10): p. 6839-6849.