The specificity of SRC Homology 3 (SH3) domain interactions is modulated by their protein context *in vivo*

Ugo Dionne[1,2,3,7,9], Émilie Bourgault[3,6,7,9,12], Alexandre K Dubé[3,5,6,7,9,12], David Bradley[3,5,6,7,9,12], François Chartier[1,2,3,12], Rohan Dandage[3,5,6,7,9], Soham Dibyachintan[3,5,6,7,10], Philippe C Després[3,6,7,9], Gerald D Gish[11], Jean-Philippe Lambert[1,2,8], Nicolas Bisson[1,2,3,4*] and Christian R Landry[3,5,6,7,9,13*]

[1] Centre de recherche du Centre Hospitalier Universitaire (CHU) de Québec-Université Laval, Québec, QC, Canada.
[2] Centre de recherche sur le cancer de l'Université Laval, Québec, QC, Canada.
[3] PROTEO-Regroupement québécois de recherche sur la fonction, l'ingénierie et les applications des protéines, Québec, QC, Canada.
[4] Département de biologie moléculaire, biochimie médicale et pathologie, Université Laval, Québec, QC, Canada.
[5] Département de biologie, Université Laval, Québec, QC, Canada.
[6] Département de biochimie, de microbiologie et de bio-informatique, Université Laval, Québec, QC, Canada.
[7] Centre de recherche en données massives de l'Université Laval, Université Laval, Québec, QC, Canada
[8] Département de médecine moléculaire, Université Laval, Québec, QC, Canada.
[9] Institut de biologie intégrative et des systèmes (IBIS), Université Laval, Québec, QC, Canada.
[10]Department of chemical engineering, Indian Institute of Technology Bombay (IIT), Powai, Mumbai, Maharashtra, India.
[11]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex, Toronto, ON, Canada
[12] These authors contributed equally
[13] Lead contact

*Authors for correspondence:
Christian R Landry christian.landry@bio.ulaval.ca
Nicolas Bisson Nick.Bisson@crchudequebec.ulaval.ca

Abstract:

SRC Homology 3 (SH3) domains contribute to cellular processes via their ability to support protein-protein interactions (PPIs). While the intrinsic binding specificities of SH3 domains have been studied *in vitro*, whether each domain is necessary and sufficient to define PPI specificity *in vivo* is largely unknown. We combine genome editing, deep mutagenesis scanning, PPI mapping, growth phenotyping and single cell imaging in yeast to identify SH3-dependent PPI networks and their associated phenotypes. We show that both the sequence of the SH3 host protein and the position of the SH3 domains within their host are critical for interaction specificity and for cellular processes such as endocytosis. We further validate these findings using a human multi-SH3 adaptor protein and investigating its ability to promote phase separation. Our work highlights the importance of the interplay between a SH3 domain and its host protein for the regulation of cellular and biophysical processes.

Introduction:

The adaptation of cells to changing environments depends on the detection of intra- and extracellular cues (Mayer, Blinov and Loew, 2009). The activation of signaling pathways promotes the formation of dynamic protein complexes and networks of protein-protein interactions (PPIs) (Bisson *et al.*, 2011; Ideker and Krogan, 2012). The specificity of these PPIs is crucial for cellular homeostasis despite many of them being weak and transient (Gao *et al.*, 2018). Consequently, disease-causing mutations are enriched on protein interaction interfaces and frequently affect PPIs (Bergendahl *et al.*, 2019).

Cell signaling proteins typically display a modular architecture partially defined by folded domains that bind short linear peptide motifs on their targets (Scott and Pawson, 2009). SRC Homology 3 domains (SH3s) are one of the most prevalent families of such PPI modules (Mayer, Hamaguchi and Hanafusa, 1988; Stahl *et al.*, 1988; Mayer, 2015). These ~60 amino acid domains are present on signaling proteins regulating functions such as endocytosis and actin cytoskeleton remodelling (Mirey *et al.*, 2005; Carducci *et al.*, 2012). SH3s have expanded in numbers throughout the evolution of signaling networks, with 27 in yeast and nearly 300 members in humans (Vogel and Chothia, 2006; Teyra *et al.*, 2017). Not only have genes coding for SH3-containing proteins been duplicated, but SH3s themselves have been duplicated within proteins (Vogel, Teichmann and Pereira-Leal, 2005; Vogel and Chothia, 2006; Kelil, Levy and Michnick, 2016). This led to the birth of a large number of proteins containing 2-3 SH3 copies, favoring the formation of higher order protein complexes through multivalent interactions and elaborate arrangements such as the ones leading to phase separation (Banjade and Rosen, 2014). With the expansion of SH3s and their potential partners came a requirement for higher specificity in their interactions (Kelil, Levy and Michnick, 2016).

Domain binding specificity can be encoded in the domain itself, meaning that the SH3 is necessary and sufficient to define its own binding specificity within the host protein. SH3s display a distinctive 3D structure, which contrasts with their relatively low sequence conservation (Kay, 2012). They typically bind to Pro/Arg-rich motifs on their target partners, and are denoted as class I or class II according to the position of a positively charged residue around an archetypical PXXP motif (where X represents any amino acid) (Carducci *et al.*, 2012; Teyra *et al.*, 2017). However, it was recently demonstrated that SH3s can bind to 10 different categories of peptide motifs, including atypical Pro/Arg-less peptides (Teyra *et al.*, 2017). Modular domains such as SH3s are generally considered as acting as “beads on a string” by virtue of their ability to independently fold and bind to protein partners with high intrinsic specificity (Tong *et al.*, 2002; Tonikian *et al.*, 2009; Kay, 2012; Xin *et al.*, 2013; Mayer, 2015; Kelil *et al.*, 2017; Teyra *et al.*, 2017; Ivarsson and Jemth, 2019; Cunningham *et al.*, 2020). Following this model, the large scale characterisation of the preferred binding motifs of individual *S. cerevisiae*, *C. elegans* and human SH3s has been performed systematically on isolated domains (Tong *et al.*, 2002; Tonikian *et al.*, 2009; Xin *et al.*, 2013; Teyra *et al.*, 2017), either using peptide libraries *in vitro* or yeast two-hybrid assays (Y2H). Computational models allowing the prediction of interactions mediated by SH3s have also been designed, mostly stemming from *in vitro* information (Kelil *et al.*, 2017; Cunningham *et al.*, 2020). These studies have yielded extremely rich data regarding the intrinsic binding specificities of SH3s but only represent a first possible layer of regulation. Since about half of human proteins contain at least

one classical PXXP motif (Ravi Chandra *et al.*, 2004), the binding preference of SH3 domains taken individually cannot by itself explain specificity within their host proteins. As such, a second means of regulation of SH3 specificity may arise from the cellular context, including compartmentalization and co-expression of interacting partners (Zarrinpar, Park and Lim, 2003; Ivarsson and Jemth, 2019).

A third layer of complexity may emanate from the functional relationship between SH3s and their host proteins, defined here as protein context, which hinges on domain positioning and/or intramolecular interactions, including complex arrangements for proteins with multiple SH3s (Arbesú and Pons, 2019; Laursen *et al.*, 2020). While the role of the host protein has not been investigated in detail, a few studies suggest that it is an important layer of control. Indeed, SH3s are often part of larger proteins that can comprise many functional elements. As such, the sequence of their host protein and their position within the protein varies and likely affects SH3 specificity. For instance, the SH3, SH2 and kinase domains of SRC tyrosine kinases have co-evolved and their relative position within the host protein is required for the formation of an auto-inhibited conformation (Nars and Vihinen, 2001; Yadav and Miller, 2008). Another example is the SH3 domain of the MAGuK scaffold PSD-95, which participates in the formation of a three-domain supramodule that forms a dynamic supertertiary structure regulating oligomerization and binding to protein partners in synapses (McCann *et al.*, 2012; Laursen *et al.*, 2020). Furthermore, we and others have shown that SH3 affinity for targets can be controlled via the phosphorylation of a conserved tyrosine within the domain's C-terminus. This phospho-regulation depends on the recruitment of the SH3s to cognate kinases via other domains present in the SH3-containing proteins *in vivo* (Broome and Hunter, 1997; Li *et al.*, 2001; Dionne *et al.*, 2018; Merő *et al.*, 2019). For multi SH3-containing proteins, specificity and affinity can also be increased via the cooperative binding of their SH3s to multiple sites on a given target (i.e. multivalency) (Banjade and Rosen, 2014; Banjade *et al.*, 2015; Errington, Bruncsics and Sarkar, 2019; Ivarsson and Jemth, 2019). Together, these studies exemplify that SH3 function may be regulated by other functional modules within host proteins.

Despite the accumulation of knowledge on SH3 domains, the complete extent of their interactions and functions within their protein context remains poorly characterized. In order to fully understand how PPI networks achieve specificity and how mutations, domain gains and losses alter specificity during evolution and in diseases, we sought to investigate the contribution of protein context. To reach this goal, we investigated how many PPIs are dependent on SH3s *in vivo* by individually deleting all yeast SH3s to infer a SH3-dependent PPI network (Figure 1A). We hypothesized that PPIs mediated by SH3s *in vivo* are influenced by each domain's protein context, i.e. SH3 host protein identity, and that position is critical for SH3 binding specificity (Figure 1A). We therefore swapped SH3s among proteins to test whether a SH3 placed into a new protein context changes its binding specificity (Figure 1A). We discovered that even PPIs that do not depend on an endogenous SH3 (i.e. SH3-independent PPIs) can be altered by domain swapping. By performing an *in vivo* deep mutagenesis scan (DMS) of a SH3, we found that mutating the peptide-binding residues tend to disrupt specifically SH3-dependent PPIs and that mutations of buried SH3 residues alter both SH3-dependent and SH3-independent interactions. Lastly, we used a yeast and a human protein containing multiple SH3s to systematically shuffle SH3s' positions and measure the impact on binding specificity (Figure 1A). Surprisingly, interchanging the position of the SH3s had a drastic impact on PPIs,

even for domains that bind similar motifs *in vitro*. Our results demonstrate the important contribution of protein context to the binding specificity of SH3 domains *in vivo* and highlight the crucial role of conserved domain positioning for interactions and functions.

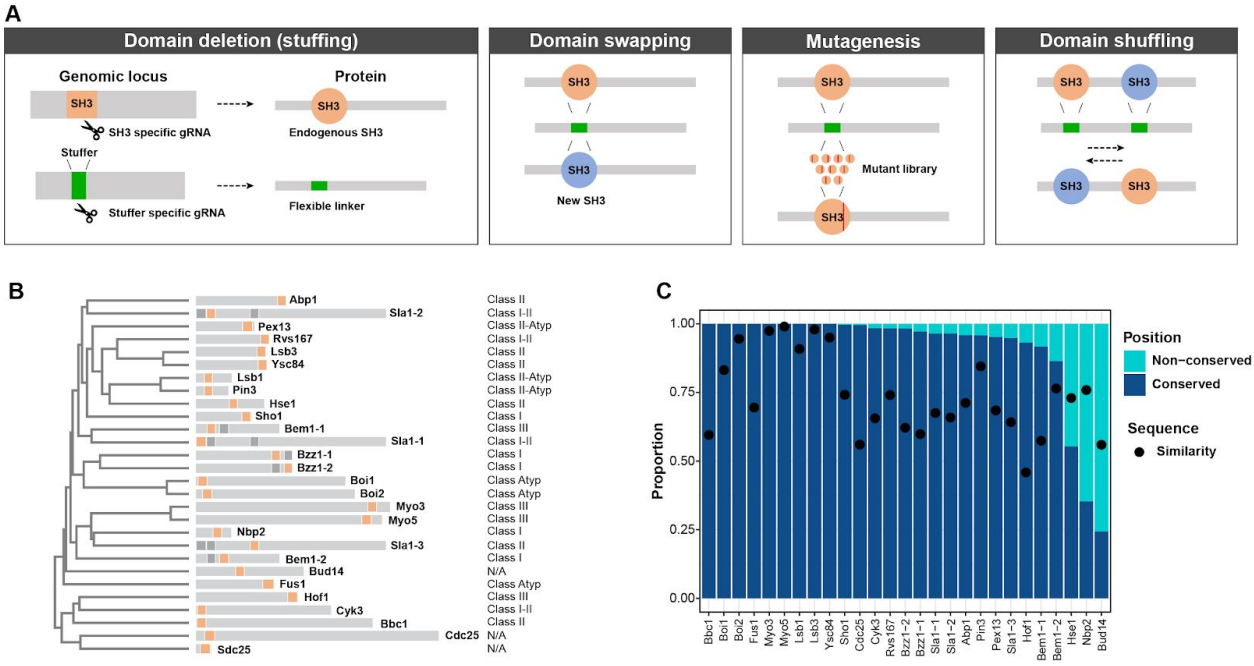

Figure 1: Experimental approach, diversity of SH3 domains and their conservation in fungi. A) CRISPR-Cas9 SH3 editing approaches to study SH3 domains in living cells. For domain deletion (stuffing), SH3 sequences are targeted by a specific gRNA using CRISPR-Cas9 and replaced by a stuffer sequence encoding a flexible linker. The stuffer DNA is then targeted by a stuffer gRNA and replaced by a SH3 from another protein (domain swapping). For deep mutagenesis in the genome, the stuffer is replaced by a library of single mutants. Domains are also moved from one position to another in proteins containing multiple SH3s, using the same tools but with multiple stuffers (domain shuffling). B) The relationships of yeast SH3 amino acid sequences. Domain position and length relative to their host proteins are illustrated. SH3s used in the phylogeny are shown in orange. Grey squares are other SH3s in the same protein. The preferred type of binding motifs from *in vitro* assays are indicated (Tonikian *et al.*, 2009). C) The y-axis shows the fraction of one-to-one orthologs (across 449 fungal species) with conserved domain positions relative to *S. cerevisiae*. The mean sequence similarity is represented by black circles. The pseudogene Scd25 was excluded from these analyses. See also Figure S1 and Data S1.

Results:

Yeast SH3 domains are necessary for a large number of PPIs and associated functions *in vivo*

The budding yeast proteome contains 27 SH3s in 23 proteins (the Sdc25 pseudogene also contains a SH3, Figure 1B, S1A and Data S1). We compared the sequences of one-to-one orthologs in 449 species of Fungi and found that the position of SH3s in their host protein is highly conserved and is generally independent of the extent of their amino acid sequence conservation (Pearson's correlation = 0.22, p-value = 0.28, Figure 1C and S1B). These results suggest that SH3 positioning is important for protein function.

To test the extent to which yeast SH3s are necessary for PPIs *in vivo*, we measured binary PPIs between 22 of the 23 WT SH3-containing proteins as baits and their putative partners as preys

using dihydrofolate reductase protein-fragment complementation assay (DHFR-PCA). DHFR-PCA allows for the detection of PPIs between proteins that are in direct or near-direct contact and that are expressed at endogenous levels *in vivo* (Tarassov *et al.*, 2008; Rochette *et al.*, 2015). Because DHFR-PCA readout is cell growth, the amount of protein complex formed correlates with colony size, allowing for the detection of quantitative changes in PPIs (Freschi *et al.*, 2013; Rochette *et al.*, 2015). We designed an array of 575 putative preys previously reported to interact physically with at least one SH3-containing protein, in addition to 40 negative controls (for a total of 615 unique preys) (Stark *et al.*, 2006). Each bait-prey pair was assayed in quadruplicate. PPI detection was highly reproducible between replicates (Pearson's correlation = 0.91-0.93). We identify 607 distinct PPIs (out of 13,530 possible PPIs) (Figure 2A and Data S2). None of the negative controls display PPIs above the significance threshold (cf. methods). About 33% (202/607) of the PPIs were observed before, mostly by direct methods such as Y2H (Figure S2A-B and Data S2) (Stark *et al.*, 2006). The 405 new PPIs we report expand the yeast SH3-containing protein interactome by ~30% of what is currently described in databases (Stark *et al.*, 2006). The PPI network is overrepresented with proteins involved in actin cytoskeleton-dependent processes, in agreement with previous reports (Boyle *et al.*, 2004; Mirey *et al.*, 2005) (Data S2).

To identify SH3-dependent PPIs, we repeated the DHFR-PCA experiments with baits in which each of the SH3s were individually removed (SH3-deleted baits) and replaced with a flexible linker (i.e. stuffer DNA) by genome editing. We identify 256 PPIs (171 unique PPIs corresponding to 28% of the WT SH3-containing proteins interactome) that are significantly weaker or completely lost upon SH3 deletion (Figure 2A), showing that they are SH3-dependent. We validated the vast majority of the quantitative changes detected following SH3 deletions by low-throughput liquid DHFR-PCA experiments (Figure S2C-D). We also verified that most of the PPIs lost are not due to changes in protein abundance by confirming that SH3-deleted baits have expression levels similar to WT (Figure S2E). Forty percent of these SH3-dependent PPIs were previously reported with the WT SH3-containing proteins and are thus high-confidence PPIs (Figure S2A-B). By examining a manually-curated list of yeast SH3-dependent PPIs, mostly determined by *in vitro* assays, we find that 23% (58/256) of the PPIs that are lost or decreased following SH3-deletions in our assay were previously reported (Kelil *et al.*, 2017). For instance, we find that Sho1-Pbs2 and Nbp2-Pbs2 are SH3-dependent PPIs, as previously described (Zarrinpar, Park and Lim, 2003; Mapes and Ota, 2004) (Figure S2D). Enriched biological processes in the network of SH3-dependent PPIs are in agreement with previously reported functions such as cytoskeleton binding and protein complex formation (Data S2) (Boyle *et al.*, 2004; Mirey *et al.*, 2005).

Interestingly, 37 PPIs are stronger (6% of the WT SH3-containing PPIs) and 75 are gained following SH3 deletions, suggesting that SH3s also preclude interactions *in vivo* (Figure 2A). Of these 112 PPIs, 22 were previously reported, indicating that many of the preys gained following SH3 deletions were already interacting directly or indirectly with the WT SH3-containing proteins (Figure S2A-B). These are enriched for PPIs reported by AP-MS, which is known to report many indirect interactions such as those found in large protein complexes (Figure S2B). SH3 domain deletion therefore likely changes the configuration of the protein within a complex or its relative binding preference, making these PPIs now detectable by DHFR-PCA. The increase in PPIs following SH3 deletions may also be explained by changes in intramolecular interactions, as

reported for the human SH3-containing SRC-family of tyrosine kinases (Engen *et al.*, 2008; Meng, Pond and Roux, 2017).

Our analyses confirm that SH3 binding motifs are enriched within SH3-dependent interaction partners. We used position weight matrices (PWMs) of *in vitro* SH3-specific binding sites (Tonikian *et al.*, 2009) to scan sequences of PPI partners (preys) that were weaker or lost following SH3 deletion. PWM peptide motifs are more frequent in the SH3-dependent PPI partners when compared to sequences from random proteins, SH3-independent PPI partners and PPI partners that are stronger or gained following SH3 deletions (p-value = 1.5×10^{-17}, p-value = 0.0038 and p-value = 2.1×10^{-08}, respectively; Mann-Whitney test, one-tailed) (Figure 2B). Intriguingly, ~50% of PPIs that are weaker or lost upon SH3 deletions do not display the *in vitro* motif targeted by the domain, highlighting the possibility that a significant fraction of SH3 domains do not bind *in vivo* to the canonical motifs that were identified *in vitro* (Figure 2B). In addition, no difference in motif enrichment is observed when SH3 PWMs are assigned randomly to the SH3s (p=0.25, Mann-Whitney test, two-tailed, Figure 2C), indicating that the SH3-specific binding motifs determined *in vitro* can partially identify SH3-bound proteins but poorly discriminate among different SH3 domains *in vivo* (Figure 2C). This observation could be explained in part by the overlap among SH3 binding motifs sequences (Tonikian *et al.*, 2009; Ivarsson and Jemth, 2019).

We next examined if changes in PPIs caused by the loss of SH3s impact cellular phenotypes. The growth of WT, SH3-deleted and knockout strains was measured under 51 different stress conditions ranging from DNA damage induction to high osmolarity (Figure 2D-E and Data S2). In most cases, gene or SH3 deletion leads to subtle phenotypes (Figure 2D), as expected given that these genes are non-essential (Giaever *et al.*, 2002; Giaever and Nislow, 2014). Nonetheless, when gene deletion results in strong phenotypes, SH3 deletion leads to similar growth defects (e.g. *NBP2*, *BBC1*, *BEM1* and *SLA1*), highlighting the domain's importance for protein function (Figure 2D). The deletion of SH3s mediating a large number of PPIs generally leads to drastic new sensitivities under multiple conditions (Figure 2D-E). Indeed, the number of SH3-dependent PPIs in a SH3-deleted strain is negatively correlated with its growth score relative to the WT strain (Pearson's correlation = -0.426, p-value = 0.038). For example, the deletion of Sla1 SH3-1 or Sla1 SH3-2 has a more severe impact on growth compared to the loss of SH3-3, which displays remarkably less SH3-dependent PPIs (Figure 2A and 2E). Interestingly, a subset of SH3 deletions leads to sensitivities under very specific conditions. For example, the Bem1 SH3-2-deleted strain is more sensitive to dimethyl sulfoxide (DMSO) than the WT. Considering that DMSO increases yeast phospholipid content (Murata *et al.*, 2003), this sensitivity is consistent with our observation that many PPIs that depend on Bem1 SH3-2 are phospholipid binding proteins (8/18) (Data S2). Overall, our results highlight that a large-fraction of PPIs depend on SH3 domains *in vivo*.

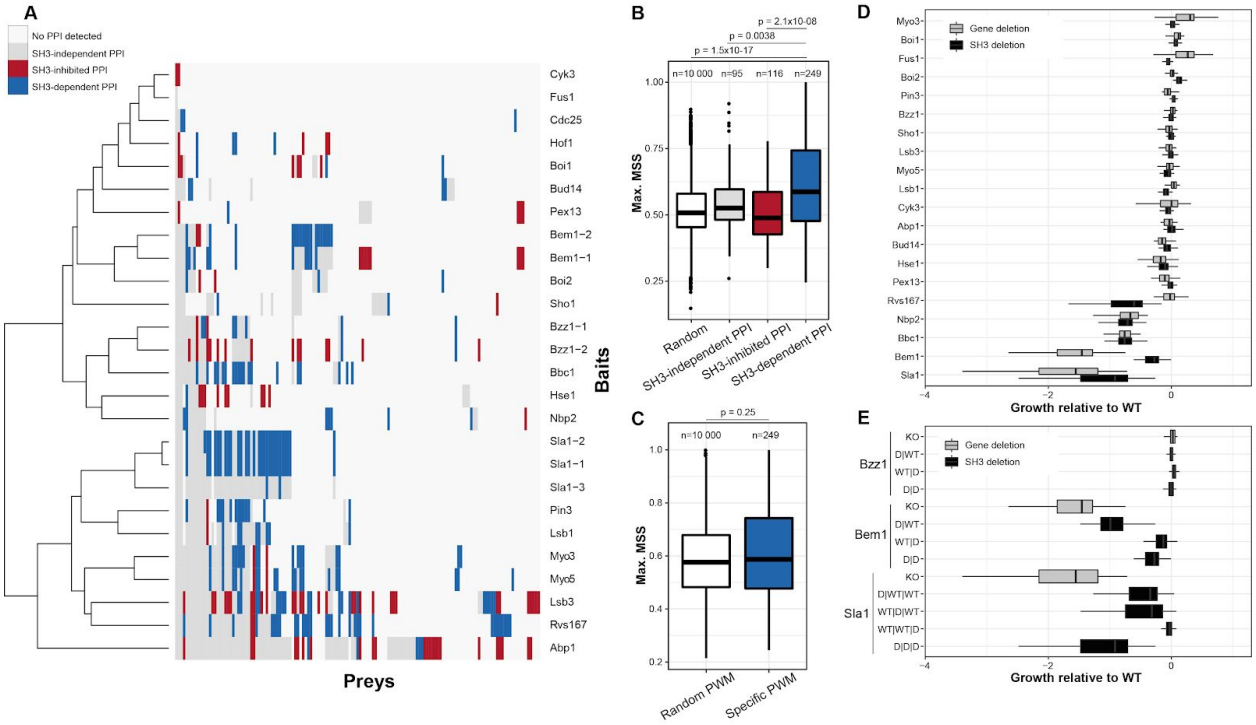

Figure 2: Identification of SH3-dependent and SH3-independent PPIs *in vivo* and their functional impact. A) PPIs detected by DHFR-PCA for WT and SH3-deleted proteins. The colors represent different types of PPIs. Each PPI was assessed in quadruplicate. B) Matrix similarity scores (MSS) of SH3-specific Position Weight Matrices (PWMs) for preys corresponding to different types of PPIs from A). SH3-dependent PPIs are enriched for SH3 binding motifs relative to the Random, SH3-independent, and SH3-inhibited PPIs ($p = 1.5x10^{-17}$, $p = 0.0038$ and $p = 2.1x10^{-08}$, respectively; Mann-Whitney test, one-tailed). C) PWM MSS for sequences of SH3-dependent interactors (as in 2B), and also for PWMs randomly assigned to different SH3s ($p = 0.25$, Mann-Whitney test, two-tailed). D) Growth of gene- or SH3-deleted strains compared to the corresponding WT under 51 stress conditions. For Bzz1, Bem1 and Sla1, the data shown is for the deletion of all SH3s. E) Same as D) but for mutants with individual or combinations of SH3 deletions. D represents deleted SH3, i.e. replaced with a linker sequence. Each strain was grown in twelve replicates per condition (D-E). All the data represented in this figure is listed in Data S2, including the names of the preys in panel A). See also Figure S2.

SH3 domains are rarely sufficient to dictate binding specificity of their host protein *in vivo*

To investigate whether SH3s are sufficient to bind specific PPI partners irrespective of their host protein, we exploited the DNA stuffer introduced at the Abp1 SH3 locus (Biot-Pelletier and Martin, 2016). We performed a second round of genome editing to individually replace Abp1 SH3 with 30 exogenous SH3s (i.e. domain swapping, Figure 1A). We selected Abp1 as a model because it nucleates a large number of PPIs (80 PPIs, 28 PPIs affected by the SH3 deletion), has a single SH3 with a conserved position at the protein C-terminus (Figure S1B), and is part of the actin protein interaction network that is typical for SH3-containing proteins. We replaced the Abp1 stuffer with its own SH3 as a control ($Abp1_{SH3}$ in Abp1), which shows an interaction profile almost identical to that of the non-edited WT strain (Kendall's $\tau = 0.93$, Figure 3A). We also introduced the 26 other yeast SH3s, SH3s from Sdc25 (annotated as a pseudogene in *S. cerevisiae* reference strain but functional in related Fungi) and Abp1's three human orthologs

CTTN, HCLS1 and DBNL. For cases where Abp1 loses most of its PPIs, we verified that Abp1 expression level was not significantly affected (Figure S3A).

The impact on Abp1 PPIs of exogenous yeast SH3s swapping into Abp1 partly correlates with SH3 amino acid sequence identity (Figure 3A-S3B and Data S3). For instance, the insertion of SH3s with low sequence identity, such as Bem1 SH3-2 (24%) and Cdc25 (22%), results in the loss of almost all of Abp1 PPIs. In contrast, Abp1 swapped with SH3s from its human orthologs, which have the highest sequence identity (human orthologs 46 to 52%, other yeast SH3s 19 to 41%), displays PPI profiles more strongly correlated to WT Abp1 than most of the other configurations (Figure 3A-S3B) (Kendall's τ: CTTN = 0.89, DBNL = 0.89 and HCLS1 = 0.86). In addition, we observe a significant correlation (cophenetic correlation = 0.20, p-value = 0.005, Figure 3B) between the similarity of PPI profiles (Figure 3A) and sequence similarity of the SH3 domains (Figure S1A). This result is illustrated by the co-clustering of interaction profiles of Abp1 swapped with SH3s from paralogous proteins (Boi1-Boi2, Myo3-Myo5, Lsb1-Pin3, Lsb3-Ysc84) (Figure S1A and 3A). Interestingly, we detected no correlation when comparing the clustering of Abp1 SH3 swapped PPI profiles with the similarity of the *in vitro* PWMs of each SH3 domain (p-value = 0.32, from permutation). The lack of correlation suggests that the previously determined binding motifs for yeast SH3s account poorly for the interaction profiles of the different Abp1 SH3 swapped proteins. This analysis indicates that the exogenous SH3 protein context (i.e. Abp1) has a greater influence on the binding specificity than the SH3 binding motifs derived from *in vitro* data.

The partial rescue of Abp1 PPIs by its human orthologous SH3s suggests that the binding specificity is at least partially conserved across more than a billion years of evolution despite relatively little sequence conservation (46-52%). However, a subset of SH3-dependent PPIs is only observed with the endogenous domain. For instance, Hua2 and App1, both well characterized partners of Abp1 (Drees *et al.*, 2001; Fazi *et al.*, 2002; Ho *et al.*, 2002; Tong *et al.*, 2002; Landgraf *et al.*, 2004; Yu *et al.*, 2008; Tonikian *et al.*, 2009) are only detected when Abp1 contains its own SH3 (WT or reintroduced). Interestingly, several SH3s inhibit most of Abp1 PPIs, including SH3-independent interactions (Figure 3A-S3B). It is possible that these exogenous SH3s create intramolecular interactions within their new host protein that prevent PPIs with SH3-independent partners. They could also nucleate new PPIs undetected in our DHFR-PCA assay but preventing interactions with established WT Abp1 partners. Overall, two thirds (21/30) of the SH3s swapped into Abp1 lead to at least 1 PPI not previously detected with WT Abp1. However, these gained PPIs are rarely mediated by SH3 intrinsic specificities, as they were not initially identified in our SH3 deletion screen as being SH3-dependent (Figure 3C). We conclude that most SH3s are not sufficient to bring their endogenous specificity to a new protein context (Figure S3C). Consistent with these results, protein partners gained by Abp1 are not enriched for SH3-specific PWMs relative to unaffected PPIs (p=0.52, Mann-Whitney test, one-sided) (Figure S3D). Exceptions are $Sho1_{SH3}$ and $Nbp2_{SH3}$ in Abp1 leading to an increase in Abp1 association with Pbs2, which binds both Sho1 and Nbp2 in a SH3-dependent manner (Zarrinpar, Park and Lim, 2003; Mapes and Ota, 2004) (Figure S2B and Data S2). The observation that SH3s are rarely autonomous in dictating their host PPI partners, but rather alter PPIs in a manner that cannot be predicted from their intrinsic specificity suggests the presence of complex interactions between a SH3 domain and its host protein.

We further investigated the impact of Abp1 SH3 swapping by measuring growth under stress conditions, as described above. Domain swapping does not lead to significantly enhanced sensitivities to the different conditions tested (Figure S3E). However, we observe some correspondence between PPIs, SH3 sequence similarity and growth profiles. For example, the WT Abp1 growth profile clusters with the control $Abp1_{SH3}$ in Abp1 and with Abp1 containing its closest human orthologous SH3 from CTTN (Figure S3E). In addition, $Myo3_{SH3}$ or $Myo5_{SH3}$ in Abp1 co-cluster as they do in the PPI similarity cluster (Figure 3A and S3E). One strong phenotype obtained from *ABP1* deletion is an increased hygromycin resistance (Barreto *et al.*, 2011). We show that this phenotype is dependent on Abp1 SH3 (Figure S3F), as SH3 or gene deletion lead to the same level of resistance. None of the swapped SH3s, except for the $Abp1_{SH3}$ in Abp1 control, fully reproduces the WT sensitivity to hygromycin, confirming our observation with PPI patterns (Figure S3F). However, a subset of swapped SH3s display intermediate phenotypes, including the orthologous human SH3s $CTTN_{SH3}$, $HCLS1_{SH3}$ and $DBNL_{SH3}$ in Abp1; this is consistent with PPI profiles and sequence similarity clusters (Figure 3A, S3F). Our results show that the functions of Abp1 strongly depend on its endogenous SH3 domain and associated PPIs, notwithstanding that many of Abp1 PPIs are not SH3-dependent.

We next examined whether Abp1 SH3 is autonomous for establishing its own PPI specificity when swapped into a new host protein. We introduced Abp1 SH3 at each of Sla1's three SH3 positions, individually (Fig 3D). Inserting Abp1 SH3 at either of the first two positions only slightly affects Sla1 PPIs. Interestingly, SH3 swapping at the third position led to the detection of seven new PPIs, although few Sla1 partners depend on the SH3 at this position (Figure 2A and 3D). None of the PPIs gained by Sla1 were originally found to be dependent on Abp1 SH3, further supporting our finding that the ability of a domain to dictate the PPI partners of its host protein is highly dependent on the identity of the host. Overall, the SH3 swapping results suggest that SH3 binding specificity and/or affinity is generally dictated by host protein sequences and by SH3 positioning within the host protein, thus uncovering possibly complex assemblies.

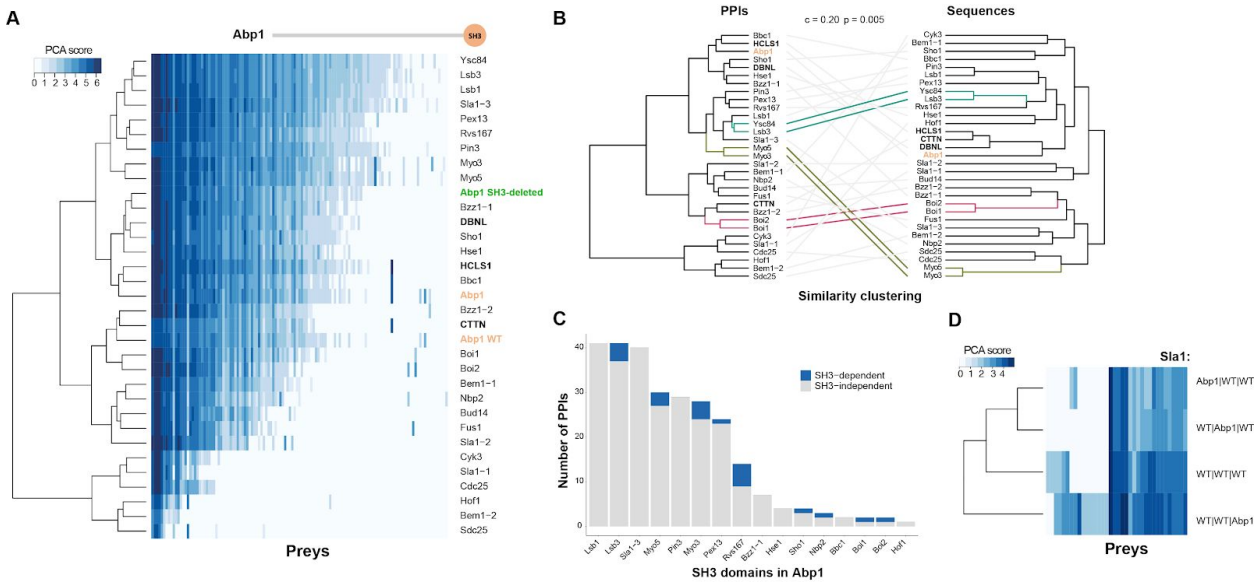

Figure 3: Swapping SH3 domains in a new host protein shows that they are rarely sufficient to bring the specificity of their endogenous host protein. A) PPIs of Abp1 with its endogenous SH3 domain and with domains from other yeast proteins and its human orthologs. Blue shades correspond to PPI interaction scores measured by DHFR-PCA (PCA score). A scaled cartoon representation of Abp1 is shown above the heatmap. Bold SH3 domains represent human orthologous, orange SH3s are controls, and green is for the SH3-deleted protein. Each PPI was assessed in quadruplicate. B) Cophenetic correlation between Abp1 SH3 swapped interaction clusters with the SH3 domain sequence similarity clusters. The empirical p-value (p = 0.005) was obtained from permutation. C) Number of PPIs gained by Abp1 upon domain swapping. PPIs originally detected as dependent on the domain when present in its host protein are shown in blue. D) Sla1 PPIs with the SH3 from Abp1 inserted at each of the three SH3 positions. Each SH3 is either the WT Sla1 domain (WT|WT|WT) or is Abp1 SH3 (Abp1|WT|WT is Sla1 with its SH3-1 swapped with Abp1 SH3). PPIs were measured in quadruplicate. See Data S3 for A-C), and see Data S5 for D). See also Figure S3.

Abp1 SH3 domain sequence can affect both SH3-dependent and SH3-independent interactions

Abp1 SH3 swapping revealed that PPIs can be affected by the replacement of the endogenous domain by a new SH3, even for SH3-independent PPIs. These results indicate that the sequence composition of a SH3 can alter PPIs in ways that may not depend directly on canonical SH3-ligand interactions. To evaluate the contribution of each amino acid position in an unbiased manner, we performed an *in vivo* deep mutagenesis scan (DMS) on Abp1 SH3 and then examined Abp1 association with SH3-dependent interaction partner Hua2 (Figure 4A) and SH3-independent partner Lsb3 (Figure 4B). Each residue of the SH3 genomic sequence was mutated to every other possible amino acid using genome editing. The effect on both PPIs was measured in living cells by DHFR-PCA (Figure 4A-B and S4A-B).

The impact of mutations on binding is detected (Figure 4A-C and S4C) by the enrichment of alleles surviving or being eliminated during the DHFR-PCA selection (Diss and Lehner, 2018), resulting in a DMS score that represents the impact of individual mutations. Using synonymous mutations as a reference, we identified a majority of amino acid residues that significantly decrease binding and a few that increase it (Figure 4A-C). For Abp1-Hua2 PPI, there is a strong correlation between the results of a smaller scale validation experiment and DMS assays,

confirming the reproducibility of the DMS approach (Figure S4A, B and D). The profiles of mutation sensitivity for the two PPI partners are overall highly correlated (Kendall's τ = 0.59, p-value = 1.0×10^{-201}) but also show significant differences (Figure 4A-C). Some positions are sensitive to any mutation for both PPIs (e.g. L17), while others are specific to one PPI (e.g. D32). We found that there is a strong overlap between regions where mutations affect both PPIs and those that are highly conserved across SH3s, for instance residues W35, G47, and P50, which are key to SH3 functions (Larson and Davidson, 2000) (Figure 4A-B). In addition, buried sites tend to be more sensitive to mutations when compared to solvent-exposed positions, for both SH3-dependent and -independent PPIs (Figure S4E). Destabilising mutants are enriched for variants predicted to be deleterious on the basis of Abp1 sequence conservation (Hua2: p-value = 2.5×10^{-30}; Lsb3: p-value = 2.8×10^{-16}, Fisher's exact test) and structure (Hua2: p-value = 1.3×10^{-45}; Lsb3: p-value = 2.3×10^{-58}, Fisher's exact test), as previously estimated by (Wagih *et al.*, 2018). Only a subset of destabilising variants (24.9% for Abp1-Hua2 and 26.5% for Abp1-Lsb3) were not predicted to be deleterious on the basis of Abp1 structure or sequence conservation (Wagih *et al.*, 2018).

Using silent substitutions as neutral sets, non-silent mutations can be divided into four main categories based on their DMS scores: (1) destabilizing both PPIs, (2) specifically destabilizing Hua2 PPI or (3) Lsb3 PPI and (4) neutral or slightly increasing binding. These categories correspond to the different quadrants of Figure 4C. We found that mutations affecting both PPIs correspond to buried residues that are distal to the Abp1-bound peptide and are likely affecting protein folding rather than residue-residue interactions (Stollar *et al.*, 2009) (Figure 4C-E). Mutations that specifically destabilize Abp1-Lsb3 tend to have smaller effects on the interaction (DMS scores closer to 0) and are distant from the SH3 binding peptide (Figure 4A-D). However, mutations that are specifically destabilizing the Abp1-Hua2 PPI have a generally greater impact on the interaction (DMS scores more negative) and are more likely to lie at the interface with the peptide. Such interface residues include Y7, E13, E16, V31, D32, D34, L48 and Y53, supporting their specific role in binding (Figure 4A-D and S4F). Interestingly, a subset of mutations at positions 15 and 52, two positions in contact with the SH3 target peptide, are specifically enriched in the Abp1-Hua2 PPI, suggesting that this interaction could in principle have higher affinity (Figure 4A-B and S4F). This suboptimal affinity could be important for specificity and fitness, as it was previously reported for Sho1-Pbs2 SH3-dependent PPI (Zarrinpar, Park and Lim, 2003). Very few of the positions specifically affecting the Abp1-Hua2 PPI are conserved between Abp1 SH3 and other yeast SH3s. This explains in part why none of the other SH3s can complement the loss of Abp1-Hua2 interaction upon domain swapping (Figure 3A).

Overall, DMS analysis of Abp1 SH3 PPIs helps discriminate the residues defining SH3 binding specificity from the positions regulating the core function of the domain. The DMS results explain in part how replacing Abp1 SH3 with other SH3s resulted in the loss of PPIs that were found to be Abp1 SH3-independent. Interface residues may specifically affect the binding of the host protein with its partners that are SH3-dependent. In addition, other residues may affect PPIs in a way that is independent from the intrinsic binding specificity of the domain but involves other interactions within the host protein.

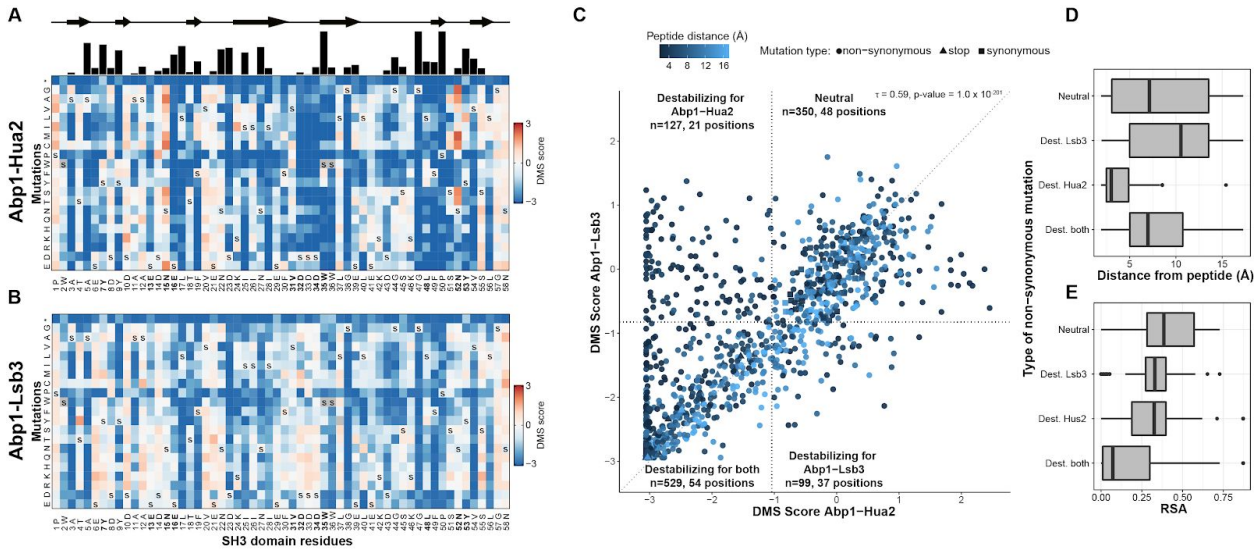

Figure 4: Comprehensive mutagenesis of Abp1 SH3 domain illustrates the contribution of its amino acid composition to SH3-dependent and -independent PPIs. A-B) Impact of mutations on binding represented in terms of Deep Mutagenesis Scanning (DMS) score i.e. log2 ratio of mutant counts in the test condition (DHFR-PCA) relative to the reference condition. A) Tolerance to amino acid mutations of Abp1 SH3 for the SH3-dependent interaction with Hua2. Shades of blue indicate low DMS score i.e. reduced interaction strength and shades of red indicate increased interaction strength. The horizontal black line above represents the secondary structure of Abp1 SH3. The black bars on the top of the heatmap represent the level of conservation of the positions across the 27 *S. cerevisiae* SH3 domains. Positions in bold represent residues in contact with the SH3 ligand (see Figure S4F). B) Same as A) but for the SH3-independent PPI Abp1-Lsb3. The two experiments were performed in two biological replicates and the data represents the average. C) DMS scores for the two PPIs. DMS scores below the 1st percentile of synonymous variant scores at the codon level (n=120) are defined as deleterious variants. This percentile is shown on the plot as dashed lines for each PPI. The four quadrants defined by these lines are referred to in the text as the four categories of mutations. The minimum distance of each position relative to a binding peptide is shown in blue. Distances were calculated from the resolved structure of Abp1 SH3 in complex with a peptide of Ark1 (PDB: 2RPN, (Stollar *et al.*, 2009)). D) Mutations from each quadrant are represented based on their minimum distance to the Abp1-bound peptide. Dest is for destabilizing mutations. E) The relative solvent accessibility (RSA) of the different sensitive positions is represented. See also Figure S4 and Data S4.

SH3 domain positions in multi-SH3 proteins affect their PPIs and *in vivo* functions

Proteins containing multiple SH3s can mediate the formation of multivalent interactions, bringing an additional level of complexity to PPI regulation. Only three yeast proteins have multiple SH3s, namely Bem1, Bzz1 and Sla1 (Figure 1B). We focus on Sla1, a cytoskeleton binding protein that has three SH3s displaying low sequence identity (Figure 1B and S1A). We identified significant differences in PPIs that depend on each of Sla1 SH3s (Figure 2A), as well as when comparing the impact of individual SH3 deletions on growth phenotypes (Figure 2E). The difference between Sla1 SH3-1/SH3-2 and SH3-3 is particularly striking.

To investigate whether the function of each of Sla1 SH3s depend on their position within the protein, we performed domain shuffling by constructing all possible domain-position permutations within Sla1, and measured the impact on PPIs as described (Figure 1A and S5A). Should Sla1 SH3s behave as beads on a string independently of one another, we would expect SH3 shuffling not to affect Sla1 functions. In addition, because Sla1 SH3-1 and SH3-2 bind to

the same peptide motifs *in vitro*, they are expected to be interchangeable if intrinsic specificity is their major contribution to binding (Tonikian *et al.*, 2009). Strikingly, Sla1 PPIs are highly dependent on SH3 positions (Figure 5A and S5A). Many PPIs are lost when the SH3-1 position is occupied by SH3-2 (2|2|3) but maintained with SH3-3 (3|2|3) (Figure 5A), even if there are few PPIs that depend on SH3-3 in WT Sla1 (Figure 2A and 5A). Remarkably, shuffling two or all three Sla1 SH3s leads to the loss of most Sla1 PPIs, despite that all three SH3s are still present albeit in a different order (Figure S5A). For example, shuffling the first two domains (2|1|3) results in the loss of ~80% of Sla1 PPIs (Figure S5A). This is surprising considering that Sla1 SH3-1 and SH3-2 bind to the same peptide motifs *in vitro* (Tonikian *et al.*, 2009); these two SH3s can directly bind three proline-rich motifs in any orientation (type 1 and 2) present on Las17 (Feliciano and Di Pietro, 2012). However, shuffling SH3-1 with SH3-2 completely disrupts Sla1 association with Las17, suggesting that these SH3s do not behave independently but rather depend on their position *in vivo* (Figure S5A and Data S5). To further assess the impact of Sla1 SH3 shuffling, we examined the consequences of these changes on growth. Sla1 SH3 shuffling leads to severe growth defects under multiple conditions. The similarity of PPI profiles among mutants is significantly correlated with growth phenotypes (cophenetic correlation = 0.28, p-value = 0.00002, from permutation) (Figure S5B). These results indicate that the position of Sla1 SH3s is important for most of its biological functions and that domains binding the same peptides *in vitro* are not strictly interchangeable *in vivo*.

We extended these analyses by directly examining at the single-cell level Sla1's specific function in clathrin-mediated endocytosis (Figure 5B) (Feliciano and Di Pietro, 2012; Mund *et al.*, 2018; Tolsma, Cuevas and Di Pietro, 2018; Manenschijn *et al.*, 2019). During endocytosis, Sla1 plays a central role as an adaptor protein linking cargos to clathrin, in addition to participating in the recruitment of the actin machinery to internalization sites via SH3-dependent PPIs (Figure 5B) (Feliciano and Di Pietro, 2012; Tolsma, Cuevas and Di Pietro, 2018). Sla1 appears as compartmentalized into foci at the plasma membrane where it is present until cell membrane invagination is complete. During this process, vesicles are fully internalized following scission, before complex disassembly. It was previously reported that Sla1 deletion results in altered endocytosis dynamics and defective actin polymerisation (Holtzman, Yang and Drubin, 1993; Kaksonen, Toret and Drubin, 2005; Feliciano and Di Pietro, 2012; Tolsma, Cuevas and Di Pietro, 2018). We found that Sla1 SH3 shuffling alters its interaction with partners that are themselves associated with different phases of endocytosis, such as clathrin (Chc1), Las17, Pan1, End3 and subunits of the Arp2/3 complex (ARC proteins) (Figure 5C). We performed live fluorescent imaging of individual yeast cells to follow endocytosis dynamics by tagging WT and mutant Sla1 alleles with GFP (Figure 5D and S5C). This approach allowed measuring the distance traveled by Sla1-GFP-labeled vesicles during endocytosis, as foci moved from the cell periphery towards the cell center before their disassembly (Figure 5D and S5C).

We found that Sla1-GFP foci movements toward the center of the cell (up to ~80 nm in 40 seconds) before vesicle disassembly is mostly regulated by Sla1 two N-terminal SH3s, as previously reported (Figure S5D) (Rodal *et al.*, 2003; Feliciano and Di Pietro, 2012). Remarkably, shuffling Sla1 SH3s in any combination results in a drastic decrease in the effective distance traveled by Sla1-positive foci, as well as their persistence in time, likely due to the inability to complete the endocytic process (Figure 5E-F and S5E-G). This observation is consistent with a >50% decrease in the number of Sla1-GFP-labeled particle internalizations per

cell and an increase in the number of incomplete events, i.e. events that were not completed before the end of the image acquisition period (Figure 5G and S5E). In addition, the linearity of Sla1-GFP foci trajectories is drastically perturbed in all Sla1 SH3-shuffled strains, suggesting inefficient vesicle internalization (Figure S5F). The observed phenotypes could be linked to the loss of Sla1 PPI with partners that are important for the stimulation of actin nucleation and polymerisation, such as Myo5 and Las17 (Figure 5C). Overall, our results demonstrate that the identity and position of each of Sla1 SH3 domain regulate its PPIs. In turn, PPI alterations quantitatively impact the cell ability to tolerate stress and dynamic processes such as endocytosis, demonstrating the important contribution of SH3 positioning to protein functions.

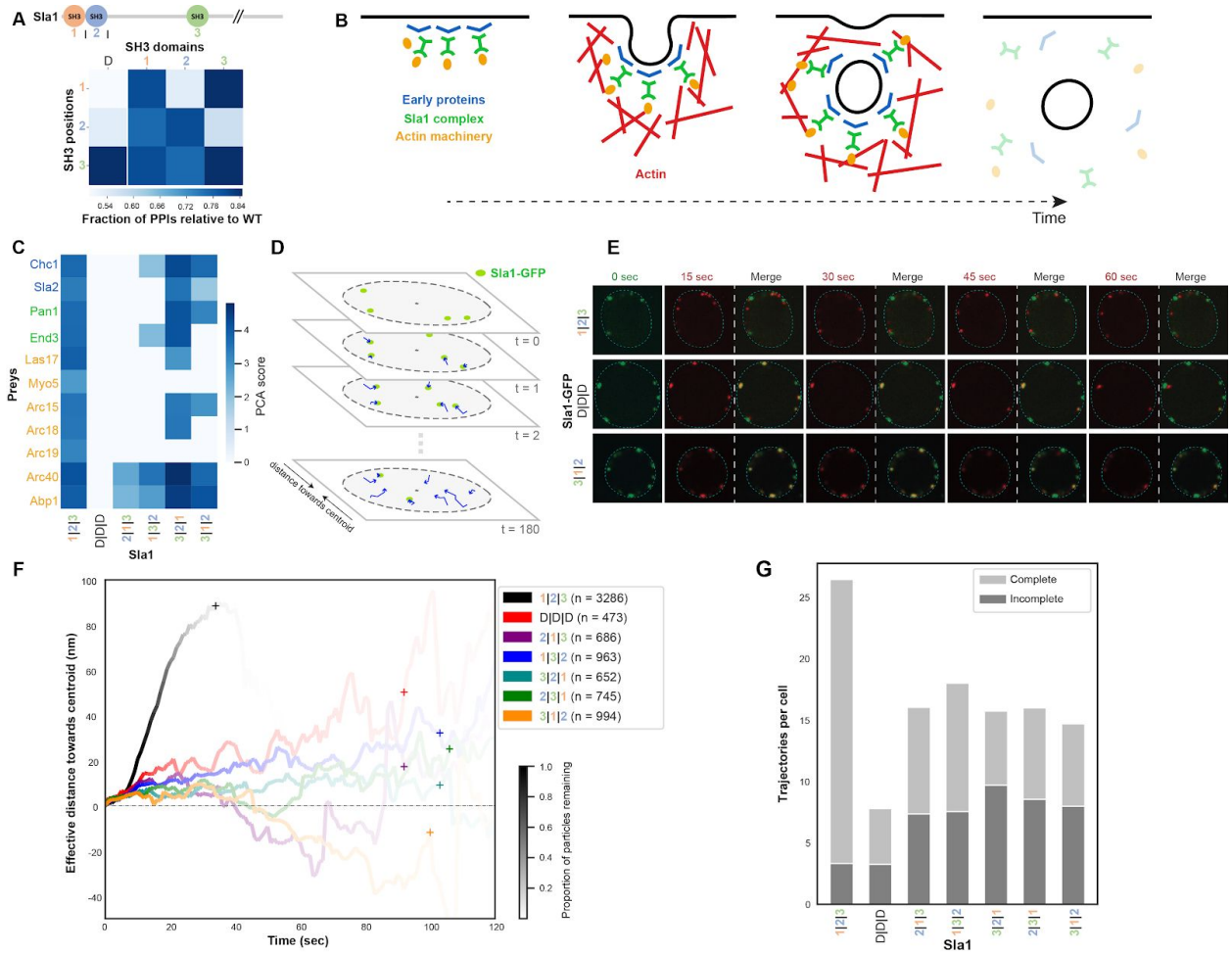

Figure 5: Sla1 SH3 domain shuffling greatly alters its interactome *in vivo* with strong impacts on endocytosis. A) PPIs detected by DHFR-PCA relative to Sla1 WT for SH3-shuffling or -deletion per SH3 position. PPIs were measured in quadruplicate. B) Scaled cartoon representation of Sla1 and illustration of Sla1 function as an adaptor linking early proteins with the actin machinery in yeast clathrin-mediated endocytosis. C) PPIs of Sla1 SH3-shuffled or -deleted with partners implicated in clathrin-mediated endocytosis. Each PPI was assessed in quadruplicate by DHFR-PCA. The color code represents the strength of the PPIs detected by DHFR-PCA (PCA score). Sla1 1|2|3 represents the WT protein. D) Schematic representation of Sla1 foci assembled at the cell membrane and their movement toward the center of the cell during internalization before disassembly, in time. E) Representative fluorescence microscopy images of cells expressing different Sla1-GFP proteins at multiple time points. Foci from WT Sla1-GFP, the negative control D|D|D and a Sla1-GFP shuffle (3|1|2) are shown, in time. The first time frame is green and the others are red. A merge with the starting point is shown for each time frame. F) Average effective distance travelled by Sla1-GFP

particles towards the centroid of the cell through time for WT Sla1 (1|2|3), SH3-deleted control or shuffled strains. Color transparency represents the proportion of events that are not completed yet that are used to calculate the average distance at the different time points. Plus signs (+) represent the moment in time when 95% of the foci have disassembled. G) Average number of complete or incomplete Sla1-GFP endocytosis events per cell is shown. See also Figure S5 and Data S5.

PPI specificity is also regulated by the positioning of SH3 domains in multi-SH3 containing proteins in human cells

Multi-SH3 domain proteins are more prevalent in metazoans (62/216 SH3-containing proteins in humans) than in yeast (3/23). NCK adaptors (NCK1 and NCK2) are archetypal examples of human proteins exclusively composed of modular binding domains such as SH3s (Buday, Wunderlich and Tamás, 2002; Chaki and Rivera, 2013) (Fig 6A). As for yeast Sla1, NCKs are involved in signaling pathways modulating the actin cytoskeleton. NCK proteins contain a phosphotyrosine-binding SH2 module, adding an additional layer of regulation to their three SH3 domains. NCK SH3s bind to similar types of peptide motifs but are known to interact with different partners. For example, NCK2 SH3-2 and SH3-3 bind to type II motifs but associate with different peptides *in vitro* while NCK2 SH3-1 interacts with type I/II binding sites (Liu *et al.*, 2006; Teyra *et al.*, 2017). To test whether SH3 shuffling impacts NCK2 PPIs in human HEK293T cells, we combined affinity purification with sequential window acquisition of all theoretical mass spectra (AP-SWATH) quantitative proteomics (Gillet *et al.*, 2012; Collins *et al.*, 2013; Lambert *et al.*, 2013).

We first generated a high confidence WT NCK2 interactome and validated that it significantly overlaps with previously reported PPIs (71%), displaying about a third of all reported NCK2 interactors (Stark *et al.*, 2006). The SH3-inactive (W38K/W148K/W234K) negative control displayed almost no PPIs, as expected. Overall, we identified 360 PPIs (FDR < 1%) across all constructs tested, 79 of which are non-redundant, including 39 PPIs not previously reported (Stark *et al.*, 2006). By comparing via AP-SWATH the abundance of PPI partners in the WT and SH3-shuffled NCK2 complexes (Figure 6A), we detect significant changes including the complete loss of up to 17 PPIs (30% of WT NCK2 PPIs) and the gain of up to 13 new partners (Figure 6A). Interestingly, NCK2 PPI profiles cluster based on the identity of the SH3 at the third position, suggesting its prevalent importance for the regulation of the NCK2 interactome (Figure 6A). A more detailed analysis of three well-characterized NCK2 partners that each associate with a specific SH3 reveals that shuffling alters their association with NCK2, a result that is mostly dependent on the position of each target cognate SH3 (Figure 6B). In particular, NCK2 SH3-2 association with PAK1, one of its best characterized PPI (Bokoch *et al.*, 1996; Bisson *et al.*, 2007), is significantly impaired when the domain is at the third position (Figures 6B, S6A-B). In contrast, NCK2 interaction with WASL, which is mediated by multiple SH3-PXXP interactions likely involving all three NCK2 SH3s, is only slightly altered by shuffling (Figure 6C) (Rivero-Lezcano *et al.*, 1995; Rohatgi *et al.*, 2001), suggesting that in this case shuffling may have led to combinations of gain- and loss-of-function with a net neutral result. Together, these results demonstrate that maintaining the position of human SH3 domains is necessary to preserve interaction profiles, in agreement with yeast experiments.

To test whether shuffling affects the intrinsic SH3 affinity to targets via a direct effect on peptide binding, we performed *in vitro* fluorescence polarization binding assays using purified full length

recombinant NCK2 and peptides from established direct SH3 partners CD3Ɛ (SH3-1) (Gil *et al.*, 2002) and PAK1 (SH3-2) (Bokoch *et al.*, 1996). Shuffling leads to modest variations in the dissociation constant (K_D) upon SH3-1 shuffling for the interaction with CD3Ɛ peptide, with a ~2-fold increase when SH3-1 was moved to the third position (2|3|1 or 3|2|1, Figure 6D and S6C). Similarly, NCK2 binding to its PAK1 target peptide decreases by ~2-3-fold when SH3-2 is inserted at the SH3-3 site (1|3|2 or 3|1|2) (Figure 6D and S6D). The same shuffling combinations lead to a more important decrease of PAK1 association with NCK2 in live cells (Figure 6A-B). These results indicate that shuffling SH3s in their host protein might only slightly affect the position and/or accessibility of their binding pocket. Interestingly, changing the positions of NCK2 SH3s results in the complete loss of a subset of PPIs in cells, even though the binding pocket of the domains appears to be accessible using purified proteins. The consequences of SH3 shuffling on NCK2 interactome are thus possibly caused by factors other than binding pocket availability *in vivo*.

We next assessed whether SH3 shuffling leads to longer-range disruptions to another protein interaction domain, namely NCK2 SH2. AP-SWATH analysis reveals that GIT1 association with NCK2 SH2 is significantly modulated in several SH3-shuffled mutants (Figure 6C) (Frese *et al.*, 2006). Consistent with this, P130CAS/BCAR1 association with NCK2 increases in shuffled mutants relative to WT (Figure S6E-F) (Schlaepfer, Broome and Hunter, 1997). These results indicate that SH3 positioning can also affect interactions that are mediated by other modular domains, suggesting the presence of intramolecular interactions and/or intricate associations in protein complex assemblies. Such indirect regulation of non-SH3 domains by adjacent SH3 positioning may also explain why many SH3-independent PPIs were altered by domain shuffling and swapping in our yeast experiments (Figure 3 and 5).

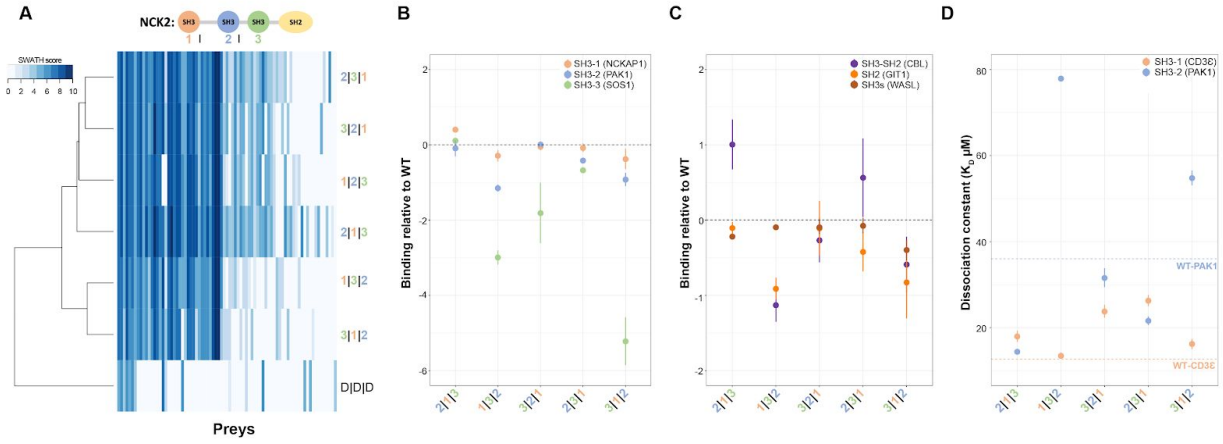

Figure 6: The NCK2 human interactome is modulated by SH3 domain shuffling. A-C) AP-SWATH quantitative MS data of NCK2 PPIs. All experiments were done in biological duplicates. True positive PPIs were determined using SAINT with four replicates of a FLAG-GFP construct as a negative control (FDR<1%). A) Scaled cartoon representation of NCK2 with its three SH3s and single SH2. Log2 of the average spectral count (SWATH score) for the different interaction partners of NCK2 SH3-shuffled proteins is shown. The different baits were clustered based on the similarity of their interaction profiles. The color code represents the PPI strength (SWATH score). The bait nomenclature 1|2|3 is for WT NCK2 and D|D|D is for the triple SH3-inactive negative control (3W/K). B-C) NCK2 PPI levels for selected partners with known binding sites on NCK2 are shown. The spectral count from the two biological replicates was compared to the WT NCK2 score (Ratio NCK2mut/NCK2WT) for each interaction. The log2 average

ratio is shown (WT NCK2 ratio = 0). Error bars indicate the SD. D) Fluorescence polarization *in vitro* binding assays of NCK2 full length recombinant proteins with a SH3-1 direct (CD3ε) or SH3-2 direct (PAK1) partner. Measured dissociation constants (K_D) are shown. Each binding assay was performed in independent triplicates. Error bars indicate the SD. See also Figure S6 and Data S6.

NCK2 SH3 domain positioning modulates its ability to promote phase separation *in vitro*

The separation of phases in the cytosol is a feature of multiple cellular processes. Phase separation mediated by NCK2's close paralog NCK1 was previously shown to depend on both its number of functional domains (Li *et al.*, 2012; Banjade and Rosen, 2014), and its capacity to self-associate (Banjade *et al.*, 2015). In the high-confidence SWATH dataset, many PPI partners affected by domain shuffling, such as CBL, SYNJ1, WIPF3, MINK1, GIT1 and GIT2 are predicted to undergo phase separation based on their amino acid sequences (Data S6) (Sun *et al.*, 2019). To better understand the biophysical consequences of NCK2 SH3 shuffling, we examined the ability of NCK2 SH3-shuffled proteins to undergo phase separation *in vitro* via self-association.

Remarkably, NCK2 SH3 domain positioning has a profound impact on its ability to promote phase separation (Figure 7A). At equivalent molar concentrations, the different NCK2 SH3-shuffled mutants initiate varying levels of phase separation (Figure 7A-B). The most striking differences occur when the SH3-2 is shuffled at the third position, which results in a significant ~2-fold increase in phase separation relative to WT (pairwise ANOVA, NCK2 1|3|2 p-value = 0.00067, NCK2 3|1|2 p-value = 0.0054, Figure 7A-B). Consistent with this observation, phase separation is inhibited with the addition of a PAK1 peptide, presumably by competing homotypic binding to NCK2 SH3-2 (Figure 7B). Based on previous reports on NCK1, we hypothesized that phase separation is dependent on NCK2 SH3-2 electrostatic interaction in *trans* with the first interdomain region of another NCK2 molecule (Banjade *et al.*, 2015). Consistent with this hypothesis, the propensity of the NCK2 SH3-shuffled proteins to form phase-separated condensates is altered by increasing NaCl concentration; constructs bearing the SH3-2 at the third position are the most resilient (Figure S7A). The addition of 1,6-hexanediol, a small aliphatic alcohol thought to disrupt the weak hydrophobic interactions often present in protein condensates (Kroschwald *et al.*, 2015), also disrupts all NCK2 SH3-shuffled proteins in a similar manner (Figure S7B). Taken together, these results demonstrate that NCK2 phase separation, which appears to be dependent on its electrostatic interaction with other NCK2 molecules, is highly dependent on the position of its SH3 domains. Strikingly, shuffling NCK2 SH3-2 at the third position highly disrupts NCK2 PPIs in cells (Figure 6A-C and S6A-B and S6E-F) while it stabilizes NCK2 condensates *in vitro*, suggesting that NCK2's capacity to phase separate might impact its potential to nucleate PPIs *in vivo*.

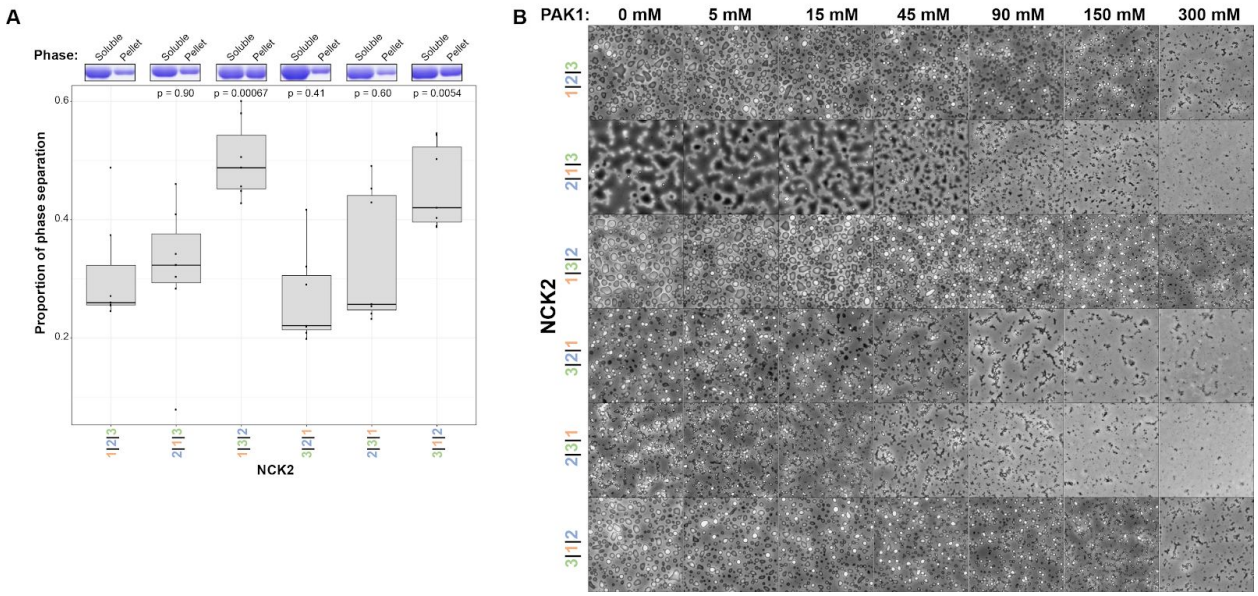

Figure 7: The ability of NCK2 to promote phase separation *in vitro* via self-assembly depends on the position of its SH3 domains. A) NCK2 phase separation was quantified following 24 hours of incubation. Soluble and phase-separated (pellet) proteins were resolved on 10% polyacrylamide gels and quantified via Coomassie staining. Coomassie staining of the different phases for one replicate is shown as an example above the total quantification box plot. Proportion of proteins in the pellet compared to the total protein content (soluble + pellet) is shown for every replicate. The experiment was repeated in seven replicates. Proportions of phase separated proteins were compared to NCK2 WT (pairwise ANOVA, p-values: 2|1|3 = 0.90, 1|3|2 = 0.00067, 3|2|1 = 0.41, 2|3|1 = 0.60 and 3|1|2 = 0.0054). B) An increasing amount of Pak1 peptide known to bind NCK2 SH3-2 was added, as indicated, to the different NCK2 samples and incubated for 24 hours before observing phase separation using phase contrast microscopy. The experiment was performed in three replicates. See also Figure S7 and Data S6.

Discussion:

PPI networks dictate how signaling pathways and biological processes are physically organized in the cell (Landry *et al.*, 2013). In human cells, it has been estimated that around half a million one-to-one PPIs occur (Stumpf *et al.*, 2008), and this figure largely ignores the fact that a given individual might express nearly one million distinct proteoforms (Aebersold *et al.*, 2018). It is therefore an outstanding challenge to understand how proteins find their cognate PPI partners while avoiding spurious interactions (Ladbury and Arold, 2012). The issue of specificity is especially important for domains binding to peptide sequences with recognition motifs that are highly degenerate and therefore present at high frequency throughout the proteome (Ladbury and Arold, 2012). This raises the question as to whether short domains such as SH3s encode enough information on their own to avoid spurious interactions. One possibility is that binding specificity could be largely enhanced by contextual specificity, which includes subcellular co-localization and co-expression of the interacting proteins. However, the contribution of SH3 domain protein context to the specificity of SH3-containing proteins *in vivo* remained to be quantified. Here, we combined genome editing and deep mutational scans with PPI network mapping and cellular phenotyping to elucidate the impact of protein context on SH3 domains *in vivo* functions.

We demonstrate that most SH3 domains mediate specific interactions that are dependent on their host proteins. We reveal that SH3s are often necessary but rarely sufficient for host proteins to target their cognate partners. This is supported by our observation that SH3s can bring unexpected PPI partners when swapped into new host proteins. We also show that the evolutionary conserved SH3 positioning is highly important for PPIs and functions of multi -SH3-containing proteins. It was recently predicted that intrinsically disordered regions such as linker sequences between modular domains can have altered conformations depending on their protein context (Mittal *et al.*, 2018). Overall, these observations imply that the description of protein domains as “beads on a string” is too simplistic and that the context of functional elements such as disordered regions and structured domains is of great importance. A more comprehensive view of proteins as a combination of functional regions co-evolving together to perform a specific task in cells will allow us to better understand their function, and to assess how binding specificity of short domains is tuned by their host protein sequence. In conclusion, our work yields new insights into the roles of SH3 domains *in vivo* and the contribution of protein context to SH3 PPIs and functions. We assert that our findings may be extended to other families of protein binding modules, and therefore contribute to a better understanding of complex signaling networks in normal and disease states. Our observations also suggest that the evolution of protein domains by gains and losses can have complex effects that will depend on the sequence of the domains themselves, their positions and how they interact within the host protein.

Acknowledgments

We thank Bisson and Landry lab members for discussions; Véronique L'Italien for expert assistance with figure artwork; N. Aubin-Horth, S. Elowe, I. Gagnon-Arsenault and C. Gagné-Thivierge for critical reading of the manuscript. This work was supported by a Canadian Institutes of Health Research (CIHR) Foundation grant 387697 and a HFSP grant (RGP0034/2018) to CRL, CIHR Project grant 162439 to NB, and Discovery Grant 1304616-2017 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to JPL. CRL holds the Canada Research Chair in Evolutionary Cell and Systems Biology. NB holds the Canada Research Chair in Cancer Proteomics. JPL is supported by a Junior 1 salary award from the Fonds de Recherche du Québec-Santé (FRQ-S). U.D. holds an Alexander-Graham-Bell Canada graduate scholarship and the Louis-Poirier scholarship of the Fondation du CHU de Québec-Université Laval. EB was supported by a NSERC graduate scholarship. RD is funded by a FRQ-S Postdoctoral fellowship. SD was supported by a MITACS globalink scholarship. PD is supported by a CIHR Vanier Canada Graduate Scholarship.

Author Contributions

UD, AKD, FC, RD, NB and CRL designed experiments. UD, EB, AKD, RD, SD, FC and GDG performed experiments. DB, PD, RD, SD and CRL designed and performed computational analyses. UD, EB, DB, RD, PCD, CRL, and JPL analysed results. UD, NB and CRL wrote the manuscript with contributions from all authors. NB and CRL supervised the research.

Declaration of Interests

The authors have no conflicts of interest to declare

METHODS

Saccharomyces cerevisiae strains

For the DHFR-PCA experiments, bait strains were constructed and preys were either constructed or retrieved from the yeast protein interactome collection (Horizon) (Tarassov *et al.*, 2008). BY4741 (MATa, *his3Δ leu2Δ met15Δ ura3Δ*) strains with a specific gene of interest fused at its 3' to the DHFR F[1,2] with a nourseothricin-resistance marker (NAT 100 µg/ml, Werner BioAgents) were used as baits. BY4742 (MATα, *his3Δ leu2Δ lys2Δ ura3Δ*) prey strains are each expressing a gene of interest fused at its 3' end to the DHFR F[3] and are resistant to hygromycin B (HPH 250 µg/ml, Bioshop Canada). The same BY4741 strains were used in the growth assays with the addition of knockout (KO) strains from the yeast knockout collection (Horizon) (Winzeler *et al.*, 1999; Giaever *et al.*, 2002; Giaever and Nislow, 2014). For the fluorescence microscopy experiments, the different Sla1 engineered proteins were constructed using the BY4741 Sla1-GFP strain from the GFP collection (Thermo Fisher Scientific) (Ghaemmaghami *et al.*, 2003; Huh *et al.*, 2003). See supplementary table Data S7 for the complete list of all strains used for this study.

Saccharomyces cerevisiae growth conditions

Cells were grown with their specific selection antibiotics in YPD medium (1% yeast extract, 2% tryptone, 2% glucose, and 2% agar (for solid medium)) and with the combination of NAT and HPH for diploid selection. Synthetic medium (0.67% yeast nitrogen base without amino acids and without ammonium sulfate, 2% glucose, 2.5% noble agar, drop-out without adenine, methionine and lysine, and 200 µg/ml methotrexate (MTX, Bioshop Canada) diluted in dimethyl sulfoxide (DMSO, Bioshop Canada)) was used for the DHFR-PCA experiments (PCA selection). For western blot experiments, cells were grown in synthetic complete medium (0.174% yeast nitrogen base without amino acids, without ammonium sulfate, 2% glucose, 0.134% amino acid dropout complete, 0.5% ammonium sulfate). Sla1-GFP strains were grown in synthetic complete medium without tryptophan to reduce background fluorescence. Cells were grown in synthetic complete medium with monosodium glutamate (0.1% MSG) instead of ammonium sulfate in hygromycin B growth experiments. See supplementary table Data S7 for all the growth media used.

Cells

Human embryonic kidney 293T cells (HEK293T) were grown in Dulbecco's Modified Eagle's medium (DMEM) high glucose complemented with 10% fetal calf serum under 5% CO_2 at 37°C. Transient expression of cDNA in cells was performed by polyethylenimine (PEI, Millipore Sigma) transfections. Enrichment of phospho-dependent PPIs was done by treating HEK293T cells with the tyrosine phosphatase inhibitor pervanadate, freshly made by mixing 4 mM orthovanadate (Millipore Sigma) with 30% H_2O_2 (Millipore Sigma) in water in a 200:1 volume ratio.

Strain construction

The complete list of strains, primers and gRNAs used in this study can be found in the supplementary table Data S7. All constructed yeast strains were validated by PCR and confirmed by Sanger sequencing of the locus of interest. Bait strains for the DHFR-PCA experiments were constructed as follows. The genes of interest were fused at their genomic 3' end by replacing the stop codon with a linker (GGGGSGGGGS), the DHFR F[1,2] followed by a smaller linker (GGGGS), a 1X FLAG tag (DYKDDDDK) and the NATMX4 resistance cassette via homologous recombination (all amplified from pAG25-DHFR[1,2]-linker-FLAG). The set of 615 prey strains (DHFR F[3] fusions in the BY4742 background) were retrieved from the Yeast Protein Interactome Collection (Tarassov *et al.*, 2008) with the exception of 73 that were reconstructed (Data S7). Knockout strains were taken from the yeast knockout collection except for *ABP1*, *BEM1*, *BZZ1* and *SLA1* which were reconstructed by replacing the ORFs with the NATMX4 resistance cassette by homologous recombination in the BY4741 background.

Yeast SH3 domain genomic sequences were replaced by CRISPR-Cas9 genome editing. Each SH3 domain (as defined by the SMART database V8.0 except for Ysc84 SH3 (as defined by PROSITE), see Data S1) was specifically targeted by one gRNA as described in (Ryan, Poddar and Cate, 2016). Yeast competent cells were co-transformed with a pCAS plasmid (Addgene plasmid 60847) expressing both the gRNA of interest and *Streptococcus pyogenes* Cas9 and a donor DNA sequence (stuffer) with 60bp homology arms surrounding the SH3 DNA sequence. The Stuffer DNA was either PCR amplified or digested from a plasmid. The stuffer DNA sequence (GGCGGAAGTTCTGGAGGTGGTGGT) codes for a small flexible linker (GGSSGGGG) inspired from protein linkers used in structural biology (Reddy Chichili, Kumar and Sivaraman, 2013; Li *et al.*, 2016). This type of linker is frequently used for fusing proteins without affecting the structure of the linked proteins.

The stuffing of *YSC84* SH3 rendered the strain sterile and was therefore not used for this study as the assays require crosses. For genes containing two SH3s (*BZZ1* and *BEM1*), a combination of this first stuffer with a second one (GGTGGCTCAGGAGGAGGTGGTGGA) coding for a highly similar linker (GGSGGGGG) was used, which allowed targeting each SH3 position independently. The two N-terminal domains of *SLA1*, which are separated by six nucleotides, were stuffed by a single stuffer. Cells containing pCAS were selected via the G418 resistance cassette from the plasmid on YPD plates (YPD + NAT 100 µg/ml (DHFR F[1,2] selection) + G418 200 µg/mL (Bioshop Canada)). Colonies were randomly selected and grown in YPD media without G418 to allow plasmid loss. The correct insertion of the stuffer was validated by PCR and Sanger sequencing. The SH3-swapped, -shuffled and -mutated strains were constructed following a second CRISPR-Cas9 editing step with the same procedures. The stuffer sequences were targeted by specific gRNAs (one gRNA per stuffer) and replaced by PCR amplified specific SH3 (with 40bp homology arms) or libraries of mutated SH3 domains for deep mutational scanning (DMS) with 78bp flanking homology regions. Validation of the SH3 domain insertions was done as we previously described except for DMS libraries (see below).

Cloning

The fragment of plasmid pAG25-DHFR[1,2]-linker-FLAG was synthesized by Synbio Technologies. A GGGGS sequence was added after the DHFR F[1,2] followed by one repeat of the FLAG epitope (DYKDDDDK) in the parental plasmid pAG25-DHFR[1,2] (Tarassov *et al.*, 2008). Cloning of the gRNAs in pCAS was performed as described in (Ryan, Poddar and Cate, 2016). The cloning of the gRNAs sequences was performed via full plasmid PCR amplification with the KAPA (Roche) polymerase using 60-mer oligonucleotides (Eurofins Genomics). The resulting PCR products were digested for 1h at 37°C with DpnI (NEB) and transformed in competent *Escherichia coli* MC1061 cells. Positive colonies were selected on 2YT plates (1% yeast extract, 1.6% tryptone, 0.2% glucose, 0.5% NaCl, and 2% agar) with 100 µg/ml of kanamycin. The SH3 domain DNA sequence of *ABP1* used in the DMS experiments was cloned into pUC19 (NEB) via restriction enzyme cloning with 78bp homology arms surrounding the SH3 DNA (from genomic DNA). The sequences of CTTN, DBNL and HCLS1 SH3 domains (gBlocks, Integrated DNA Technologies) were cloned into pUC19 in the same way. Positive clones were selected on 2YT plates with 100 µg/ml of ampicillin. For NCK2 constructs, the mouse mNCK2 cDNA sequence was cloned via restriction enzymes in frame with a N-terminal 3XFLAG tag in pcDNA3.1(-) (ThermoFisher Scientific) for western blot (WB) experiments and into the pMSCVpuro vector (Clontech) for mass spectrometry (MS) experiments. For bacterial recombinant protein expression and purification, the cDNA was inserted in frame with a 6xHis-GST TEV-cleavable tag into a modified version of pET-30b (Novagen). The NCK2 SH3 shuffled chimeras were constructed using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Inc). The different functional regions of NCK2 are based on NCBI (NP_035009.3) and Uniprot (O55033) definitions and defined as SH3-1 (2-59), linker 1 (60-113), SH3-2 (114-170), linker 2 (171-197), SH3-3 (198-256), Linker 3 (257-282), SH2 (283-380). All cloning was validated by Sanger sequencing.

PCA experiments and analysis

The overall DHFR-PCA pipeline is based on (Tarassov *et al.*, 2008) and the procedures have been previously described (Rochette *et al.*, 2015). The screens were performed with robotically manipulated pin tools (BM5-SC1, S&P Robotics Inc). First, the preys were cherry-picked from 96-well plates to omnitrays in a 384 format on YPD + HPH solid media. Plates were then combined into the 1,536 format on YPD + HPH. The array was constituted of 575 preys previously shown to interact physically with at least one SH3-containing protein as reported in BioGRID V3.4.16 along with 40 negative controls (Stark *et al.*, 2006). The prey array comprised a border of control strains of two rows and two columns corresponding to the PPI between LSM8-DHFR F[1,2] and CDC39-DHFR F[3]. This double border was used to remove border effects. Preys are present in duplicate at a randomized position in the 1,536 array (except for YAL041W which is present four times per array). Baits were crossed with two prey plates, resulting in four independent bait-prey tests for each PPI. Baits were grown in liquid YPD + NAT and then transferred on solid YPD + NAT omnitrays. They were then replicated on YPD + NAT omnitrays in 1,536 array format. Mating of DHFR F[1,2] baits with DHFR F[3] preys was done on YPD followed by incubation at 30°C for 48 hours. Two diploid selection steps were performed by replicating the plates on YPD + NAT + HPH and incubating them for 48 hours at 30°C. Pictures were taken at the end of the second diploid selection for quality control. All images were

acquired with a EOS Rebel T5i camera (Canon). Finally, the diploid cells were replicated on omnitrays containing solid PCA selection media and incubated for 4 days as a first selection step in a spImager custom robotic platform (S&P Robotics Inc). Cells were replicated for a second PCA selection step and incubated in the same manner. The last time point of the second PCA selection step was used for the data analysis as in previous studies using this method.

Colony sizes from pictures were measured with the R package 'Gitter' (Wagih and Parts, 2014). The output was validated by manual inspection. Colonies that did not grow in the diploid selection were removed from the analysis. Diploids that were still present with at least two replicates were kept for downstream analysis. The colony sizes were log2-transformed and each plate was normalized to its background level to allow for comparison among plates. The median log2 normalized colony size was then calculated for each PPI (PCA score). A cut-off was determined based on the PCA score distribution of all PPIs and on the comparison with controls to identify true PPIs (see below). To be called as a true PPI, a pair of bait-prey needed to have the median of its replicates (PCA score) above the threshold (see below). None of the baits used in the DHFR-PCA screens were found to interact with the 40 negative control preys. The first DHFR-PCA screen included WT, SH3-deleted and Abp1 SH3 swapped strains. The plates $Sla1_{SH3-1}$ in Abp1, in which Sla1 SH3-1 was used to replace Abp1 SH3, had no detectable level of background and was assigned the measured background of Abp1 WT plates. The threshold of this screen was determined based on the distribution of the PCA score of every PPI and corresponds to the top 7.3% of all detected colonies in the DHFR-PCA experiment. PPIs above the threshold of the Abp1 WT bait are highly similar (Kendall's r = 0.93) to those of the $Abp1_{SH3}$ in Abp1 control bait, which suggest that the determined threshold allows the detection of true interactions. In addition, we found that the true PPIs of the WT SH3-containing baits are enriched in PPIs that were previously reported (33.2%) when compared to the PPIs that did not pass the threshold (7.7% previously reported PPIs) (Stark *et al.*, 2006). The vast majority of the true PPIs of this screen are from four replicates (93.6%). To determine which PPIs were weaker or stronger following SH3 domain deletions, a ratio was calculated to compare the PCA scores of the PPIs of SH3-deleted baits with the PCA score of the WT baits (PCA score ratio). The same threshold was applied to the PCA score ratio, with the top 7.3% identified as stronger PPIs and the bottom 7.3% as weaker PPIs following SH3 deletions. The same principle was applied to identify PPIs that were stronger or weaker following domain swapping in Abp1. The PCA scores of the PPIs of the Abp1 SH3-swapped baits were compared to the PPIs of the $Abp1_{SH3}$ in Abp1 bait.

Liquid DHFR-PCA low-throughput experiments were performed to validate the PPIs identified as affected by the SH3 domain deletions in the high-throughput solid DHFR-PCA screen (256 PPIs weaker or lost and 116 PPIs stronger or gained following domain deletions). Mating and diploid selection was performed as described for the large scale experiment, with the exception that the plates were in 384 format. Following the second diploid selection, diploid cells were inoculated in a 96 v-shaped well plate using a robotically manipulated 96 pin tool in synthetic complete medium with MSG at pH 6.0 with NAT and HPH. After 48h of growth, optical density (OD) was read using a TECAN Infinite F200 Pro and dilutions to 1 OD in sterile nanopure water were prepared. Cells were then diluted to 0.1 OD by combining 25 µl of cell suspension at OD = 1

and 225 µl of PCA selection media without agar in a 96 wells plate. Plates were incubated at 30°C in a TECAN Infinite M Nano and OD was monitored each 15 mins for 72 hours.

A separate DHFR-PCA screen was performed for the Sla1 SH3 shuffling experiment, with the same analysis as previously described. Three plates had no detectable level of background, which correspond to the second plate of the three Sla1 SH3 WT re-insertion controls. They were attributed the average of the background of the three first plates of the same baits. The threshold was also determined based on the distribution of the PCA scores of all PPIs and corresponds to the top 2.75% of all colonies detected in this experiment. The PPIs above the threshold of the Sla1 WT bait are highly similar (Kendall's r = 0.89, 0.90 and 0.87) to those of the WT SH3 re-insertion control baits, indicating that the threshold allows the detection of true interactions. Almost all true PPIs are from four replicates (96.2%). The Sla1 SH3 shuffled strain SH3-2|SH3-3|SH3-1 DHFR F[1,2] was not successfully constructed and was excluded from the experiment.

Growth assays in stress conditions

All WT, SH3-deleted, KO (except for *CDC25* which is the only essential SH3-containing gene in yeast), *ABP1* SH3-swapped and *SLA1* SH3-shuffled strains were grown on omnitrays in 1,536 format and robotically manipulated with a pin tool (the DHFR F[1,2] bait strains were used except for the KO strains). Each strain was randomly positioned in 6 replicates per plate and two plates per condition were performed, for 12 replicates per strain in total. A border of two rows and two columns of WT strains (BY4741) around the plate was used to remove border effects. Strains were grown on YPD media for two days at 30°C. They were then replicated on 51 different media that are described in the table Data S7 and were incubated at 37°C in a spImager custom platform with images acquired every two hours for four days. The colony sizes from plates pictures were also measured with the R package 'Gitter' and the output was manually inspected. Positions with abnormality in colony circularity were removed. The colony sizes after 74 hours of incubation at 37°C were log2 transformed and used for the analyses. First, the differences between the two plate replicates were corrected to remove plate effects. The median colony size was calculated for each strain per plate. Strains with a difference in their median colony size of more than 2 between the two plate replicates were removed. Finally, the average of the median colony size was calculated for each strain per condition (growth score). For *BEM1* SH3-2 deletion analysis, the strain was more sensitive than the parental strain (BY4741) and the WT strain (WT *BEM1*) in 6% DMSO and was more sensitive compared to the WT *BEM1* strain at 8% DMSO.

For the hygromycin B resistance assays in liquid cultures, the *ABP1* SH3 swapped strains including the *ABP1* SH3-deleted and *ABP1* KO strain were grown overnight at 30°C as pre-cultures in synthetic complete medium with MSG. The next day, OD was measured with a TECAN Infinite F200 Pro plate reader and the cultures were diluted to OD of 1.0 in sterile nanopure water. The different strains were further diluted at 0.1 OD by adding 25 µl of cultures to 225 µl of synthetic complete medium with MSG or with MSG and Hygromycin B 120 µg/mL in 96 wells plate (final concentration of 108 µg/mL). Cells were incubated in a TECAN Infinite M Nano plate reader at 37°C for 24 hours and OD was measured every 15 min.

Deep Mutational Scanning

Single site mutation (SSM) libraries were generated by a PCR-based saturation mutagenesis method (Li, Acevedo-Rocha and Reetz, 2018) as described below. The mutagenesis was carried out on the pUC19 plasmid containing the SH3 domain sequence of *ABP1* flanked by its homology arms subsequently used in the CRISPR mediated genomic re-integration. We used oligos containing degenerate nucleotides (NNN) to carry out the mutagenesis at each codon of the domain sequence. In the first step of the 2-step PCR procedure (short PCR step), an amplicon was generated with an oligo containing the degenerate codon which is positioned within the domain sequence, and another oligo lying outside Abp1 sequence in the plasmid. Short PCR step was carried out with the following settings: 5 min at 95°C, 35 cycles: 20 sec at 98°C, 15 sec at 60°C, 30 sec at 72°C and a final extension of 1 min at 72°C. The amplicon generated in this step was used as a megaprimer in the second step (long PCR step) to amplify the whole plasmid. Long PCR step was carried out with the following settings: 5 min at 95°C, 22 cycles: 20 sec at 98°C, 15 sec at 68°C, 3 min 30 sec at 72°C and a final extension of 5 min at 72°C. The long PCR product was digested for 2 hours at 37°C with DpnI. The digestion product was then transformed into *E.coli* MC1061 competent cells. In order to obtain most of the 64 possible codons, we recovered ~1000 colonies from each plate to retrieve the plasmid single mutant libraries. Saturation mutagenesis was carried out in a codon-position wise manner. Mutants at each position were stored separately. Library quality control was assessed by amplifying the purified plasmid DNA preparations by PCR followed by illumina sequencing as described below. The genomic insertion of the mutant libraries (one library per amino acid position) was performed by targeting the stuffer sequence in the *ABP1* SH3-deleted strain as described in the strain construction section. *ABP1* SH3-deleted yeast competent cells were co-transformed with pCAS containing the stuffer specific gRNA and PCR amplified *ABP1* SH3 single position mutated libraries from the pUC19 preparations. As for the library generation, ~1000 colonies from each plate were retrieved. Glycerol stocks of yeast cells were kept for each position (58 different stocks corresponding to the 58 amino acid positions of Abp1 SH3). The stocks were also validated by Illumina sequencing (see below).

Liquid DHFR-PCA with librairies of mutants

Liquid DHFR-PCA experiments were performed following the generation of the *ABP1* SH3 single position mutated yeast strains and were performed in biological independent duplicates. An overnight culture was started in YPD + NAT by adding the same number of cells from each yeast mutant strains (master pool). Then, the master pool and Hua2-DHFR F[3] or Lsb3-DHFR F[3] preys strains were mixed in YPD (2:1) for mating and incubated for 8h at 30°C. Diploid cells were selected a first time by transferring the mixture in a YPD + NAT + HPH (OD 0.5) for 16h culture at 30°C. The next day, the first diploid selection was transferred in SC complet pH 6.0 + NAT + HPH at 0.5 OD and the culture was grown for 24h at 30°C. The validation of the yeast mutant libraries was done the next day with a fraction of the diploid cells (first DHFR-PCA experiment time point without PCA selection, reference condition S2) by amplifying *ABP1* SH3 genomic sequence followed by illumina sequencing as described in the section below. Following the selection, diploid cells optical density (OD) was read using a TECAN Infinite F200 Pro and dilutions to 0.1 OD in PCA selection media without agar in a 50 mL tube was done. Tubes were incubated at 30°C for 72 hours (second time point of the DHFR-PCA experiment, first PCA

selection). A fraction of the cells (5U OD) were collected for illumina sequencing. The cells from the first PCA step were grown for a second PCA selection step using the same procedures. Finally, the genomic DNA of the cells (5U OD) after two cycles of PCA selection was extracted and the surviving *ABP1* SH3 mutant strains were detected by sequencing as described below.

Single mutant libraries sequencing

For the validation of the plasmid single mutant libraries, the plasmid DNA was extracted from bacteria cells following transformation using a mini-prep plasmid extraction kit (FroggaBio). Yeast genomic DNA extraction of the library of *ABP1* SH3 mutants was performed using a standard phenol/chloroform protocol. Sequencing of yeast genomic DNA was performed at three different time points during the liquid DHFR-PCA experiments. After diploid selection of the bait-prey strains (single mutant library validation, reference condition S2) and following each of the two PCA selection rounds. For the final analysis presented in this article, the first and last time points were used. The libraries for sequencing were prepared by 3 successive rounds of PCR. The first PCR was performed with primers to amplify *ABP1* SH3 from saturation mutagenesis minipreps (4.5 ng of plasmid) or on genomic DNA extracted from yeast (90 ng of genomic DNA) (PCR program : 3 min at 98°C, 20 cycles : 30 sec at 98°C, 15 sec at 60°C and 30 sec at 72°C, final elongation 1 min at 72°C). The second PCR was performed to increase diversity in the libraries by adding Row and Column barcodes (Yachie *et al.*, 2016) for well identification in a 96 well plate (PCR program : 3 min at 98°C, 15 cycles : 30 sec at 98°C, 15 sec at 60°C and 30 sec at 72°C, final elongation 1 min at 72°C). The first PCR served as a template for the second one (2.25 µl of 1/2500 dilution). After the second PCR, 2 µl of the products were run on a 1.5% agarose gel and band intensity was estimated using Image Lab (BioRad Laboratories). The PCR products were mixed based on their intensity on agarose gel to roughly have equal amounts of each in the final library. Mixed PCRs were purified on magnetic beads and quantified using a NanoDrop (ThermoFisher). The third PCR was performed on 0.0045 ng of the purified pool from the second PCR to add a plate barcode and Illumina adapters (PCR program : 3 min at 98°C, 15 cycles : 30 sec at 98°C, 15 sec at 61°C and 35 sec at 72°C, final elongation 1 min 30 sec at 72°C). Each reaction for the third PCR was performed in triplicate, combined and purified on magnetic beads. After purification, libraries were quantified using a NanoDrop (ThermoFisher). Equal amounts of each library were combined and sent to the Genomic Analysis Platform (IBIS, Quebec, Canada) for pair-end 300bp sequencing on a MiSeq (Illumina).

Analysis of DMS data

The raw data generated from deep sequencing (fastq format) was first demultiplexed based on the barcode sequences using custom scripts. Forward and reverse reads corresponding to each sample were merged using PEAR (Zhang *et al.*, 2014). The merged fastq reads were filtered to remove reads with average Phred score of less than 30 and the nucleotides with Phred score of less than 30 using fastp (Chen *et al.*, 2018). The reads were aligned to their corresponding reference sequence using bowtie2 (Langmead and Salzberg, 2012). Global alignment was carried out using the command: “bowtie2 -p 6 --end-to-end --very-sensitive --no-discordant --no-mixed -x $referencep -U $fastqp -S $samp”, where $referencep is path to the bowtie2 built reference, $fastqp is the path to the unaligned fastq file and $samp is the path to the aligned file in sam format. The aligned sam file was used to count the number of mutations using samtools

(Li *et al.*, 2009) and pysam (https://github.com/pysam-developers/pysam). In order to normalise the counts by the depth of sequencing, the counts of the mutations were divided by the depth of sequencing at the position of the mutation. The normalised counts are referred to as frequencies. Next, the log2 ratios (pseudolog with pseudocount of 0.5) of frequencies in the test condition (second PCA selection) were normalized to the reference condition corresponding to the selection of diploids (before the PCA selections, S2) for each PPI. The python based code used for the analysis of the deep mutational scanning data is available in the `align` module of 'rohan' package (Dandage, 2020b). Finally, the ratio of each mutation was scaled by subtracting the median ratio for synonymous codon substitutions (n=120) for each PPI (DMS score). Variants decreasing binding were defined based on the DMS score distribution for synonymous variants at the codon level. Mutations with a DMS score below the 1st percentile of this distribution represent 'deleterious' variants. Likewise, mutations improving PPIs have a DMS score above the 99th percentile of this distribution. For the structural analysis, the relative solvent accessibility (RSA) values were calculated by dividing surface accessibilities by the maximum surface accessibility for the relevant amino acid. Surface accessibilities for PDB: 2RPN were calculated using the DSSP webserver (Kabsch and Sander, 1983), and the 20 maximum surface accessibility values (empirical) taken from (Tien *et al.*, 2013). For the conservation analysis, the 27 yeast SH3s (excluding *SDC25*-YLL017W) were aligned using the MAFFT L-INS-i method (Katoh *et al.*, 2005), and then the per-site conservation calculated with a BLOSUM62 matrix using the *conserv()* function of the R package 'bio3d'. For the comparison of the DMS results with a variant effect predictor, a ΔΔG threshold > 2 was used to define deleterious mutations on the basis of structure, and a SIFT score < 0.001 for variants on the basis of sequence conservation (Wagih *et al.*, 2018).

Cellular lysis and protein immunoprecipitation

Pre-cultures of yeast cells expressing baits fused with the DHFR F[1,2]-Flag were grown overnight at 30°C with agitation in synthetic complete medium. The next day, cells were expanded in synthetic complete medium to reach their exponential phase (OD of around 1.0), pelleted and frozen at -80°C. The equivalent of fifty OD600 of cells were resuspended in ice cold yeast lysis buffer (10 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5 mM EDTA (Millipore Sigma), 1% triton X-100 (Millipore Sigma) and one cOmplete Mini Protease Inhibitor Cocktail (Millipore Sigma) per 10 mL) with 425–600 M glass beads (Millipore Sigma) and disrupted at 4°C by 10 cycles of 2 min vortexing and 2 min cooldown. The equivalent of one hundred OD600 of lysed cells was used for the immunoprecipitation of the bait proteins using FLAG-M2 agarose beads (Millipore Sigma). The beads were incubated with the cell lysates for 120 mins at 4°C with slow rotation, followed by three washes with the yeast lysis buffer.

For the experiments with HEK293T cells (1x10cm confluent dish for WB and 3x15cm for MS), cells were washed in ice-cold PBS and lysed in lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% sodium deoxycholate (Millipore Sigma), 10 mM beta-glycerophosphate (Millipore Sigma), 10 mM sodium pyrophosphate (Millipore Sigma), 50 mM NaF (Millipore Sigma)). Protease inhibitors were added to the lysis buffer (P8340 (Millipore Sigma) for WB and PMSF 1 mM, aprotinin 10 µg/mL, leupeptin 10 µg/mL and pepstatin 10 µg/mL (all Millipore Sigma) for MS). Cells were scraped in the lysis buffer and lysed at 4°C for 20 mins and centrifuged for another 20 mins at 20,000 g at 4°C. Bait proteins from the supernatant were

immuno-precipitated for 90 mins at 4°C with FLAG-M2 agarose beads (Millipore Sigma) and washed three times in the lysis buffer. Beads were then used for WB or MS experiments (see below).

Western blotting and antibodies

Protein extracts and affinity-purified baits from yeast and human lysed cells were migrated on 10% polyacrylamide gels. The separated proteins were then transferred on nitrocellulose membranes. Validation of correct protein loading was done with ponceau staining of the membranes. Before the primary antibody incubations, membranes were blocked in 5% milk. Protein signals were detected on an Amersham Imager 600RGB (GE Healthcare) following the incubation of the membranes with the appropriate HRP-conjugated secondary antibodies and the Clarity Western ECL Substrate (Bio-Rad). The antibodies used in this study are: M2-HRP (A8592, Sigma), PAK (SC-881, Santa Cruz), p130cas/BCAR1 (SC-860, Santa Cruz), actin (Cell Signaling Technology), anti-mouse HRP (Cell Signaling Technology) and anti-rabbit HRP (Cell Signaling Technology). The quantification of the WB signals was performed with the Amersham Imager 600 analysis software (GE Healthcare).

Live imaging of endocytosis

Overnight cultures of Sla1-GFP cells were diluted to an OD600 of 0.15 and grown in synthetic complete medium without tryptophan until they reached an OD600 of ~0.4-0.5 at 30°C. The cells were then seeded on concanavalin A 0.05 mg/mL (Millipore Sigma) coated 8-well glass-bottom chamber slides (Sarstedt). Image acquisition was performed using a Perkin Elmer UltraVIEW confocal spinning disk unit attached to a Nikon Eclipse TE2000-U inverted microscope equipped with a Plan Apochromat DIC H 100×/1.4 oil objective (Nikon), and a Hamamatsu Orca Flash 4.0 LT+ camera. Imaging was done at 25°C in an environmental chamber. The software NIS-Elements (Nikon) was used for image capture. For each field, one brightfield and a series of fluorescence (GFP) images were taken. Cells were excited with a 488 nm laser and emission was filtered with a 530/630 nm filter. GFP time laps were acquired continuously at a rate of 1 frame/sec for 3 minutes.

Analysis of microscopy images

Bright field images were used to segment cells using YeastSpotter (Lu et al. 2019). Cells were filtered based on circularity, solidity and the normalized difference between minor and major axis lengths to remove poorly detected cells. Out of focus cells were also filtered out based on size and brightness. The centroids of the segmented cells were used to identify the locations of each cell in the image using Scikit-Image (van der Walt et al. 2014). The location of each cell was used to isolate individual cell time-frames of fluorescence (GFP) images. Each cell time frame of the fluorescence images was processed through a python based single particle tracking tool - Trackpy (Allan et al. 2019). Trackpy detected the locations of Sla1-GFP positive foci in each frame and linked them together to provide particle-wise trajectories. Trajectories detected in less than 10 frames were considered as spurious and were filtered out. Segments of the trajectories spuriously indicating movement of the particles back to the cell membrane were trimmed. The preprocessed trajectories of the particles were then used for the calculation of distances using the `spatial distance` module of 'SciPy' (Jones et al. 2016). The python

based code used for the analysis of microscopy images are included in the `endocytosis` module of 'htsimaging' package (Dandage 2020a).

Protein purifications

E. coli BL21 (DE3) cells were transformed with mNCK2 constructs in the pET-30b vector and grown overnight at 37°C in LB with 100 µg/ml kanamycin. The next day, cultures were expanded in LB with kanamycin until they reached their exponential growth phase (OD of around 1.0). Protein production was then induced by the addition of 1 mM IPTG (Bio Basic) and incubation at 16°C overnight with agitation. Cultures were pelleted with a centrifugation step (3,200 g at 4°C, 30 mins), washed with ice-cold PBS and pellets were kept at -80°C. Bacteria were thawed in GST buffer (20 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1 mM EDTA and 1 mM DTT (Millipore Sigma)) with cOmplete protease inhibitors (Millipore Sigma) and disrupted with sonication. Lysates were then centrifuged at 20,000 g for 20 min at 4°C and the supernatant loaded to a GSTrap FF 1 mL column (GE Healthcare). Following a washing step with 10 mL GST-Buffer, the GST-Tagged proteins were eluted with 15 mM GSH (Bio-basic) and cleaved in solution using purified recombinant His-Tagged TEV protease (Thermo Fisher Scientific). Tag-free NCK2 proteins were further purified using a 1 ml HisTrap FF column (GE Healthcare) to remove free His-tagged GST and TEV as well as remaining impurities. The purified proteins were concentrated with Amicon 10K filters (Millipore Sigma) and their concentration assessed using Coomassie and BCA assays (Thermo Fisher Scientific).

Fluorescence polarization

Increasing amounts of NCK2 full-length recombinant proteins in GST buffer (final concentrations of 0-150 µM for CD3ε and 0-175 µM for PAK1) were used for binding assays with a constant amount of fluorescein isothiocyanate (FITC)-conjugated peptides (resuspended in 17 mM Tris-HCl pH 8, 100 mM NaCl and 0.5% Brij L23 (Millipore Sigma), peptides final concentration of 10 nM). Cd3ε (RGQNKERPPPVPNPDY) and PAK1 (DIQDKPPAPPMRNTST) peptides were used as NCK2 SH3-1 and NCK2 SH3-2 ligands, respectively. Binding assays were performed in the FP buffer (17 mM Tris-HCl pH 8 and 100 mM NaCl) and fluorescence polarization was measured on a Synergy H1 multi-mode plate reader (Bio Tek) at 535 nm with 485 nm excitation. Calculation of the dissociation constants was performed with a one-site total binding model and nonlinear regression in GraphPad Prism version 7.

Phase separation

Purified NCK2 proteins were diluted in GST buffer to identical concentrations. At 500 mM NaCl present in this buffer, all NCK2 constructs were completely soluble. Phase transition was initiated by dilution in imidazole buffer (10 mM imidazole (Bio Basic) pH 7.0, 1 mM DTT, 1mM EGTA (Millipore Sigma), 1 mM MgCl2, 2.5% Glycerol (Bioshop Canada), 3% Dextran (Millipore Sigma)) containing various amount of NaCl so that the final concentration of NaCl would correspond to those in Figure S7A. For all the other experiments, phase transition of NCK2 proteins was assessed at 60 mM NaCl. NCK2 at 8 µg/µL (40 µg in 5uL) was added to 35 uL imidazole buffer in a plastic 96 well plate that had first been pre-treated with a 0.2% BSA solution to reduce nonspecific binding. The plate assay was then incubated overnight (24 hours) at 4°C before imaging. Phase contrast images of each well were acquired using a Zeiss Axio Vert.A1 inverted microscope with a LD A-Plan 40x/0.55 Ph1 objective (Zeiss) in Zen Blue

software (version 2.3.69.1000). After imaging, the soluble protein fraction (supernatant) was recovered for SDS-PAGE analysis. Protein droplets attached at the bottom of the well (pellet) were washed once with imidazole buffer with 60 mM NaCl and then solubilized in Laemmli buffer. The pellets to supernatant fractions were used to quantify phase transition at 60 mM NaCl (7 replicates) via Coomassie protein quantification. The involvement of weak hydrophobic interactions in NCK2 driven phase transition was assessed using 1,6-Hexanediol (Millipore Sigma) at 60 mM NaCl. Imidazole buffer containing 0, 1%, 5% and 10% 1,6-Hexanediol was first prepared and phase transition was initiated and analyzed as described above. Finally, competition of SH3-2 dependent homotypic NCK2 interaction with a SH3-2 binding peptide was assessed in the presence of 5 µM, 15 µM, 45 µM, 90 µM, 150 µM, or 300 µM PAK1 peptide in imidazole buffer with 60 mM NaCl. The impact of PAK1 peptide addition was observed via phase contrast microscopy as described above.

Experimental design for mass spectrometry experiments

NCK2 baits purified on beads from HEK293T cells (as previously described) were washed two additional times in 20 mM Tris-HCl pH 7.4 and eluted in 50 mM phosphoric acid. Proteins were then digested with trypsin (Promega) on-beads as previously detailed (Beigbeder *et al.*, 2016). Briefly, eluted proteins were concentrated on ZipTip®-SCX (Millipore Sigma), reduced with TCEP 100 mM (Millipore Sigma) and alkylated during the tryptic digestion with iodoacetamide 10 mM (Millipore Sigma). Peptides were then eluted in ammonium bicarbonate 200 mM (Millipore Sigma) and desalted on Stage Tips C18 columns (Rappsilber, Mann and Ishihama, 2007). The final eluted peptides were dried using an Eppendorf Vacufuge™ Concentrator. For each bait, two biological replicates were processed independently. These were analyzed alongside four negative controls corresponding to transiently expressed 3×FLAG-GFP purified from HEK293T cells (as described for NCK2 constructs). These control cell lines were grown in parallel to those expressing baits and treated in the same manner. The equivalent of one third of each sample was used for the SWATH-libraries generation and another third for SWATH quantification. Data dependent acquisition (DDA) and data independent acquisition (DIA) analysis of each sample was performed back-to-back without washes to diminish the instrument time required to complete the analysis. To minimize carry-over issues during liquid chromatography, extensive washes were performed between each sample; and the order of sample acquisition on the mass spectrometer was also reversed for the second biological replicates to avoid systematic bias. We validated that the same quantity of baits was used for each NCK2 chimeras by quantifying NCK2 peptides in each sample (Data S6).

Proteins identification by mass spectrometry

The analyses were performed at the proteomic platform of the Quebec Genomics Center. Peptide samples were separated by online reversed-phase (RP) nanoscale capillary liquid chromatography (nanoLC) and analyzed by electrospray mass spectrometry (ESI MS/MS). The experiments were performed with a Dionex UltiMate 3000 nanoRSLC chromatography system (Thermo Fisher Scientific) connected to a TripleTOF 5600+ mass spectrometer (Sciex) equipped with a nanoelectrospray ion source. Peptides were trapped at 20 µl/min in loading solvent (2% acetonitrile, 0.05% TFA) on a 5 mm x 300 µm C_{18} pepmap cartridge pre-column

(Thermo Fisher Scientific) during 5 minutes. Then, the pre-column was switch online with a self-made 50 cm x 75 um internal diameter separation column packed with ReproSil-Pur C_{18}-AQ 3-µm resin (Dr. Maisch HPLC) and the peptides were eluted with a linear gradient from 5-40% solvent B (A: 0,1% formic acid, B: 80% acetonitrile, 0.1% formic acid) in 120 minutes, at 300 nL/min. In DDA, the instrument method consisted of one 250 milliseconds (ms) MS1 TOF survey scan from 400–1300 Da followed by 20 100 ms MS2 candidate ion scans from 100–2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 5+ that exceeded a threshold of 150 cps were selected for MS2, and former precursors were excluded for 20 s after one occurrence. In DIA, acquisition consisted of one 50 ms MS1 scan followed by 32 × 25 a.m.u. isolation windows covering the mass range of 400–1250 a.m.u. (cycle time of 3.25 s); an overlap of 1 Da between SWATH was preselected. The collision energy for each window was set independently as defined by CE = 0.06 × m/z + 4, where m/z is the centre of each window, with a spread of 15 eV performed linearly across the accumulation time.

Data Dependent Acquisition (DDA) MS analysis

Mass spectrometry data was stored, searched and analyzed using the ProHits laboratory information management system (LIMS) platform (Liu *et al.*, 2016). Sciex .wiff mass spectrometry files were converted to mzML and mzXML using ProteoWizard (3.0.4468;(Kessner *et al.*, 2008)). The mzML and mzXML files were then searched using Mascot (v2.3.02) and Comet (v2012.02 rev.0) (Eng, Jahan and Hoopmann, 2013). The spectra were searched with the RefSeq database (version 57, January 30th, 2013) acquired from NCBI against a total of 72,482 human and adenovirus sequences supplemented with “common contaminants” from the Max Planck Institute (http://141.61.102.106:8080/share.cgi?ssid=0f2gfuB) and the Global Proteome Machine (GPM; http://www.thegpm.org/crap/index.html) (Beavis, 2006). Charges +2, +3 and +4 were allowed and the parent mass tolerance was set at 12 ppm while the fragment bin tolerance was set at 0.6 amu. Carbamidomethylation of cysteine was set as a fixed modification. Deamidated asparagine and glutamine and oxidized methionine were allowed as variable modifications. The results from each search engine were analyzed through TPP (the Trans-Proteomic Pipeline (v4.6 OCCUPY rev 3), (Deutsch *et al.*, 2015) via the iProphet pipeline (Shteynberg *et al.*, 2011)).

Data independent acquisition (DIA) MS analysis with MSPLIT-DIA

DIA MS data was analysed using MSPLIT-DIA (version 1.0; (Wang *et al.*, 2015)) implemented in ProHits 4.0 (Liu *et al.*, 2016). To generate a sample-specific spectral library for the FLAG AP-MS dataset, peptide-spectrum matches (PSMs) from matched DDA runs (18 runs) were pooled by retaining only the spectrum with the lowest MS-GFDB (Beta version 1.0072 (6/30/2014) (Kim *et al.*, 2010)) probability for each unique peptide sequence and precursor charge state, and a peptide-level false discovery rate (FDR) of 1% was enforced using TDA (Elias and Gygi, 2007). The MS-GFDB parameters were set to search for tryptic cleavages, allowing no missed cleavage sites, 1 C_{13} atom per peptide with a mass tolerance of 50 ppm for precursors with charges of 2+ – 4+ and a tolerance of ±50 ppm for fragment ions. Peptide length was limited to 8–30 amino acids. Variable modifications were deamidated asparagine and

glutamine and oxidized methionine. The spectra were searched with the NCBI RefSeq database (version 57, January 30th, 2013) against a total of 36,241 human and adenovirus sequences supplemented with "common contaminants" from the Max Planck Institute (http://141.61.102.106:8080/share.cgi?ssid=0f2gfuB) and the Global Proteome Machine (GPM; http://www.thegpm.org/crap/index.html). Decoys were appended using the decoy library command built in to MSPLIT-DIA, with a fragment mass tolerance of ±0.05 Da. The spectral library was then used for protein identification by MSPLIT as previously described (Wang *et al.*, 2015) with peptides identified by MSPLIT-DIA passing a 1% FDR subsequently matched to genes using ProHits 4.0 (Liu *et al.*, 2016). The MSPLIT search parameters were as follows: parent mass tolerance of ±25 Da and fragment mass tolerance of ±50 ppm. When retention time was available within the spectral library, a cut-off of ±5 min was applied to spectral matching as previously described (Wang *et al.*, 2015).

MS data analysis with SAINTexpress

SAINTexpress version 3.6.1 (Teo *et al.*, 2014) was used as a statistical tool to calculate the probability value of each potential protein-protein interaction compared to background contaminants using default parameters. The four control samples were used in uncompressed mode. Two unique peptide ions and a minimum iProphet probability of 0.95 were required for protein identification prior to SAINTexpress in DDA mode while only 2 unique peptides were required in DIA mode.

Position and sequence conservation of orthologs

The UniProt database was used to retrieve full-length sequences of all *S. cerevisiae* SH3 proteins except for the Sdc25 (YLL017W) pseudogene (UniProt Consortium, 2019). One-to-one orthologs for each protein were then retrieved from fungal species using Ensembl Compara (Herrero *et al.*, 2016). This process was automated for each SH3 protein using the Ensembl REST API (Yates *et al.*, 2015). The *hmmscan* function of HMMER v3.3 was then used for the domain annotation of all orthologs using default significance thresholds of 0.01 (E = 0.01 and domE = 0.01) and the Pfam-A HMM library (El-Gebali *et al.*, 2019). Any domain belonging to the Pfam SH3 clan (CL0010) was assigned as an SH3 domain. SH3 domain positions for each orthologous sequence were taken by dividing the domain start point ('env_from') by the full sequence length. This calculation was repeated for each *S. cerevisiae* homolog, and orthologous SH3 domains were considered positionally conserved if their domain positions were within a 10% sequence length window of the corresponding *S. cerevisiae* domain.

The sequence conservation of each SH3 domain was calculated using the same set of orthologous sequences as described directly above. Domain sequences within an orthologous group were aligned together using the MAFFT L-INS-i method (Katoh *et al.*, 2005). The pairwise similarity between domain sequences was determined from the *seqidentity()* function in the R package 'bio3d' (Grant *et al.*, 2006), and we took the mean pairwise sequence similarity between the *S. cerevisiae* sequence and all its orthologs as a measure of sequence conservation for each domain.

Tree comparison analysis

In Figure 3B, the relationship between SH3 domain sequences and their PPI profiles in Abp1 was explored by constructing dendrograms for both features and then quantifying the similarity between dendrograms. For the interaction similarities (Figure 3B, left), a Euclidean distance matrix of the PCA scores was first constructed using the *dist()* function in R, and then the distance matrix used to generate a dendrogram from the *hclust()* function in R with default parameters. For the sequence similarities (Figure 3B, right), the SH3 domain sequences were first aligned using the MAFFT L-INS-i method (Katoh *et al.*, 2005). A distance matrix was then constructed from the *seqidentity()* function in the R package 'bio3d' (Grant *et al.*, 2006) — by taking *distance* = 1−*similarity* — and then hierarchical clustering performed in the same manner as for the DHFR-PCA PPIs. Side-by-side visualisations of the dendrograms were generated using the R package 'dendextend' (Galili, 2015). The overall similarity between dendrograms was quantified using the cophenetic correlation, which correlates the tree-wise distances for both dendrograms across all possible SH3 pairs. This was calculated using the *cor.dendlist()* function in the 'dendextend' package. To test for statistical significance, we permuted the data by randomising the assignment of SH3 domains to different DHFR-PCA PPIs profiles and then recalculated the cophenetic correlation using all steps described directly above. This was repeated 10,000 times to generate a random distribution of cophenetic correlation scores.

This tree comparison procedure was repeated when comparing SH3 domain PPI profiles with the SH3 position weight matrices (PWMs) as described in the section above. To quantify similarity between PWMs, we calculated PWM-target similarity scores (using matrix similarity score (MSS)) for each PWM across ~2,672 putative SH3 ligands predicted for *S. cerevisiae* from a previous study (Tonikian *et al.*, 2009). Pearson's correlations were then generated for each PWM-PWM pair and used to calculate a Euclidean distance matrix. Hierarchical clustering was performed using the *hclust()* function in R with default parameters to generate the dendrogram of PWM similarity. Finally, a similar tree-based approach was used to compare PPI profiles with growth profiles for Sla1 constructs generated from the domain shuffling experiments (Figure S5B). Both the DHFR-PCA interaction data and the Sla1 growth data was clustered as described above, by generating a Euclidean distance matrix and then performing hierarchical clustering.

PWMs enrichment analysis

All yeast PWMs were taken from a 2009 study of SH3 domain specificity (Tonikian *et al.*, 2009). The similarity between SH3 PWMs and known interactors was assessed using a matrix similarity score (MSS) derived from the MATCH algorithm (Kel *et al.*, 2003). This scoring method assigns a score of 1 to perfect sequence matches to the PWM and vice versa. For each interactor assigned to an SH3 domain, MSS scores for the corresponding PWM were calculated for all possible sequence *k*-mers (*k* = number of PWM columns) and then the maximum taken (Max. MSS). This procedure was repeated for all SH3 domains with an assigned PWM and then the results pooled to generate Figures 2B, 2C, and Figure S3D. Some SH3 domains were represented by more than one PWM from (Tonikian *et al.*, 2009), reflecting multiple specificities (Gfeller *et al.*, 2011). In these cases, MSSs were calculated for both PWMs and then the overall maximum score taken forward for further analysis.

In Figure 2, maximum MSS scores were calculated for sequences belonging to the 'Random', 'SH3-independent PPI', 'SH3-inhibited PPI', and 'SH3-dependent PPI' category of interactor sequences. For the 'Random' category, random peptides were generated by sampling amino acids according to their background frequency in the *S. cerevisiae* proteome. For SH3-independent PPI, we included only those interactors that were found to be unaffected by *any* SH3 domain deletions given that *in vitro*-derived SH3 PWMs overlap strongly, which could lead to spurious MSS enrichment for SH3-independent PPIs. In Figure 2C, maximum MSSs were also re-calculated for the 'SH3-dependent PPI' category of interactor sequences after randomly assigning PWMs to each SH3 domain; this procedure was repeated until 10,000 maximum MSS scores were sampled. For Figure S3D, the 'PPI gained by Abp1' correspond to PPIs gained by Abp1 after domain swapping, and the 'PPI unaffected' correspond to PPIs of Abp1 that do not change after domain swapping. The 'Random' category was generated using the same approach described directly above for Figure 2.

Comparison of interactome with literature

To find previously reported PPIs, we parsed a recent release (v 3.5.16) of the BioGRID and searched for reports of physical interactions between baits and preys (Stark *et al.*, 2006). PPIs were searched in the two directions possible (A-B and B-A) in the database.

GO enrichment

To determine the biological processes enriched in our DHFR-PCA data, our list of positive interactions partners were analyzed against our list of unique preys using the Gene Ontology Term Finder of the SGD database, only considering enrichment with p-value <= 0.01 (Boyle *et al.*, 2004).

QUANTIFICATION AND STATISTICAL ANALYSIS

All the statistical details of the different experiments can be found in the figure legends, figures and in the results sections.

References:

Aebersold, R. *et al.* (2018) 'How many human proteoforms are there?', *Nature chemical biology*, 14(3), pp. 206–214.

Allan, D. *et al.* (2019) 'soft-matter/trackpy: Trackpy v0.4.2'. doi: 10.5281/zenodo.3492186.

Arbesú, M. and Pons, M. (2019) 'Integrating disorder in globular multidomain proteins: Fuzzy sensors and the role of SH3 domains', *Archives of biochemistry and biophysics*, 677, p. 108161.

Banjade, S. *et al.* (2015) 'Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck', *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), pp. E6426–35.

Banjade, S. and Rosen, M. K. (2014) 'Phase transitions of multivalent proteins can promote clustering of membrane receptors', *eLife*, 3. doi: 10.7554/eLife.04123.

Barreto, L. *et al.* (2011) 'A genomewide screen for tolerance to cationic drugs reveals genes important for potassium homeostasis in Saccharomyces cerevisiae', *Eukaryotic cell*, 10(9), pp. 1241–1250.

Beavis, R. C. (2006) 'Using the global proteome machine for protein identification', *Methods in molecular biology* , 328, pp. 217–228.

Beigbeder, A. *et al.* (2016) 'Sample Preparation for Mass Spectrometry Analysis of Protein-Protein Interactions in Cancer Cell Lines and Tissues', *Methods in molecular biology* , 1458, pp. 339–347.

Bergendahl, L. T. *et al.* (2019) 'The role of protein complexes in human genetic disease', *Protein science: a publication of the Protein Society*, 28(8), pp. 1400–1411.

Biot-Pelletier, D. and Martin, V. J. J. (2016) 'Seamless site-directed mutagenesis of the Saccharomyces cerevisiae genome using CRISPR-Cas9', *Journal of biological engineering*, 10, p. 6.

Bisson, N. *et al.* (2007) 'EphA4 signaling regulates blastomere adhesion in the Xenopus embryo by recruiting Pak1 to suppress Cdc42 function', *Molecular biology of the cell*, 18(3), pp. 1030–1043.

Bisson, N. *et al.* (2011) 'Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor', *Nature biotechnology*, 29(7), pp. 653–658.

Bokoch, G. M. *et al.* (1996) 'Interaction of the Nck adapter protein with p21-activated kinase (PAK1)', *The Journal of biological chemistry*, 271(42), pp. 25746–25749.

Boyle, E. I. *et al.* (2004) 'GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes', *Bioinformatics* , 20(18), pp. 3710–3715.

Broome, M. A. and Hunter, T. (1997) 'The PDGF receptor phosphorylates Tyr 138 in the c-Src

SH3 domain in vivo reducing peptide ligand binding', *Oncogene*, 14(1), pp. 17–34.

Buday, L., Wunderlich, L. and Tamás, P. (2002) 'The Nck family of adapter proteins: regulators of actin cytoskeleton', *Cellular signalling*, 14(9), pp. 723–731.

Carducci, M. *et al.* (2012) 'The protein interaction network mediated by human SH3 domains', *Biotechnology advances*, 30(1), pp. 4–15.

Chaki, S. P. and Rivera, G. M. (2013) 'Integration of signaling and cytoskeletal remodeling by Nck in directional cell migration', *Bioarchitecture*, 3(3), pp. 57–63.

Chen, S. *et al.* (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics* , 34(17), pp. i884–i890.

Collins, B. C. *et al.* (2013) 'Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system', *Nature methods*, 10(12), pp. 1246–1253.

Cunningham, J. M. *et al.* (2020) 'Biophysical prediction of protein-peptide interactions and signaling networks using machine learning', *Nature methods*, 17(2), pp. 175–183.

Dandage, R. (2020a) 'rraadd88/htsimaging: Single particle tracking to monitor endocytosis in yeast'. doi: 10.5281/zenodo.3697135.

Dandage, R. (2020b) 'rraadd88/rohan v0.2.0'. doi: 10.5281/zenodo.3697136.

Deutsch, E. W. *et al.* (2015) 'Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics', *Proteomics. Clinical applications*, 9(7-8), pp. 745–754.

Dionne, U. *et al.* (2018) 'Direct Phosphorylation of SRC Homology 3 Domains by Tyrosine Kinase Receptors Disassembles Ligand-Induced Signaling Networks', *Molecular cell*, 70(6), pp. 995–1007.e11.

Diss, G. and Lehner, B. (2018) 'The genetic landscape of a physical interaction', *eLife*, 7. doi: 10.7554/eLife.32472.

Drees, B. L. *et al.* (2001) 'A protein interaction map for cell polarity development', *The Journal of cell biology*, 154(3), pp. 549–571.

El-Gebali, S. *et al.* (2019) 'The Pfam protein families database in 2019', *Nucleic acids research*, 47(D1), pp. D427–D432.

Elias, J. E. and Gygi, S. P. (2007) 'Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry', *Nature methods*, 4(3), pp. 207–214.

Engen, J. R. *et al.* (2008) 'Structure and dynamic regulation of Src-family kinases', *Cellular and molecular life sciences: CMLS*, 65(19), pp. 3058–3073.

Eng, J. K., Jahan, T. A. and Hoopmann, M. R. (2013) 'Comet: an open-source MS/MS sequence database search tool', *Proteomics*, 13(1), pp. 22–24.

Errington, W. J., Bruncsics, B. and Sarkar, C. A. (2019) ‘Mechanisms of noncanonical binding dynamics in multivalent protein-protein interactions’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(51), pp. 25659–25667.

Fazi, B. *et al.* (2002) ‘Unusual binding properties of the SH3 domain of the yeast actin-binding protein Abp1: structural and functional analysis’, *The Journal of biological chemistry*, 277(7), pp. 5290–5298.

Feliciano, D. and Di Pietro, S. M. (2012) ‘SLAC, a complex between Sla1 and Las17, regulates actin polymerization during clathrin-mediated endocytosis’, *Molecular biology of the cell*, 23(21), pp. 4256–4272.

Freschi, L. *et al.* (2013) ‘qPCA: a scalable assay to measure the perturbation of protein--protein interactions in living cells’, *Molecular bioSystems*. Royal Society of Chemistry, 9(1), pp. 36–43.

Frese, S. *et al.* (2006) ‘The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains’, *The Journal of biological chemistry*, 281(26), pp. 18236–18245.

Galili, T. (2015) ‘dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering’, *Bioinformatics* , 31(22), pp. 3718–3720.

Gao, A. *et al.* (2018) ‘Evolution of weak cooperative interactions for biological specificity’, *Proceedings of the National Academy of Sciences of the United States of America*, 115(47), pp. E11053–E11060.

Garcia, B., Stollar, E. J. and Davidson, A. R. (2012) ‘The importance of conserved features of yeast actin-binding protein 1 (Abp1p): the conditional nature of essentiality’, *Genetics*, 191(4), pp. 1199–1211.

Gfeller, D. *et al.* (2011) ‘The multiple-specificity landscape of modular peptide recognition domains’, *Molecular systems biology*, 7, p. 484.

Ghaemmaghami, S. *et al.* (2003) ‘Global analysis of protein expression in yeast’, *Nature*, 425(6959), pp. 737–741.

Giaever, G. *et al.* (2002) ‘Functional profiling of the Saccharomyces cerevisiae genome’, *Nature*, 418(6896), pp. 387–391.

Giaever, G. and Nislow, C. (2014) ‘The yeast deletion collection: a decade of functional genomics’, *Genetics*, 197(2), pp. 451–465.

Gil, D. *et al.* (2002) ‘Recruitment of Nck by CD3 epsilon reveals a ligand-induced conformational change essential for T cell receptor signaling and synapse formation’, *Cell*, 109(7), pp. 901–912.

Gillet, L. C. *et al.* (2012) ‘Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis’, *Molecular & cellular proteomics: MCP*, 11(6), p. O111.016717.

Grant, B. J. *et al.* (2006) ‘Bio3d: an R package for the comparative analysis of protein structures’, *Bioinformatics* , 22(21), pp. 2695–2696.

Herrero, J. *et al.* (2016) 'Ensembl comparative genomics resources', *Database: the journal of biological databases and curation*, 2016. doi: 10.1093/database/baw053.

Holtzman, D. A., Yang, S. and Drubin, D. G. (1993) 'Synthetic-lethal interactions identify two novel genes, SLA1 and SLA2, that control membrane cytoskeleton assembly in Saccharomyces cerevisiae', *The Journal of cell biology*, 122(3), pp. 635–644.

Ho, Y. *et al.* (2002) 'Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry', *Nature*, 415(6868), pp. 180–183.

Huh, W.-K. *et al.* (2003) 'Global analysis of protein localization in budding yeast', *Nature*, 425(6959), pp. 686–691.

Ideker, T. and Krogan, N. J. (2012) 'Differential network biology', *Molecular systems biology*, 8, p. 565.

Ivarsson, Y. and Jemth, P. (2019) 'Affinity and specificity of motif-based protein-protein interactions', *Current opinion in structural biology*, 54, pp. 26–33.

Jones, E. *et al.* (2016) 'SciPy: Open source scientific tools for Python, 2001'.

Kabsch, W. and Sander, C. (1983) 'Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features', *Biopolymers*, 22(12), pp. 2577–2637.

Kaksonen, M., Toret, C. P. and Drubin, D. G. (2005) 'A modular design for the clathrin- and actin-mediated endocytosis machinery', *Cell*, 123(2), pp. 305–320.

Katoh, K. *et al.* (2005) 'MAFFT version 5: improvement in accuracy of multiple sequence alignment', *Nucleic acids research*, 33(2), pp. 511–518.

Kay, B. K. (2012) 'SH3 domains come of age', *FEBS letters*, 586(17), pp. 2606–2608.

Kel, A. E. *et al.* (2003) 'MATCH: A tool for searching transcription factor binding sites in DNA sequences', *Nucleic acids research*, 31(13), pp. 3576–3579.

Kelil, A. *et al.* (2017) 'Exhaustive search of linear information encoding protein-peptide recognition', *PLoS computational biology*, 13(4), p. e1005499.

Kelil, A., Levy, E. D. and Michnick, S. W. (2016) 'Evolution of domain-peptide interactions to coadapt specificity and affinity to functional diversity', *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), pp. E3862–71.

Kessner, D. *et al.* (2008) 'ProteoWizard: open source software for rapid proteomics tools development', *Bioinformatics* , 24(21), pp. 2534–2536.

Kim, S. *et al.* (2010) 'The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search', *Molecular & cellular proteomics: MCP*, 9(12), pp. 2840–2852.

Kroschwald, S. *et al.* (2015) 'Promiscuous interactions and protein disaggregases determine the

material state of stress-inducible RNP granules', *eLife*, 4, p. e06807.

Ladbury, J. E. and Arold, S. T. (2012) 'Noise in cellular signaling pathways: causes and effects', *Trends in biochemical sciences*, 37(5), pp. 173–178.

Lambert, J.-P. *et al.* (2013) 'Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition', *Nature methods*, 10(12), pp. 1239–1245.

Landgraf, C. *et al.* (2004) 'Protein interaction networks by proteome peptide scanning', *PLoS biology*, 2(1), p. E14.

Landry, C. R. *et al.* (2013) 'Extracting insight from noisy cellular networks', *Cell*, 155(5), pp. 983–989.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, 9(4), pp. 357–359.

Larson, S. M. and Davidson, A. R. (2000) 'The identification of conserved interactions within the SH3 domain by alignment of sequences and structures', *Protein science: a publication of the Protein Society*, 9(11), pp. 2170–2180.

Laursen, L. *et al.* (2020) 'Functional interplay between protein domains in a supramodular structure involving the postsynaptic density protein PSD-95', *The Journal of biological chemistry*, 295(7), pp. 1992–2000.

Letunic, I. and Bork, P. (2018) '20 years of the SMART protein domain annotation resource', *Nucleic acids research*, 46(D1), pp. D493–D496.

Letunic, I., Doerks, T. and Bork, P. (2015) 'SMART: recent updates, new developments and status in 2015', *Nucleic acids research*, 43(Database issue), pp. D257–60.

Li, A., Acevedo-Rocha, C. G. and Reetz, M. T. (2018) 'Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy', *Applied microbiology and biotechnology*. Springer, 102(14), pp. 6095–6103.

Li, G. *et al.* (2016) 'Construction of a linker library with widely controllable flexibility for fusion protein design', *Applied microbiology and biotechnology*, 100(1), pp. 215–225.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* , 25(16), pp. 2078–2079.

Li, P. *et al.* (2012) 'Phase transitions in the assembly of multivalent signalling proteins', *Nature*, 483(7389), pp. 336–340.

Li, S. *et al.* (2001) 'Tyrosine phosphorylation of Grb2 by Bcr/Abl and epidermal growth factor receptor: a novel regulatory mechanism for tyrosine kinase signaling', *The EMBO journal*, 20(23), pp. 6793–6804.

Liu, G. *et al.* (2016) 'Data Independent Acquisition analysis in ProHits 4.0', *Journal of proteomics*, 149, pp. 64–68.

Liu, J. *et al.* (2006) 'Structural insight into the binding diversity between the human Nck2 SH3 domains and proline-rich proteins', *Biochemistry*, 45(23), pp. 7171–7184.

Lu, A. X. *et al.* (2019) 'YeastSpotter: accurate and parameter-free web segmentation for microscopy images of yeast cells', *Bioinformatics* , 35(21), pp. 4525–4527.

Manenschijn, H. E. *et al.* (2019) 'Type-I myosins promote actin polymerization to drive membrane bending in endocytosis', *eLife*, 8. doi: 10.7554/eLife.44215.

Mapes, J. and Ota, I. M. (2004) 'Nbp2 targets the Ptc1-type 2C Ser/Thr phosphatase to the HOG MAPK pathway', *The EMBO journal*, 23(2), pp. 302–311.

Mayer, B. J. (2015) 'The discovery of modular binding domains: building blocks of cell signalling', *Nature reviews. Molecular cell biology*, 16(11), pp. 691–698.

Mayer, B. J., Blinov, M. L. and Loew, L. M. (2009) 'Molecular machines or pleiomorphic ensembles: signaling complexes revisited', *Journal of biology*, 8(9), p. 81.

Mayer, B. J., Hamaguchi, M. and Hanafusa, H. (1988) 'A novel viral oncogene with structural similarity to phospholipase C', *Nature*, 332(6161), pp. 272–275.

McCann, J. J. *et al.* (2012) 'Supertertiary structure of the synaptic MAGuK scaffold proteins is conserved', *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp. 15775–15780.

Meng, Y., Pond, M. P. and Roux, B. (2017) 'Tyrosine Kinase Activation and Conformational Flexibility: Lessons from Src-Family Tyrosine Kinases', *Accounts of chemical research*, 50(5), pp. 1193–1201.

Merő, B. *et al.* (2019) 'Structural insights into the tyrosine phosphorylation-mediated inhibition of SH3 domain-ligand interactions', *The Journal of biological chemistry*, 294(12), pp. 4608–4620.

Mirey, G. *et al.* (2005) 'SH3 domain-containing proteins and the actin cytoskeleton in yeast', *Biochemical Society transactions*, 33(Pt 6), pp. 1247–1249.

Mittal, A. *et al.* (2018) 'Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins', *Journal of molecular biology*, 430(16), pp. 2403–2421.

Mund, M. *et al.* (2018) 'Systematic Nanoscale Analysis of Endocytosis Links Efficient Vesicle Formation to Patterned Actin Nucleation', *Cell*, 174(4), pp. 884–896.e17.

Murata, Y. *et al.* (2003) 'Dimethyl sulfoxide exposure facilitates phospholipid biosynthesis and cellular membrane proliferation in yeast cells', *The Journal of biological chemistry*, 278(35), pp. 33185–33193.

Nars, M. and Vihinen, M. (2001) 'Coevolution of the domains of cytoplasmic tyrosine kinases', *Molecular biology and evolution*, 18(3), pp. 312–321.

Rappsilber, J., Mann, M. and Ishihama, Y. (2007) 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nature protocols*,

2(8), pp. 1896–1906.

Ravi Chandra, B. *et al.* (2004) 'Distribution of proline-rich (PxxP) motifs in distinct proteomes: functional and therapeutic implications for malaria and tuberculosis', *Protein engineering, design & selection: PEDS*, 17(2), pp. 175–182.

Reddy Chichili, V. P., Kumar, V. and Sivaraman, J. (2013) 'Linkers in the structural biology of protein-protein interactions', *Protein science: a publication of the Protein Society*, 22(2), pp. 153–167.

Rivero-Lezcano, O. M. *et al.* (1995) 'Wiskott-Aldrich syndrome protein physically associates with Nck through Src homology 3 domains', *Molecular and cellular biology*, 15(10), pp. 5725–5731.

Rochette, S. *et al.* (2015) 'Genome-wide protein-protein interaction screening by protein-fragment complementation assay (PCA) in living cells', *Journal of visualized experiments: JoVE*, (97). doi: 10.3791/52255.

Rodal, A. A. *et al.* (2003) 'Negative regulation of yeast WASp by two SH3 domain-containing proteins', *Current biology: CB*, 13(12), pp. 1000–1008.

Rohatgi, R. *et al.* (2001) 'Nck and phosphatidylinositol 4,5-bisphosphate synergistically activate actin polymerization through the N-WASP-Arp2/3 pathway', *The Journal of biological chemistry*, 276(28), pp. 26448–26452.

Ryan, O. W. *et al.* (2014) 'Selection of chromosomal DNA libraries using a multiplex CRISPR system', *eLife*, 3. doi: 10.7554/eLife.03703.

Ryan, O. W., Poddar, S. and Cate, J. H. D. (2016) 'CRISPR-Cas9 Genome Engineering in Saccharomyces cerevisiae Cells', *Cold Spring Harbor protocols*, 2016(6). doi: 10.1101/pdb.prot086827.

Schlaepfer, D. D., Broome, M. A. and Hunter, T. (1997) 'Fibronectin-stimulated signaling from a focal adhesion kinase-c-Src complex: involvement of the Grb2, p130cas, and Nck adaptor proteins', *Molecular and cellular biology*, 17(3), pp. 1702–1713.

Scott, J. D. and Pawson, T. (2009) 'Cell signaling in space and time: where proteins come together and when they're apart', *Science*, 326(5957), pp. 1220–1224.

Shteynberg, D. *et al.* (2011) 'iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates', *Molecular & cellular proteomics: MCP*, 10(12), p. M111.007690.

Stahl, M. L. *et al.* (1988) 'Sequence similarity of phospholipase C with the non-catalytic region of src', *Nature*, 332(6161), pp. 269–272.

Stark, C. *et al.* (2006) 'BioGRID: a general repository for interaction datasets', *Nucleic acids research*, 34(Database issue), pp. D535–9.

Stollar, E. J. *et al.* (2009) 'Structural, functional, and bioinformatic studies demonstrate the crucial role of an extended peptide binding site for the SH3 domain of yeast Abp1p', *The Journal*

of biological chemistry, 284(39), pp. 26918–26927.

Stumpf, M. P. H. *et al.* (2008) 'Estimating the size of the human interactome', *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), pp. 6959–6964.

Sun, T. *et al.* (2019) 'Prediction of liquid-liquid phase separation proteins using machine learning', *bioRxiv*. doi: 10.1101/842336.

Tarassov, K. *et al.* (2008) 'An in vivo map of the yeast protein interactome', *Science*, 320(5882), pp. 1465–1470.

Teo, G. *et al.* (2014) 'SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software', *Journal of proteomics*, 100, pp. 37–43.

Teyra, J. *et al.* (2017) 'Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities', *Structure* , 25(10), pp. 1598–1610.e3.

Tien, M. Z. *et al.* (2013) 'Maximum allowed solvent accessibilites of residues in proteins', *PloS one*, 8(11), p. e80635.

Tolsma, T. O., Cuevas, L. M. and Di Pietro, S. M. (2018) 'The Sla1 adaptor-clathrin interaction regulates coat formation and progression of endocytosis', *Traffic* , 19(6), pp. 446–462.

Tong, A. H. Y. *et al.* (2002) 'A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules', *Science*, 295(5553), pp. 321–324.

Tonikian, R. *et al.* (2009) 'Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins', *PLoS biology*, 7(10), p. e1000218.

UniProt Consortium (2019) 'UniProt: a worldwide hub of protein knowledge', *Nucleic acids research*, 47(D1), pp. D506–D515.

Vogel, C. and Chothia, C. (2006) 'Protein family expansions and biological complexity', *PLoS computational biology*, 2(5), p. e48.

Vogel, C., Teichmann, S. A. and Pereira-Leal, J. (2005) 'The relationship between domain duplication and recombination', *Journal of molecular biology*, 346(1), pp. 355–365.

Wagih, O. *et al.* (2018) 'A resource of variant effect predictions of single nucleotide variants in model organisms', *Molecular systems biology*, 14(12), p. e8430.

Wagih, O. and Parts, L. (2014) 'gitter: a robust and accurate method for quantification of colony sizes from plate images', *G3* , 4(3), pp. 547–552.

van der Walt, S. *et al.* (2014) 'scikit-image: image processing in Python', *PeerJ*, 2, p. e453.

Wang, J. *et al.* (2015) 'MSPLIT-DIA: sensitive peptide identification for data-independent acquisition', *Nature methods*, 12(12), pp. 1106–1108.

Winzeler, E. A. *et al.* (1999) 'Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis', *Science*, 285(5429), pp. 901–906.

Xin, X. *et al.* (2013) ‘SH3 interactome conserves general function over specific form’, *Molecular systems biology*, 9, p. 652.

Yachie, N. *et al.* (2016) ‘Pooled-matrix protein interaction screens using Barcode Fusion Genetics’, *Molecular systems biology*, 12(4), p. 863.

Yadav, S. S. and Miller, W. T. (2008) ‘The evolutionarily conserved arrangement of domains in SRC family kinases is important for substrate recognition’, *Biochemistry*, 47(41), pp. 10871–10880.

Yates, A. *et al.* (2015) ‘The Ensembl REST API: Ensembl Data for Any Language’, *Bioinformatics* , 31(1), pp. 143–145.

Yu, H. *et al.* (2008) ‘High-quality binary protein interaction map of the yeast interactome network’, *Science*, 322(5898), pp. 104–110.

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

Zhang, J. *et al.* (2014) ‘PEAR: a fast and accurate Illumina Paired-End reAd mergeR’, *Bioinformatics* , 30(5), pp. 614–620.

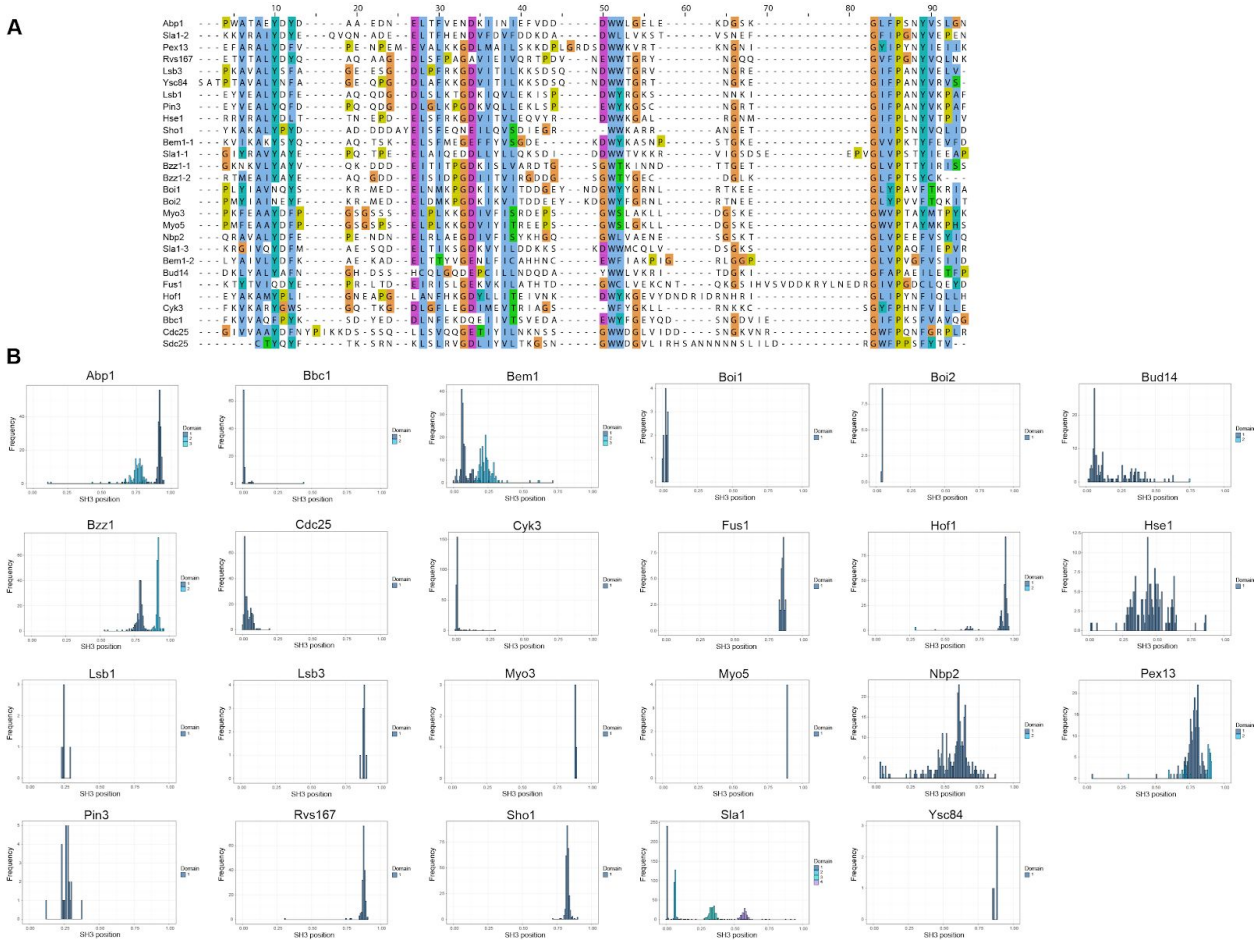

Figure S1: Yeast SH3 domain sequence and position in their host protein: A) Multiple sequence alignment of yeast SH3 domains. B) Each panel represents the distribution of SH3 domain positions for the one-to-one orthologs of a given *S. cerevisiae* SH3. For each ortholog sequence, SH3 domain start positions were identified from SH3 hidden Markov models (HMMs), and then the start positions compiled across all orthologs to derive a frequency distribution. The SH3 domains are numbered in cases where more than one SH3 was identified across orthologs, where '1' represents the domain that is also found in the *S. cerevisiae* copy. For multi-SH3 proteins (Bem1, Bzz1, and Sla1) the numbers correspond to the domain numbering given in Figure 1B.

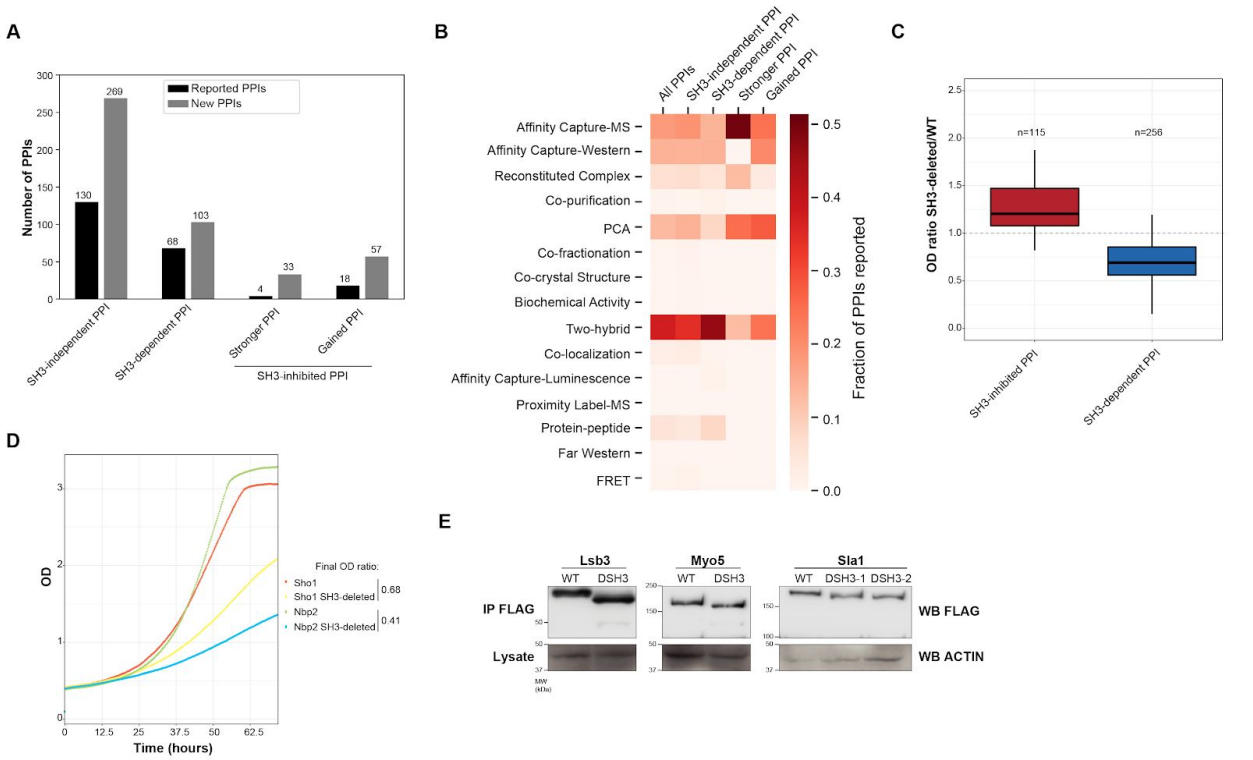

Figure S2: Number and types of PPI changes in response to SH3 deletion and validation. A) The proportion of PPIs previously reported for the different types of PPIs affected or not by SH3 deletion (Stark *et al.*, 2006). B) The fraction of known PPIs per method of detection is shown for the same categories of interactions as in A) (Stark *et al.*, 2006). C) Validation of the PPIs altered by the deletion of yeast SH3s using low-throughput liquid DHFR-PCA. The ratios (SH3-deleted/WT) of the optical density from the last time point of the experiment for the growth curves of each PPI are shown. A ratio higher than 1 represents an interaction that is stronger upon SH3 deletion. D) Examples of two SH3-dependent PPIs DHFR-PCA liquid assay. Growth curves are shown for the WT and SH3-deleted baits for Sho1-Pbs2 and Nbp2-Pbs2 PPIs. The ratios of the final optical density (SH3-deleted/WT) as represented in panel C) are also indicated. E) Western blot analysis of the expression level of SH3-deleted proteins that lost many PPIs upon their SH3-deletion. All baits have a C-terminal 1xFLAG tag allowing their immuno-detection. See also Data S2.

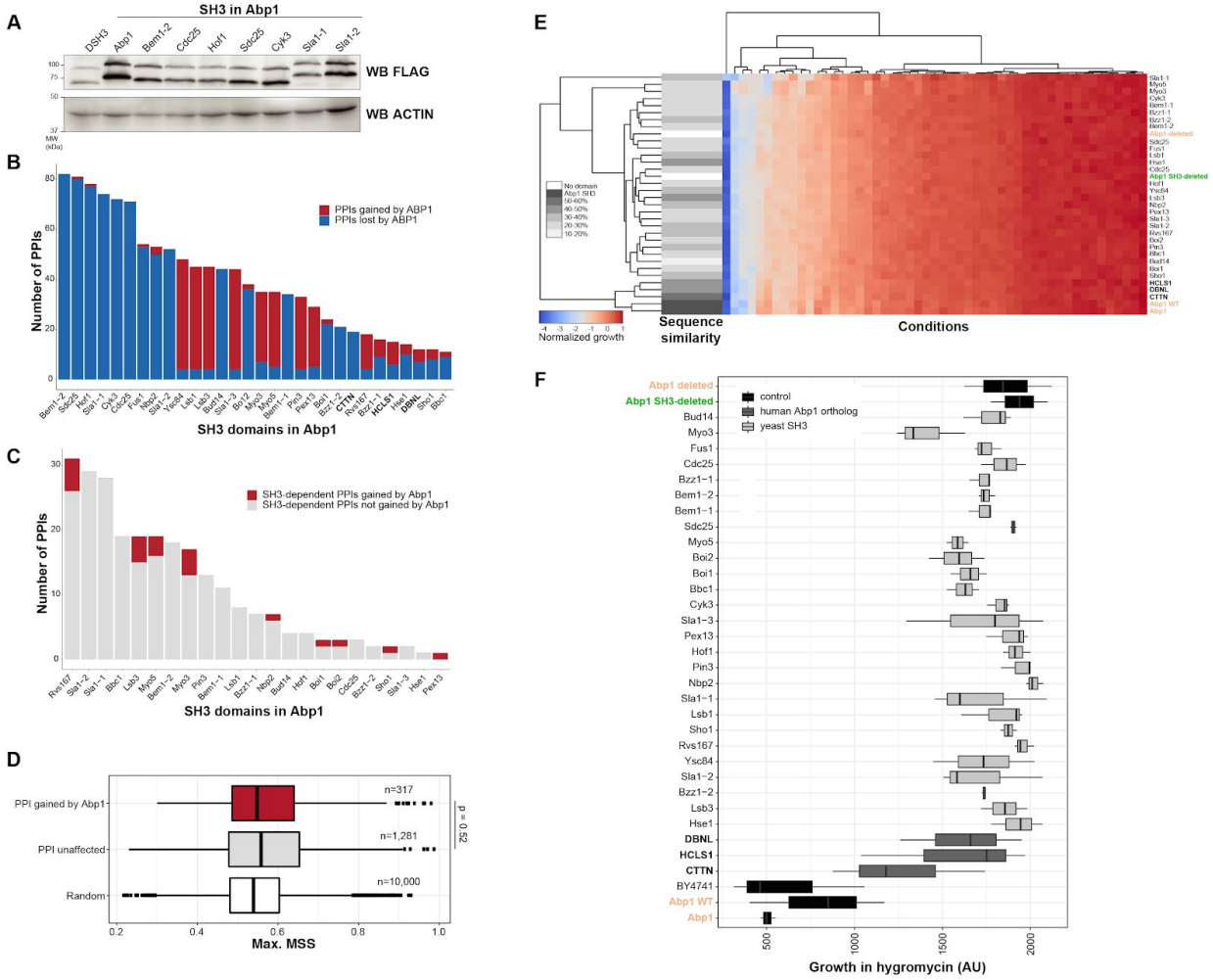

Figure S3: Characterization of Abp1 SH3 domain swapping. A) Protein expression levels of the Abp1 SH3 swapped proteins for which the smallest number of DHFR-PCA PPIs were measured as compared to the $Abp1_{SH3}$ in Abp1 and SH3-deleted strains. Two bands are detected for Abp1, as previously observed (Garcia, Stollar and Davidson, 2012). B) Number of PPIs that were affected by Abp1 SH3 domain swapping. C) Number of SH3-dependent PPIs that were gained by each SH3 swapped to Abp1 (as determined in Figure 2A) in comparison to the ones not gained. D) PWMs analysis of PPIs gained by Abp1 using the matrix similarity scores (MSS) between the PWM and sequence. This analysis shows no enrichment of the predicted SH3 motifs in the gained PPIs relative to the unaffected Abp1 PPIs (p = 0.52, Mann-Whitney test, one-sided). E) Growth in stress conditions for the Abp1 SH3 swapped strains. The growth values were scaled per strain (row). Blue to red represents the normalized growth per strain after 74 hours (log2 colony size). The sequence similarity of each SH3 to Abp1 SH3 is represented in grey scale. Each strain was grown in twelve replicates in each condition. F) Growth of Abp1 SH3 swapped stains in liquid medium with hygromycin is represented (area under the curve (AU)). The control strains, shown in black, highlight that the resistance to the drug is dependent on Abp1 SH3. Domains are sorted by increasing sequence similarity with Abp1 from the top. Growth rates were measured in triplicates. For E and F) the orange SH3s are controls and the SH3-deleted protein is in green. In all panels, human SH3s are shown in bold. See also Data S3.

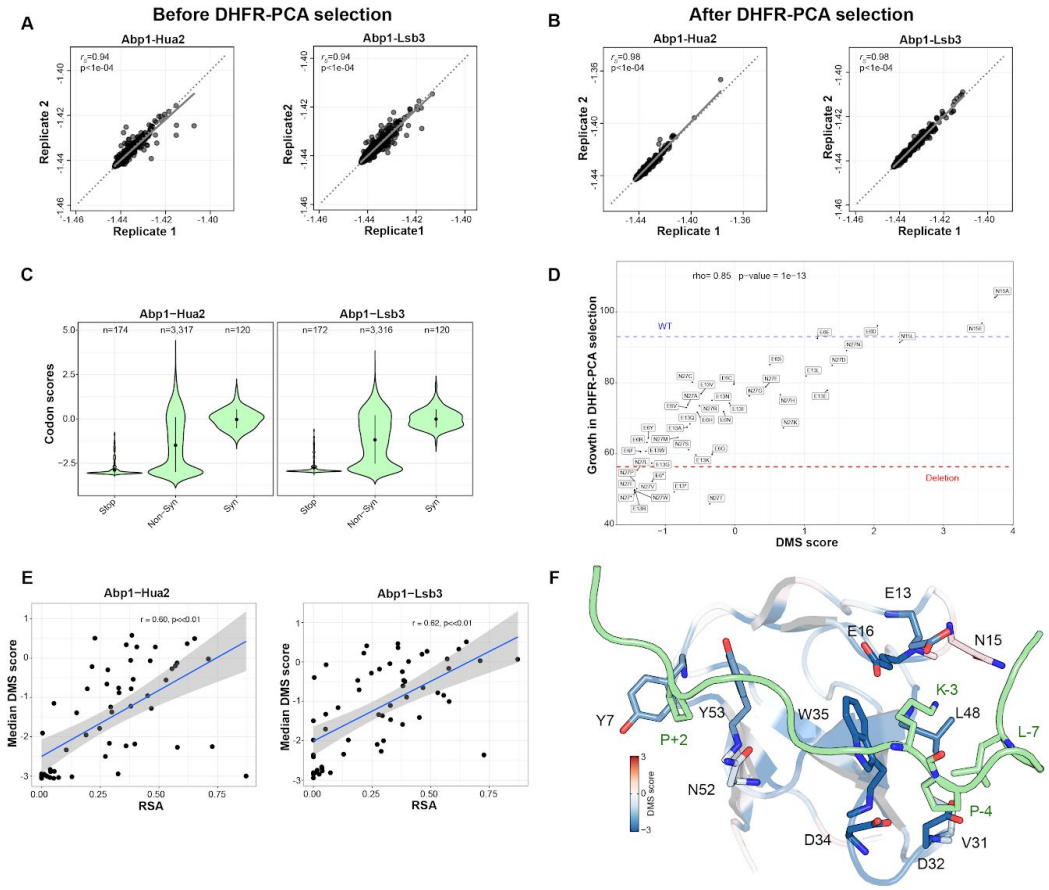

Figure S4: Characterization and validation of DMS mutants of Abp1 SH3 domain. A-B) Reproducibility between biological replicates of DMS DHFR-PCA experiments. The correlations for the two interactors are shown (A is for the reference condition and B for the DHFR-PCA condition). The frequencies of mutants (shown on log2 scale) were compared in terms of Spearman's rank correlation coefficient (r_s). Associated p-value is shown on each plot. C) Distribution of the codon log2 average sequence counts for each type of mutation of Abp1 SH3 for the two PPIs. Syn is for synonymous. D) Selected non-conserved positions (E6, E13, N15 and N27) sensitivities to mutations for the Abp1-Hua2 PPI were tested in a low-throughput liquid DHFR-PCA assay. The growth of the mutants in the low-throughput experiment (average of four replicates of the total OD for each growth curve, y axis) is compared to the mutants DMS score (x axis). The growth of WT Abp1 and Abp1 SH3-deleted are represented by horizontal dash lines. E) The median DMS score of the 58 different Abp1 SH3 positions for both PPIs in relation to their relative solvent accessibility (RSA) is shown. F) Abp1 SH3 residues in close proximity (within 4 Å) to the most important binding residues of the Ark1 peptide (P+2, K-3, P-4, and L-7) are mapped on the structure of Abp1 SH3 in complex with the peptide (PDB: 2RPN, (Stollar *et al.*, 2009)). Ark1 peptide is shown in green, oxygen atoms are in red and nitrogen atoms are in blue. The colors of Abp1 SH3 residues that are mapped on the structure represent their average DMS score for Abp1-Hua2 PPI. See also Data S4.

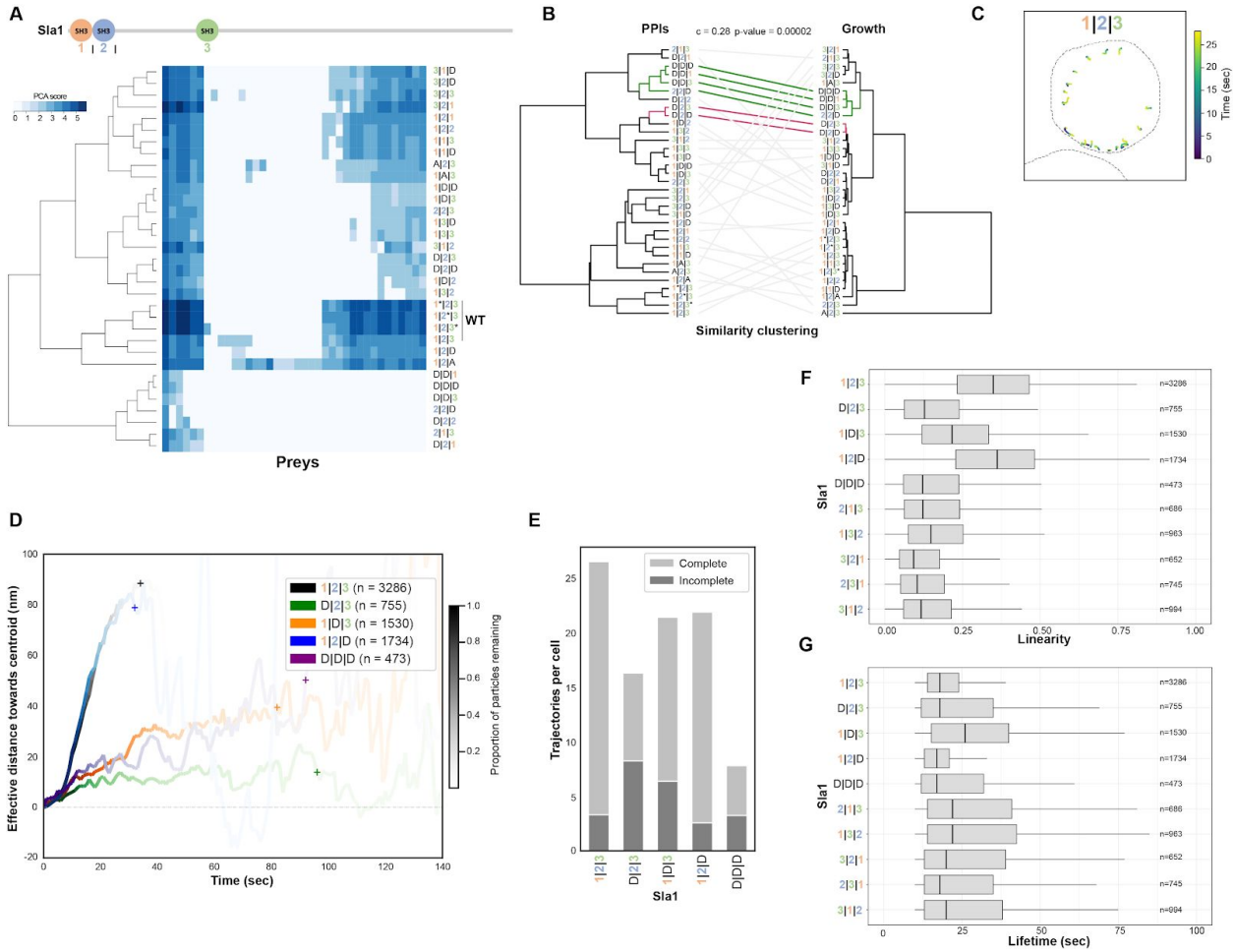

Figure S5: SH3 domains shuffling impacts Sla1 PPIs, cells growth in different stress conditions and clathrin-mediated endocytosis. A) Sla1 SH3-deleted or -shuffled PPIs as detected by DHFR-PCA. The color code represents PPI strength as detected by DHFR-PCA (PCA score). All PPIs were measured in quadruplicate. A is for Abp1 SH3 and * is for the reinsertion of the WT domain as control strains. A scaled cartoon of Sla1 is also illustrated. B) Cophenetic correlation for the similarity of Sla1 SH3-shuffled PPI clusters with the growth phenotype clusters. Empirical p-value obtained from data permutation is p = 0.00002. C) Representative fluorescence microscopy timeframe analysis of a yeast cell expressing WT Sla1-GFP. The calculated trajectories for every Sla1-GFP foci detected are shown. Cell membrane is delimited with a dashed line and each Sla1-GFP particle colors represent its position in time. D) Sla1-GFP particles average effective distance travelled towards the cell center through time. Sla1 SH3-deleted strains are shown. The proportion of events that are not completed yet is represented by the color transparency of the curves (+ represent the time point when 95% of foci have disassembled). E) Complete or incomplete Sla1-GFP endocytosis events are shown (average per cell) for Sla1 SH3-deleted strains. F) Representation of the linearity of Sla1-GFP particle trajectories for each SH3 deletion or shuffling. G) Sla1-GFP particle lifetime (in seconds) are shown for the same strains as in F). See also Data S5.

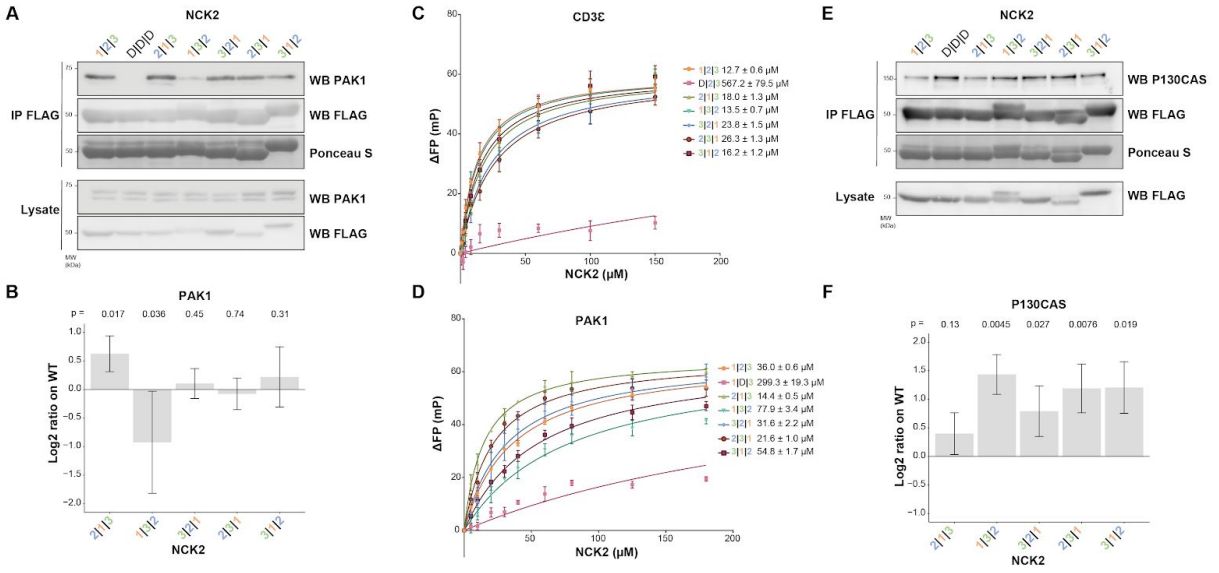

Figure S6: NCK2 SH3 shuffling affects its interactions with PAK1 and P130CAS/BCAR1 in cells but only slightly alters the availability of the SH3 binding pockets *in vitro*. A) Western blot analysis of NCK2 SH3-shuffled proteins interaction with PAK1. Each bait is N-terminally fused to a 3XFLAG tag. Loading corresponds to Ponceau staining of the nitrocellulose membrane. Experiment was performed in four replicates in HEK293T cells. B) Western blot quantification of PAK1 co-immunoprecipitation with NCK2. Average log2 ratio relative to the WT protein is shown (NCK2mut/NCK2WT, WT NCK2 log2 ratio = 0). Ratios from replicates were compared to WT NCK2 via pairwise ANOVA statistical test (p-values: 2|1|3 = 0.017, 1|3|2 = 0.036, 3|2|1 = 0.45, 2|3|1 = 0.74 and 3|1|2 = 0.31). Error bars indicate the SD. C-D) Fluorescence polarization *in vitro* curves for NCK2 binding to CD3Ɛ or PAK1. Each binding assay was executed in triplicate. Error bars indicate the SD of the average of the ∆FP for each point. Dissociation constants error values represent the SE of the value derived from the binding curve. E) Western blot analysis of NCK2 SH3-shuffled proteins association with SH2 target P130CAS/BCAR1. The experiment was performed in four replicates in HEK293T cells stimulated with the tyrosine phosphatase inhibitor pervanadate. F) Quantification of the Western blot signals of P130CAS co-immunoprecipitation with NCK2 as in B).The four ratios of the replicates were compared to WT NCK2 via pairwise ANOVA statistical test (p-values: 2|1|3 = 0.13, 1|3|2 = 0.0045, 3|2|1 = 0.027, 2|3|1 = 0.0076 and 3|1|2 = 0.019). Error bars indicate the SD. See also Data S6.

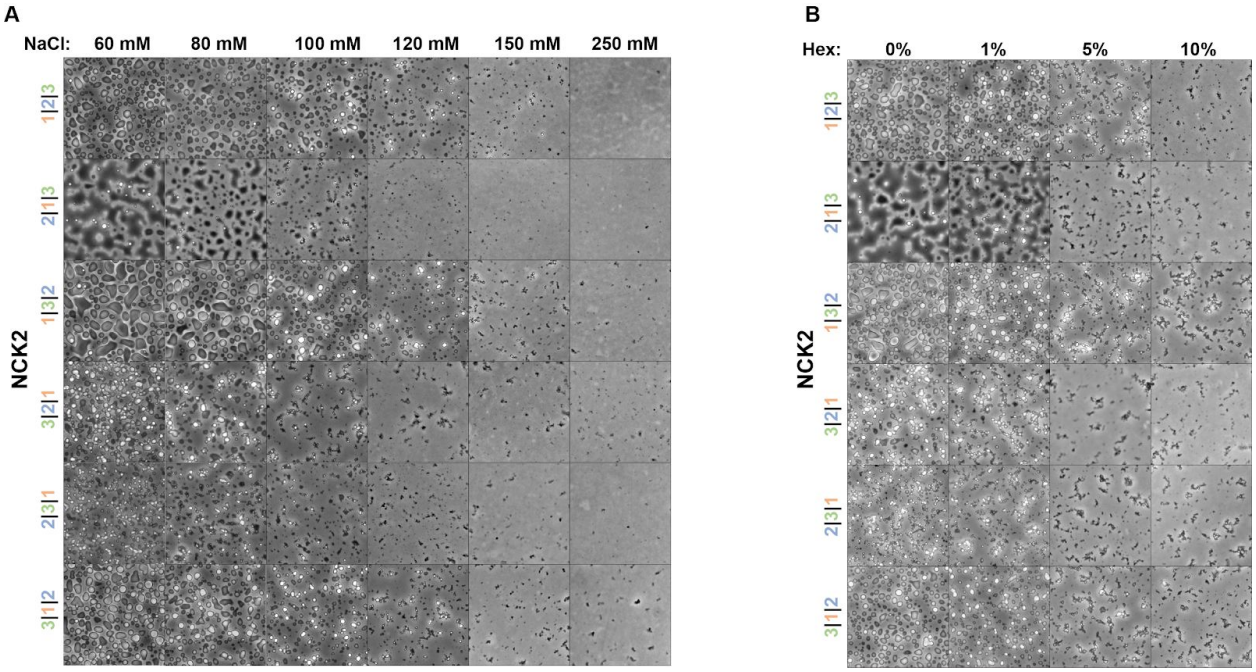

Figure S7: NCK2 SH3-shuffled proteins ability to promote phase transition is perturbed by NaCl and 1,6-hexanediol. A-B) An increasing amount of salt (A, NaCl) or 1,6-hexanediol (B, Hex) was added to the different NCK2 SH3-shuffled protein samples and incubated for 24 hours before observing phase transition using phase contrast microscopy. The experiment was performed in triplicates.