**UNIVERSITÀ DEGLI STUDI DI ROMA
"TOR VERGATA"**

FACOLTA' DI SCIENZE MATEMATICHE FISICHE E NATURALI

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE E MOLECOLARE

XXII Ciclo

The human interaction network mediated by Src Homology-3 domains

Enriching the viral-host interactomes with interactions mediated by Src Homology-3 domains

Martina Carducci

Docente Guida/Tutor: Prof. G. Cesareni

Coordinatore: Prof. G. Cesareni

Abstract

Protein-protein interactions play an essential role in the regulation of most cellular processes. The ability of proteins to form functional complexes is in part supported by families of conserved protein domains that are specialized in mediating interactions with relatively short linear peptide motifs. An important subclass of these motifs, those that are characterized by the high proline content, play a pivotal role in biological processes requiring the coordinated assembly of multi-protein complexes. This is achieved via interaction with proteins containing modules such as Src Homology-3 (SH3) that are specialized in recognizing specific proline rich patterns. To characterize the interactome defined by poly-proline binding domains we have used a variant of the WISE (Whole Interactome Scanning Experiment) approach where pre-synthesized, naturally occurring, peptides are arrayed at high density on a glass surface.

By this method we have tested a collection of 90 SH3 domains for their binding to a set of 9600 poly-proline containing peptides immobilized on a glass chip.

To evaluate the quality of the obtained dataset, we performed a series of pull down experiments. The results validated more than 90% of the peptide-domain interactions.

The protein-peptide interactions were also assayed with a library of M13 filamentous phages in which the gene for the VIII coat protein is fused to random 9-mer peptides. In order to evaluate the false positive rate of our experiments we carried out SPOT synthesis assays in which the peptides interacting with some SH3 domains were re-synthesized and probed with the protein fusions.

Peptide chips, pull down assays, SPOT synthesis and phage display experiments allowed us to determine the specificity and promiscuity of proline-rich binding domains and to map their interaction network. All the predicted interactions were collected and stored in the PepspotDB (http://mint.bio.uniroma2.it/PepspotDB/), a bioinformatic resource developed in our lab for this purpose.

Many viral pathogenic strategies involve targeting and perturbing host protein interactions. The characterization of the host protein sub-networks disturbed by invading viruses is a major goal of viral research and may contribute to reveal fundamental biological mechanisms and to identify new therapeutic strategies. To assist in this approach we have developed a database, VirusMINT, that stores in a structured format most of the published interactions between the viral and the host proteome. Although

SH3 are the most ubiquitous and abundant class of protein binding modules, VirusMINT contains only a few interactions mediated by this domain class. To overcome this limitation we have applied the WISE (Whole Interactome Scanning Experiment) approach to identify interactions between 15 human SH3 domains and viral proline-rich peptides of two oncogenic viruses, HPV (Human papillomavirus type 16) and Ad12 (Human adenovirus A type 12). This approach identified 114 new potential interactions between the human SH3 domains and proline rich regions of the two viral proteomes.

Introduction

In the post-genomic era, one of the main goals is the functional characterization of gene products and their correct positioning in the interaction web underlying cell physiology.

In fact, despite the wealth of gene and protein sequence information, we are still far from a system level understanding of how genes/proteins mediate cellular functions.

If we disregard complications, such as differential splicing and posttranslational modifications, mammalian genomes encode approximately 25000 proteins, with a variety of catalytic or structural functions. The regulation of their synthesis and activity is brought about by a complex interaction web that is still far from being fully described. In addition, the stable or transient formation of complexes, via protein interaction, mediates, in a way that is not fully understood, a non uniform distribution of specific activities in diverse organelles or in less spatially defined cellular niches.

Although a satisfactory understanding of cellular functions would require a dynamic representation of the cell interactome, we are still far from a complete, let alone quantitative description, of a static interactome, where all the possible "biochemically naïve" interactions are annotated with their dissociation constants. Protein's complexes formation is largely supported by families of conserved protein domains.

Proline rich binding domains and proline rich sequences.

Modular interaction domains are protein's conserved regions specialized in mediating the interaction with relatively short linear peptides or with other biomolecules such as lipids and nucleic acids.

Intracellular protein domains recognizing proline-rich sequences (PRS) play a pivotal role in biological processes that require the coordinated assembly of multi-protein complexes. The super-family of proline-rich sequence recognition domains consists of the SH3 (Src homology-3) (Feng et al., 1994; Yu et al., 1994), the WW (characterized by two conserved Trp residues) (Macias et al., 1996), the EVH1 (Ena/VASP Homology domain 1) (Reinhard et al., 1995), the GYF (Glycine-tYrosine-PHenylalanine) (Nishizawa et al., 1998), the UEV (Ubiquitin E2 Variant) (Pornillos et al., 2002) and profilin (Schutt et al., 1993) domains (Figure 1).

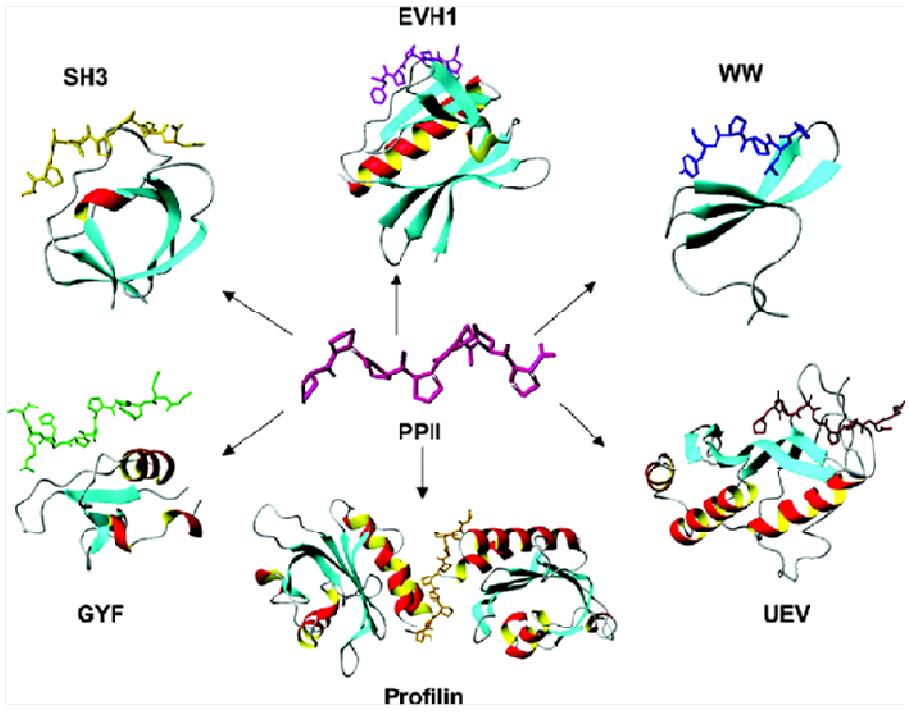


Figure 1. Structures of six proline rich binding domains in complex with their cognate ligands (Li, 2005).

The proline-rich target peptides recognized by these protein-interaction modules generally contains several consecutive proline residues, some of which are required for binding. Among the 20 naturally occurring amino acids, proline is the only one in which the side chain atoms form a pyrrolidine ring with the backbone atoms (Figure 2a). This cyclic structure confers to proline its specific properties: it induces conformational constraints among the atoms in the pyrrolidine ring, it is responsible for the slow isomerization between cis/trans conformations (Eckert et al., 2005) as well as for the secondary structure preferences of proline-rich sequences.

These peculiar properties of the proline residue cause that proline-rich sequences form either of two different secondary structures: Poli-Proline type I (PPI) helices (in which all prolines are cis isomers) and Poli-Proline type II (PPII) helices (where all prolines are trans isomers) (Figure 2b). The PPII helix is a left-handed helix with three residues per turn (Figure. 2b and

c) in which the conformational restriction of the torsion angles is given from a value of Φ =-78° and of Ψ=+146°. It has a three-fold symmetry when viewed along the helical axis and every forth residue is in the same position (at a distance of 9.3 A from each other). Along the same axis, the PPII helix also has a two-fold rotational pseudo symmetry (Zarrinpar et al., 2003). The side chains and backbone carbonyl groups are located in similar positions in both orientations along the backbone axis. This observation explains why SH3 domains may bind their PRS ligands in either of the two orientations (Feng et al., 1994). In contrast to the aqueous environment, where either PPI or PPII conformation seems possible, all the proline-rich sequence peptides characterized so far adopt a pure PPII helix upon binding to the recognition domains (Mayer et al., 1988; Nishizawa et al., 1998; Stahl et al., 1988).

PPII helices are exposed and mobile structure often found on the surfaces of proteins (Adzhubei and Sternberg, 1993; Stapley and Creamer, 1999) or at the ends of proteins (Williamson, 1994). Proline is the preferred amino acid in PPII helices but must not be necessarily present in all of them (Adzhubei and Sternberg, 1993).

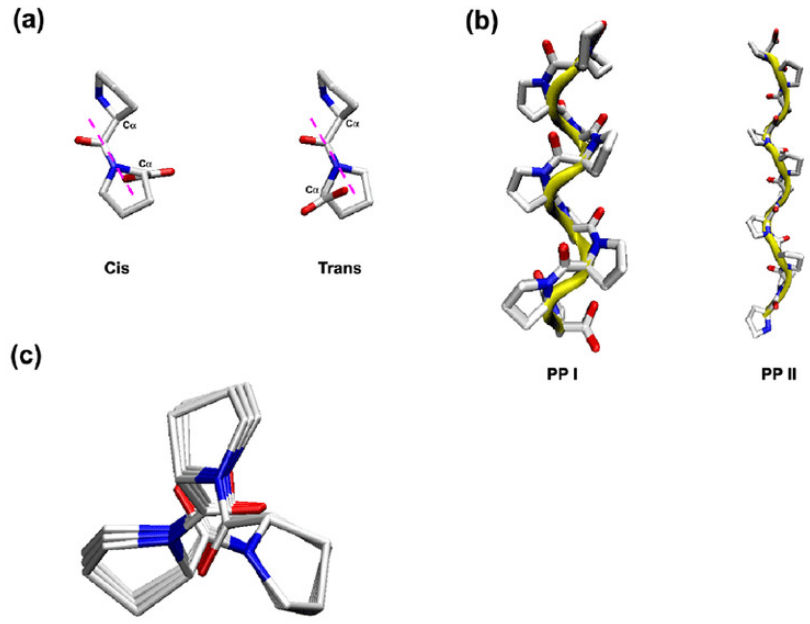


Figure 2. Structure of (a) the cis and trans proline residues; (b) the PPI and PPII helices; (c) the PPII helix viewed along the helical axis. Molecules are

shown as sticks in Corey-Pauling-Koltun colors (Gu and Helms, 2005).

SH3 domains

Proline-rich sequences are among the most abundant expressed amino acid sequence motifs in vertebrate genomes, for example, they are the most common motif in *Drosophila melanogaster* and *Caenorhabditis elegans* genomes (Rubin et al., 2000). The SH3 domains are among the most relevant domain families found in genomes. The number of SH3 domains in proteomes roughly corresponds to its complexity; in fact, the human genome encodes for about 300 SH3 domains, the *D. melanogaster* genome presents about 90 SH3 domains, while the *Saccharomice cerevisiae* proteome contains 28 members of this domain family (Karkkainen et al., 2006); furthermore SH3-like domains are found in some prokaryotes (Whisstock and Lesk, 1999).

SH3 domains are small protein-interaction modules of about 60-70 residues able to independently fold; they are present in a large number of proteins involved in several cellular processes, such as cytoskeletal arrangement, intracellular signalling, protein trafficking, degradation, immune response and many other (Li, 2005; Mayer, 2001; Zarrinpar et al., 2003). The structure of a SH3 domain is a β-barrel, formed by five or six β -strands, arranged into two perpendicular β -sheets. Each β -sheet is formed of three anti-parallel β -strands. The β -strands are linked by an RT-loop, an n-Src loop, a distal loop and a short 3_{10}-helix. Its binding surface comprises three discrete patches: two hydrophobic grooves lined mainly by aromatic residues to accommodate the xP dipeptides (x represents, in most cases, a hydrophobic amino acid) in the PPII helix and a specificity pocket, formed by residues primarily from the RT and n-Src loops (Lim et al., 1994; Musacchio et al., 1992; Yu et al., 1994; Yu et al., 1992) (Figure 3). Since loops are generally variable in sequence and flexible in structure, they play important roles in modulating the specificity of a domain (Feng et al., 1995). Positively charged residues, such as arginine and lysine, have been known to play an important part in the binding to an SH3 domain, not only by providing additional binding energy through electrostatic interactions with residues in the specificity pocket, but also by orienting the ligand with respect to the binding groove on the SH3 domain (Feng et al., 1994; Feng et al., 1995; Wu et al., 1995).

Figure 3. Structure and binding mechanism of SH3 domains. The structure of the Sem5 SH3 domain in complex with a proline-rich ligand is shown. A cartoon of the proline-binding surface of these domains docked with a ligand, showing the general mechanism of recognition, is shown below. The core recognition surface has two xP binding grooves formed by aromatic amino acids, shown in yellow, and the adjacent, less conserved specificity pockets are designated in green (Zarrinpar et al., 2003).

SH3 recognition specificity.

Proline-rich regions are present in about 25% of human proteins (Li, 2005). It is of interest to understand how hundreds of different SH3 domains encoded by the human genome, selectively bind their respective

physiological partners to execute or regulate specific cellular functions. Moreover, recent findings show that a number of SH3 domains are capable of binding also to non proline-rich sequences while others bind to proteins via tertiary contacts (Kaneko et al., 2008), rather than through a defined sequence motif. These observations significantly expanded the pool of proteins these domains can interact with.

However, most of the interactions mediated by SH3 domains occur via the recognition of short peptide ligands. Hereafter, I will describe a classification of the SH3 domain family based on binding specificity.

Class I and Class II proline-rich sequences: canonical binding motifs.

Over the last 20 years the phage display technique has been useful to identify the recognition specificity of several SH3 domains. In 1994 Feng et al described the binding modality of the Class I and Class II peptides (Feng et al., 1994).

In the following years it became clear that the SH3 domains of class I recognize sequences with a positive charge at their N terminus; the consensus sequence being (R/K)xΦPxΦP, while the SH3 domains of class II bind in reverse orientation peptides containing a positively charged residue at their C terminus; the consensus sequence of this class being ΦPxΦPx(R/K), where Φ is an hydrophobic residues while 'x' denotes any naturally occurring amino acid.

The resolved structure of an SH3 domain in complex with either a class I (Janz et al., 2007) or II (Massenet et al., 2005) peptide revealed that these two types of ligand bind to an SH3 domain in opposite orientations (Lim et al., 1994; Yu et al., 1994). Depending on the characteristics of the binding site, some SH3 domains prefer one among the two orientations, while others make little distinction between them (Figure 4).

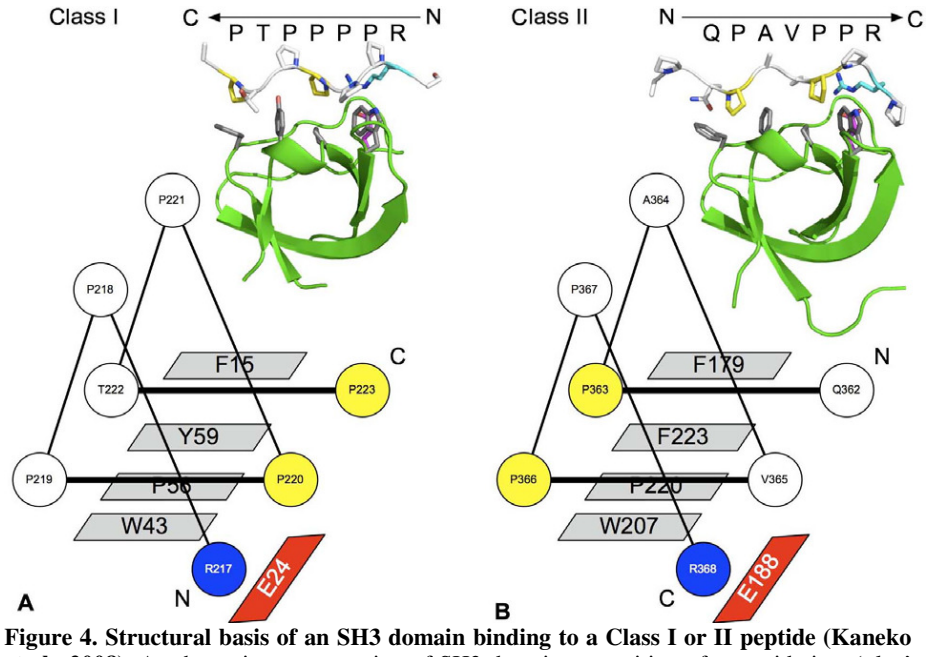


Figure 4. Structural basis of an SH3 domain binding to a Class I or II peptide (Kaneko et al., 2008). A schematic representation of SH3 domain recognition of a peptide in a 'plus' (C←N) (left) or 'minus' (N→C orientation (right) (Lim et al., 1994). The left-handed PPII helix of the peptide is shown as circles (residues) connected by sticks (amide bonds). Bold lines represent the XP dipeptide units. Conserved residues found at the ligand-binding site of the SH3 domain are shown in rectangles. (A) The beta-PIX SH3 domain in complex with a peptide derived from AIP-4 (PDB code: 2P4R) (Janz et al., 2007). The peptide contains the class I motif ([R/K]xXPxXP) and is bound in the plus orientation. (B) The $p40_{phox}$ SH3 domain in complex with a $p47_{phox}$-derived peptide (PDB code: 1W70) (Massenet et al., 2005). The peptide contains the class II motif (XPxXPx[R/K]) and is bound to the SH3 domain in the minus orientation.

The orientation of a peptide with respect to the binding site is determined by an Arg or a Lys residue that precedes or follows the PxxP core motif. In either case, the basic residue is recognized by a negatively charged pocket (specificity pocket) on the SH3 domain. The basic residue contributes significantly to domain affinity. Apart from the electrostatic interaction with acidic residues lining the specificity pocket, the side chain of an Arg or a Lys makes favorable charge-aromatic interactions and/or Van der Waals contacts with a Trp residue that is conserved at the binding site of all SH3 domains (Li, 2005).

The two XP dipeptide units (as in XPxXP) are engaged by two binding pockets on an SH3 domain through hydrophobic interactions. Although

these pockets are not deep grooves, they are remarkably specific for the XP dipeptide.

Overview about SH3 atypical motifs.

The vast majority of SH3 domains recognize Class I and/or Class II peptides, but several alternative SH3 domains binding modalities have also been described. In fact specific domains or families of SH3 domains often display and unconventional binding mode and the target peptides do not match the classical Class I and II consensus sequences.

- *PxΦxxPxxP consensus*: the SH3 domain of Abl, as first identified by phage display assays (Rickles et al., 1994), recognizes proline-rich motifs without basic residues flanking the PxxP core; it is enriched in hydrophobic amino acids.
- *PxxDY consensus*: the SH3 domain from EPS8 protein is able to bind this specific motif, lacking the canonical consensus PxxP (Mongiovi et al., 1999).
- *RxxPxxxP consensus*: the SH3 from Cortactin binds to this motif in the cytoplasmic tail of the calcium activated potassium (BK) channels (Tian et al., 2006). It is similar to Class I motifs except for the extra residues separating the two prolines
- *PxxxPR consensus*: the SH3 domains of the CIN85/CMS family recognize a proline-arginine motif (Moncalian et al., 2006) that resemble the Class II consensus. Through these motifs CIN85 SH3 domains bind effector molecules such as Cbl, p115-RhoGEF, ASAP1, SHP1 in order to control the intracellular trafficking of epidermal growth factor receptors (Kowanetz et al., 2004).
- *PxRPxR consensus*: this motif is a Class II consensus with a specific Arginine residue in the PxxP core, as first identified in peptide ligands of the SH3 domain of Amphiphysin (Cestra et al., 1999).
- *P+RPPxPconsensus*: this general consensus binds to SH3 from Endophilin proteins (Cestra et al., 1999).
- *RKxxYxxY consensus*: this motif lacks the common PxxP core and does not contain any proline. It was identified in the adaptor protein SKAP-55 to mediate its interaction with the C-terminal SH3 domain of ADAP/SLAP proteins (Kang et al., 2000). The Tyr residue, corresponding to Y294 in SKAP-55, can be phophorylated by Fyn,

which in turn abolishes the interaction between the SKAP-55 motif and the ADAP/SLAP SH3 domain (Duke-Cohan et al., 2006).

- *RxxK consensus*: first described in the deubiquitinating enzyme UBPY recognized by the STAM2 SH3 domain (Kato et al., 2000), this motif was subsequently found to mediate a physiological interaction between SLP-76 and Gads SH3 domains (Liu et al., 2003). The peculiar feature of this motif is the strongly affinity for its SH3. The affinity of the Gads C terminal SH3 for a 14-mer SLP-76 peptide shows a Kd of 8 nM (Seet et al., 2007).
- PPxVxPY consensus (Tong et al., 2002).
- RxxRxxS consensus (Tong et al., 2002).
- RxxxxY consensus (Tong et al., 2002)

Ligand-recognition via tertiary contacts and ligand-mediated SH3 dimers.

A few SH3-mediated interactions of physiological relevance couldn't be recapitulated by the corresponding SH3-peptide motif binding *in vitro*. In fact, increasing evidences suggest that, in addition to peptide recognition, SH3 domains can also associate with partner proteins via tertiary contacts that involve no defined simple linear motif. An example is the interaction of ubiquitin with the SH3 domain from Sla1 (and its mammalian homologue CIN85). The interaction between the Sla1 SH3 domain and ubiquitin does not involve a specific sequence motif; it is instead mediated by hydrophobic, tertiary contacts between the two interacting partners. This represents another layer of control over the dynamic nature of SH3-mediated protein-protein interactions in endocytosis and/or actin polarization (Stamenova et al., 2007).

The same mode of tertiary contact-mediated binding has also been observed in other SH3-domain interactions, such as the Fyn SH3 domain that binds to the SAP SH2 domain (Chan et al., 2003) or the ultra-weak interaction between NCK2 third SH3 domain and PINCH1 fourth LIM domain (Vaynberg et al., 2005).

Some SH3 domains can either omo- or hetero-dimerize. The SH3 dimers however differ from other domain dimers in that the formers are often mediated or stabilized by a peptide ligand that simultaneously engages the two SH3 domains while the latter can form directly between two omo- or hetero-domains. For instance, the two SH3 domains in the phagocyte oxidase protein p47phox, form a "super" SH3 domain, that in the inactive

state, binds to an extended C-terminal sequence in the same protein via an intra-molecular interaction (Groemping et al., 2003). The peptide fragment occupying the interaction groove, RGAPPRRSS, binds simultaneously to the two SH3 domains. Although this peptide has a mediocre but measurable affinity (Kd ~ 29 microM) for the tandem SH3 domains, it failed to bind to two SH3 domain in isolation, underlining the importance of SH3 dimerization to this interaction.

Additional examples of similar cooperative interactions, mediated by two dimerizing domains, have been recently reported in the literature. Notably, a Pro/Arg-rich peptide, PARPPKPRPRR, in the ubiquitin ligase Cbl-b has been shown to bind to two copies of either the CIN85 SH3 domain, the CMS SH3 domain or the beta-PIX SH3 domain, thereby promoting the formation of a heterotrimeric $(CIN85)_2$-Cbl-b, $(CMS)_2$-Cbl-b or $(beta\text{-}PIX)_2$-Cbl-b complex in solution. This is made possible by the pseudo-symmetrical characteristic of the peptide such that it can simultaneously engage the two SH3 domains, one in the Class I orientation and the other in the Class II orientation (Jozic et al., 2005; Moncalian et al., 2006). This peptide-induced SH3 dimerization is also observed for AMAP1 and cortactin (Hashimoto et al., 2006). In this case, a proline-rich peptide from AMAP1 mediates the binding of two cortactin SH3 domains. It is worth noting that the described peptide-mediated SH3 dimerization is different from the p47phox “super” SH3 domain. In the former cases, the two identical SH3 domains are not covalently linked and the change in affinity, from single to double SH3 binding, is not as drastic as for the p47phox SH3 domains. However, since Arg/Lys and Pro often co-exist in the proline-rich region of a protein, it is likely that other peptides that possess a palindrome-like pseudosymmetry, i.e., a pair of positively charged residues flanking a central proline-rich motif, could mediate the dimerization of additional SH3 domains. Peptide-assisted SH3 dimerization therefore provides an effective means for the formation of multimeric protein complexes.

SH3 domains are involved in virus-host interactions.

Viruses modify cell physiology, as well as its regulatory networks by manipulating critical nodes of the underlying cell machinery in order to gain control of fundamental processes, such as apoptosis, cell cycle, DNA replication and cell proliferation, and to giant themself a propagation strategy. Understanding how viral proteins interfere with cellular protein

interaction networks is therefore essential to explain the molecular mechanism of viral infection in detail and to develop efficient antiviral strategies. In addition, the identification of the cellular targets of viral proteins may contribute to reveal new key cell regulators. The ability of viruses to redirect cellular metabolism by interfering with key proteins in fundamental cellular processes have already been described with molecular details. For example, Human adenovirus E1a and E1b proteins target cell cycle regulator protein pRb (Retinoblastoma) and the genome guardian protein p53. The same crucial proteins are targets of HPV (Human Papilloma Virus), Polyomavirus, HHV-8 (Human Herpes Virus), HCMV (Human Cytomegalovirus) as well as other viruses (Dyson et al., 1992; Lazzi et al., 2002; Lechner and Laimins, 1994; Levine et al., 1991).

Our approach aims at studying the network of interactions between viral and host proteins. To this aim, we use two independent approaches. On one hand, we extracted from the literature and deposit in databases all the published information. On the other hand, we used peptide technology to identify, in a high throughput approach, novel interactions mediated by protein interaction domains. This twofold strategy is based on the hypothesis that the subversion of cell physiology induced by viral infection is mediated by alterations of the network of cellular gene products and metabolites. Such networks are responsible for emergent properties that cannot be understood by analyzing, with a reductionist approach, one protein and one interaction at a time.

MINT and VirusMINT databases.

The whole array of physical and enzymatic interactions between organic molecules is a way to understand cell physiology.
Several years ago, the group where I carried out my thesis work developed the MINT database (MINT, http://mint.bio.uniroma2.it/mint/) (Zanzoni et al., 2002). MINT was designed to collect experimentally verified protein-protein interactions (PPIs) in a binary or complex representation and collects physical interactions, direct interactions and colocalizations manually curated according to the Proteomics Standards Initiative-Molecular Interaction (PSI-MI 2.5) standards (Chatr-aryamontri et al., 2007). At now, MINT collects 83162 interactions data connecting 29701 proteins and supported by 114127 experimental evidences resulting from more than 3141 reviewed articles.

Due to the diverse sensitivity and specificity of the techniques used to describe these protein interactions and because of false positive in high throughput experiments not all interactions that are described in MINT are equally reliable. To facilitate users, MINT adopts a scoring system (MINT score) based on i) size (large and small scale) and type (co-immunoprecipitation, two hybrid, ecc) of experimental evidence, ii) number of interaction partners detected in a single purification, iii) number of different publications supporting the interaction and for interaction stored in HomoMINT (see below), iv) the degree of sequence homology.

A number of specialized databases share the MINT structure and data model but have developed specialized tools and offer dedicated web interfaces. HomoMINT is an inferred human protein interaction network where interactions discovered in model organisms, and collected in MINT, are mapped onto the corresponding human orthologs (Persico et al., 2005). DOMINO is a database for domain-protein interactions (Ceol et al., 2007), while VirusMINT is specialized in presenting protein interaction networks involving viral proteins (Chatr-aryamontri et al., 2009).

The VirusMINT database aims at annotating in a structured format all the interactions between human and viral proteins and at integrating this information in the human protein interaction network. Since its first description, VirusMINT has acquired an upgraded user interface and improved algorithms have been developed for automatically grouping interactions reported for the same proteins in different viral isolates. In addition the continuing curation effort has resulted in an increase of the stored interaction information.

The VirusMINT database contains interaction data for 418 proteins encoded by 99 different viral strains, corresponding to 1854 unique interactions supported by 5140 experimental evidences resulting from more than 1568 reviewed articles. The initial focus has been on curation of manuscripts describing interactions of viral proteins encoded by the main medically relevant viruses: Adenovirus, Simian virus 40 (SV40), Papillomavirus, Epsetin-Barr virus (EBV), Hepatitis B virus (HBV), Hepatitis C (HCV), Herpesvirus, Human Immunodeficiency virus (HIV), Influenza A virus and Vaccinia virus. The VirusMINT data set is freely accessible without limitations and can be downloaded in two formats: as a tab delimited flat files and in the PSI-2.5 XML format.

WISE (Whole Interactome Scanning Experiment)

In our work, we have been focusing on the characterization of the entire protein interaction network mediated by SH3 domains and small protein recognition modules in human cells.

To characterize the interactome defined by poly-proline binding domains we have used the WISE (Whole Interactome Scanning Experiment) approach (Landgraf et al., 2004).

Using a peptide based scanning of a proteome, the WISE method permits a rapid identification of the putative partners of any peptide recognition module; however the SPOT synthesis approach imposes some limitations on the number of peptides that can be screened in a single experiments. Thus it is important to have a selection strategy that permits to identify the peptides that are most likely to be target of an SH3 domain. This can be achieved by a combination of Phage Display screening and SPOT-synthesis method. Schematically this approach requires three steps (Figure 5):

- Identification of a consensus recognition sequence (for instance by phage display) and informatic scanning of the whole proteome in search of peptides that match the relaxed consensus.
- Synthesis by the SPOT technique of all the peptides matching the loose consensus.
- Identification of the domain binding peptides by overlay assays.

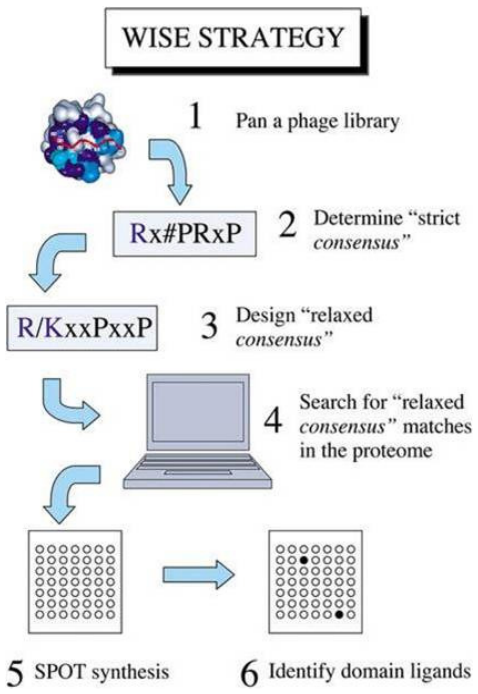


Figure 5 Wise approach (Landgraf et al., 2004).

Obviously, this strategy only helps to identify natural peptides with the potential for binding to any given recognition domain, and even if one can use this information to infer the formation of protein complexes *in vivo*, there are a number of reasons why this inference could turn out to be incorrect. For instance, a peptide could be unavailable for interaction in the native protein structural context. Alternatively, the two inferred partners could be located in different cell compartments or expressed in different tissues or at different times during development. Finally, all the interactions that are mediated by an extended region of a protein surface and that cannot be supported by a relatively short peptide will be missed by this approach.

The WISE approach was originally developed for screening the yeast proteome. In order to apply this strategy to the human proteome we have developed a high density peptide array in which the naturally occurring peptides are pre-synthesized and immobilized at high density on a glass surface, where they are probed for binding.

Searching VirusMINT (http://mint.bio.uniroma2.it/virusmint/) a database that collects protein interactions between viral and human proteins reported

in the literature, we observed that the interactions involving viral proteins and SH3 domains are underrepresented in the database. To fill this gap, we have applied the WISE (Whole Interactome Scanning Experiment) technique to identify interactions between 15 human SH3 proteins and viral proline-rich sequences from oncogenic HPV (Human papilloma virus type 16) and Ad12 (Human Adenovirus A type 12). By this approach we identified 114 new potential interactions between human SH3 domains and the proline rich regions of the two viral proteomes.

Peptide Microarray

Over the past decade, microarray technology has revolutionized the study of gene expression. Arrays of nucleic acids, comprising either single-stranded oligonucleotides or double-stranded PCR products, have been used to measure the abundance of thousands of transcripts in cell or tissue samples. The technology is well suited to large-scale, system-wide investigations for two reasons: (1) it enables many different samples to be interrogated simultaneously in a rapid and economical fashion; and (2) it enables such experiments to be performed hundreds or even thousands of times, with different cells under different conditions. These two features apply equally well to the system-wide study of protein function (Figure 6).

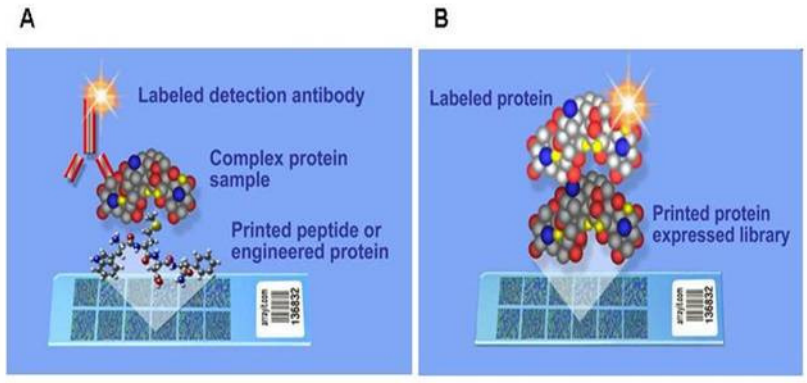


Figure 6. Peptide chip technology. Protein-peptide interactions (A) or protein-protein interactions (B) can be revealed with an antibody conjugated to a fluorophore. Otherwise it is possible to conjugate directly the fluorophore to the protein of interest.

In fact the clear advantage of the array format could then be fully exploited to study protein interaction in those cases where one of the partners participates in complex formation by docking a relatively short peptide into a receptor protein. Our partner Jerini (a company specialized in peptide technology) has developed and reported methods to fabricate microarrays of purified peptides at high spatial density on chemically-derivatized glass microscope slides. In order to perform functional assays, we have developed surfaces that permit covalent attachment of peptides and exhibit low non-specific binding properties. The microarrays of peptides can be probed with the protein of interest to identify stable interactions, so providing a robust way to study protein function in a rapid, economical, and system-wide fashion.

Aim of the work

Over the past decades a large number of reports have contributed to the discovery of a variety of interactions mediated by SH3 domains. For instance, the Domino database contains about one thousand SH3 interaction partners. However, we are still far from a global understanding of the protein mesh mediated by this domain family. More specifically it is not clear whether all or only a few of the cell functions strictly require this network. In addition a question that is often asked is "how is it possible that this class of interactions, that show often low specificity and high dissociation constant, can support pathways displaying higly specific responses to external or intracellular clues?". In other words, what is the interplay between specificity and promiscuity in the interactions mediated by the SH3 domains?

To answer these questions we would like to provide a more global view of the "interaction potential" of many members of these versatile domain family. To this aim I contributed to develop a new technology that allows the parallel screening of a large number of peptides, representing most of the proline rich domains present in the human proteome. I used this approach to profile the recognition specificity of 90 SH3 domains, covering most of the SH3 domain evolutionary tree. This systematic approach, combined with appropriate informatic analysis, is likely to bridge some gaps in our understanding of the cross talk between well-studied biological pathways. In the final part of my project I would like to demonstrate the feasibility of using the WISE approach to discover novel putative interactions between human SH3 proteins and viral proline-rich sequences of oncogenic viruses, such as HPV and Ad12.

Results

SH3 domain collection

The human proteome contains about 300 SH3 domains, belonging to about 220 proteins. I chose to profile a subset of 90 SH3 domains in order to characterize the recognition specificity and the protein interaction map of each SH3 sub-family.

To understand the biological process as well as the molecular function in which all these SH3-containing proteins are involved, we used the DAVID tool (The Database for Annotation, Visualization and Integrated Discovery) available at http://david.abcc.ncifcrf.gov/. The DAVID Bioinformatic resource is able to extract biological features/meaning associated with large gene lists. DAVID is able to handle any type of gene list, no matter which genomic platform or software package generated them (Huang da et al., 2009).

The Uniprot identifiers of all human SH3-containing proteins were used as input in DAVID. The tool associates a Molecular Function or a Biological process to each Uniprot ID and identifies the Gene Ontology terms that are enriched in the specific list of proteins that are used in input. In the pie chart in Figure 7, I plotted the molecular functions and the biological processes characterizing the human SH3 containing proteins according to Gene Ontology. These SH3-containing proteins are associated with several molecular functions (protein kinases, oxidases, G-protein modulators, cytoskeletal adaptor activity, etc.) and are involved in many cellular processes, such as general signal transduction, immunity, cell structure, cell death, etc.. However, many proteins in the family are characterize by a “binding activity”, underlying their involvement in important cellular processes as adaptor proteins (Figure 7).

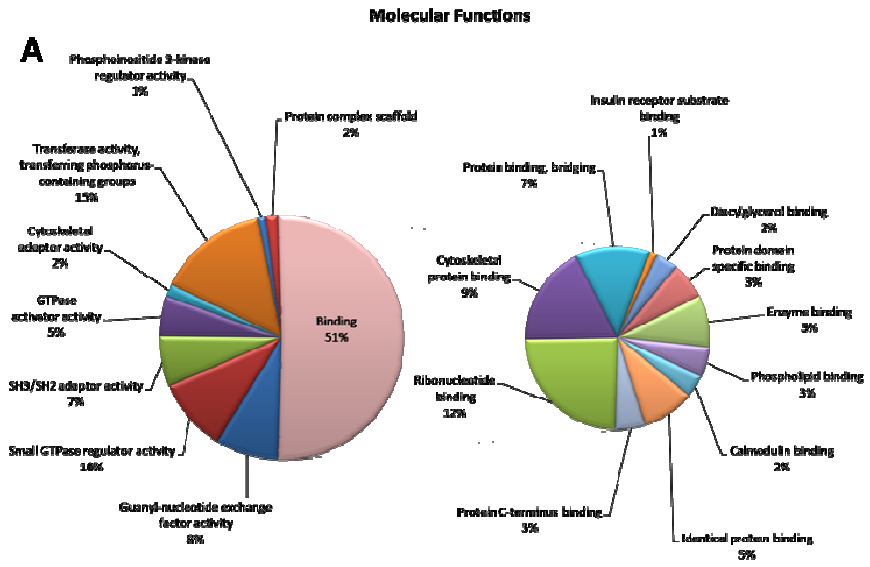


Figure 7 GO term associated to human SH3 domain containing proteins. Distributions of SH3 containing protein in different Molecular Functions (A) and Biological Process (B) according to DAVID.

In principle we wanted to apply our experimental approach to the complete list of the 300 human SH3 domains but for many reasons, such as the insolubility of some domains, their large number and the fact that many SH3 domains are very closely conserved, we focused on a subset of domains as representative of the different sub-families.

Peptide Chips

Over the past years, peptide array technology has considerably improved our ability to study protein interactions at the proteome level. Peptide libraries are a powerful tool for profiling physical interactions where one of the partners participates in the formation of a complex by docking a relatively short peptide into a receptor domain. Most dynamic functions in cells are based on phosphorylated peptides binding SH2 or PTP domains or poly-proline peptides docking their binding domains (SH3, WW, EVH1, GYF domain). The SPOT technology involves the spatially addressed automated synthesis of peptide arrays on various types of flat surfaces (cellulose, polypropylene, glass) (Schneider-Mergener, 2001). This technology provides a robust way to study protein function in a rapid, economical and system-wide fashion. Our partner Jerini has developed methods to produce microarrays of purified peptides at high spatial density on chemically-derivatized glass microscope slides. This strategy allowed the concurrent screening of several thousands peptides while decreasing the amount of the each peptide to be spotted.

In order to characterize the interactome defined by the poly-proline binding domains, we used a variant of the WISE approach (Landgraf et al., 2004). To characterize a representative sample of all the poly-proline rich regions in the human proteome, we first carried out an extensive literature mining work, in order to extract the available published information about interactions mediated by poly-proline peptides. Then we classified the motifs of the interacting proline-rich peptides and we derived 25 regular expressions representing most of the recovered peptides. These regular expressions were then used to scan the human proteome for matching peptides. The sequences of 9600 identified proline-rich peptides were synthesized on cellulose membranes, by the SPOT technique, and arrayed at high density on a glass surface. The peptide length is 15 residues. Each glass chip contains replicas of two identical subarrays (Figure 8).

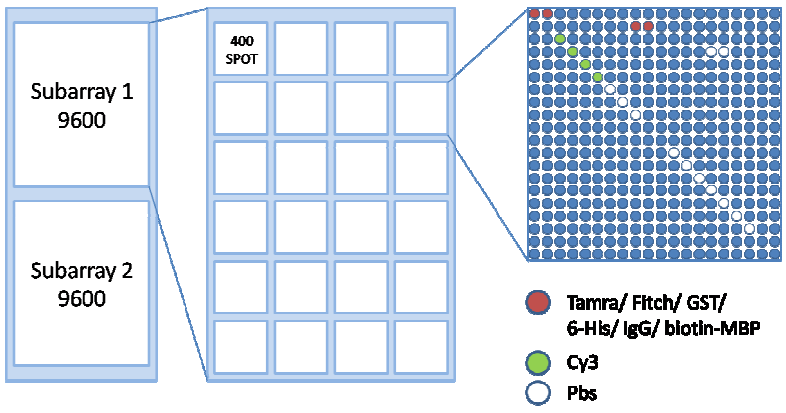


Figure 8 The structure of the proline-rich peptide chips used to profile the specificity of human SH3 domains. A) Two identical subarrays are printed on the chip surface. B) Each subarray is composed of a 6x4 grid of 400 spots. C) Each small grid unit contains the peptides and the control spots reported in legend.

The chips were probed with purified SH3 domains fused to a GST protein tag (Figure 9). Out of a collection of 90 domains, we succeeded in purifying 83 soluble domains. The remaining ones were either degraded or formed inclusion bodies in the host bacterium. 12 of the 83 soluble SH3 domains did not produce satisfactory results when used as probes on the peptide chips and were not considered further. It is possible that these domains don't fold correctly outside the native protein and/or cellular context or alternatively they bind peptides that are not present on our chip. In conclusion most of the tested SH3 domains gave satisfactory results, thus allowing the reliable identification of potential peptide ligands.

The interaction between proline-rich peptides and binding domains was revealed with an anti-GST antibody conjugated to a fluorescent group (Cyanine-5) (Figure 9). In order to exclude non-specific binding, peptide chips were probed with the GST protein alone (data not shown).

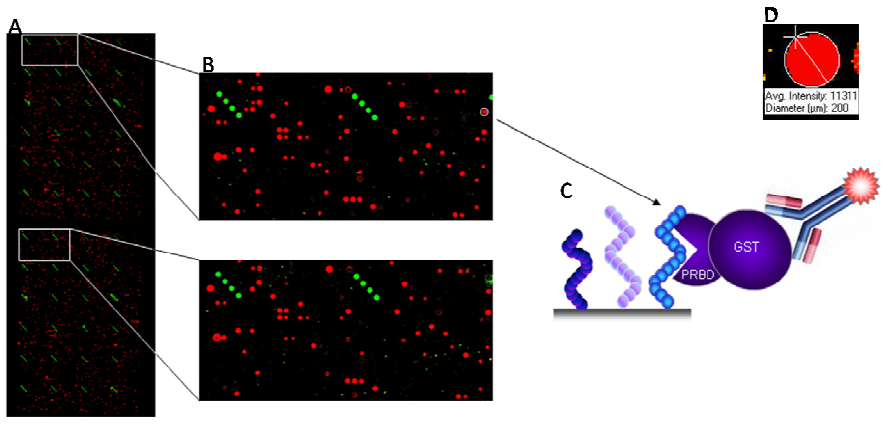


Figure 9 SH3 domain binding assay. In A) the peptide chip probed with Nphp1 SH3 domains is shown; the same two areas of duplicates are enlarged in B) to show the reproducibility of the duplicate experiment. In C) the cartoon represents the interaction between the domain and the peptides in the array: an anti-GST antibody recognize the peptide bound GST fusion domain. In D) a single spot is shown: its diameter is about 200 micrometer.

To evaluate the reproducibility between two subarrays we calculated their Pearson correlation coefficient (PCC) (it is a measure of the correlation (linear dependence) between two variables X and Y, giving a value between +1 and −1 inclusive). For most experiments the correlation between the two replica arrays is high (>0.9 PCC). Experiments with a low Pearson coefficient (<0.7) were not considered.

The SPOT experiment provides a quantitative output. To detect the peptide targets of each SH3 domain, the foreground and background intensity values for each spot were computed by averaging the values of the two replica experiments.

Classification of peptides according to their signal intensity.

We considered that a successful binding reaction between the peptide spotted on that position in the chip and the tested domain had occurred, when the signal intensity of a spot exceeded a certain threshold. For each experiment, the threshold was set to the median signal intensity plus two

times the standard deviation. This operation is tainted with a certain degree of arbitrariness, since chosen threshold is necessary arbitrary, at least to a certain extent; yet, it is a reasonable solution that works well in most practical cases. Then, in order to classify spots according to their foreground signal, we distinguished between three classes of binders: true binders, potential binders and non-binders, identified with 2, 1 and 0 respectively in the data files. True binders were defined as spots with signal intensity above the binding threshold and having their fg_flag set to 'GOOD'. Spots with either signal intensity lower that the binding threshold or fg_flag set to 'BAD' were classified as potential binders. Finally, non-binders were those spots with both low signal and low foreground intensity (fg_flag = 'BAD') (see experimental procedures).

Storage of Experimentally Identified Binders: PepspotDB

If we consider that the outcome of the profiling of just one domain includes thousands of signal intensity measures, it is evident that our project produced an enormous amount of data.

To facilitate data storage, retrieval and analysis, we developed a new database, PepspotDB, designed to store in a single integrated resource the outcome of all experiments concerning molecular interaction assays exploiting molecular array technology, along with publicly available information relevant to the issues we set out to investigate. This facilitated the fruition of the data and the formulation of new testable biological hypotheses from them.

PepspotDB (http://mint.bio.uniroma2.it/PepspotDB/) not only contains results derived from the SH3-peptide interaction project, but also stores information on protein sequences and properties imported from an external database, UniProtKB (The UniProt Consortium 2007, The UniProt Consortium 2009), the reference repository for protein records.

Figure 10 shows a screenshot of the PepspotDB web site from which it is possible to reach the Search page that allows quick retrieval of domain-peptide interactions. When the proteins containing the SH3 domain and the one containing the peptide are selected, the interactions and features involving the two proteins of interest are displayed by clicking the search button. Alternatively, from the homepage it is possible to select the “Browse Project Section” that summarizes the outcome of the SH3 Human Interactome Mapping by showing the hierarchical clustering of the human

SH3 domains. Upon mouse hovering, the logos popping up show the domain specificity. Domains similarly colored have similar consensus sequences, as apparent from the logos popping up. The white nodes indicate a SH3 domain that we have not profiled yet. A click on a tree leaf forwards the user to the search page, where the identified candidate domain targets, for that particular SH3 domain, are listed.

Figure 10 Pepspot DB: the Advanced search page. The search can be performed by name, accession number or peptide sequence of interest. Here I have shown an example of search for interactors of the Cin85 SH3 domains.

Recognition specificity

The binding of proline-rich peptides in the SH3 domains binding pocket is mediated by conserved amino acids. Most of the SH3 domains characterized so far bind to proline-rich peptides containing the core PxxP sequence, flanked by basic residues either at the carboxy or amino terminal side.

The peptide microarray experiments allow us to determine the consensus sequences of analyzed SH3 domain members.

In order to determine the sequence context of the preferred ligand of each SH3 domain we tested, the sequences of the inferred interactors were aligned to identify the preferred residues. The alignments were automatically obtained by mapping the regular expression for a particular SH3 domain to the positive peptide list. The amino acid composition of peptide binders can

be visualized with Sequence Logos, displaying the amino acid enrichment at each position in the binding peptides.

Some SH3 domains display more than one sequence preference, since they bind peptides that matched to multiple sequence motifs. By our approach we have been able to describe the sequence preferences of about 50 previously uncharacterized SH3 domains.

Results of the chip profiling experiments.

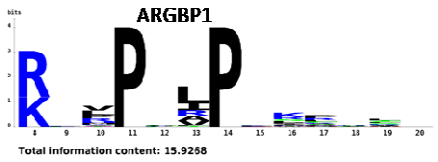
Class I and Class II peptides

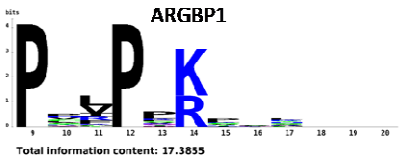
The most populated recognition specificity class is represented by domains that can bind both Class I and Class II peptides. Thirty six SH3 domains have recognized class I and II peptides (Figure 11). In most cases an equal fraction of peptides of either classes bound to the SH3 domain with equal signal intensity.

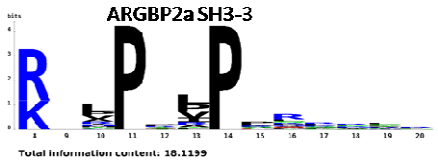
Only ten of these 36 SH3 domains have been previously characterized and some of their ligands have been described in the literature.

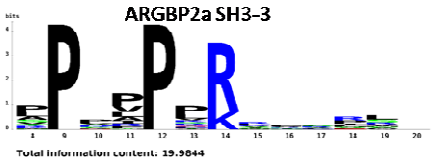
The SH3 domains belonging to c-Src selected both Class I and Class II peptides, as already demonstrated by several authors (Cheadle et al., 1994; Rickles et al., 1994; Yu et al., 1994). The SH3 domains of the Fyn, Lyn and Pik3R1 proteins bind to Class I peptides, as shown in literature (Rickles et al., 1994). In my experiments these domains showed equal preference for Class I and Class II peptides. The third SH3 domain of Nck2 protein selected Class II peptides as suggest previously (Liu et al., 2006), but also Class I peptides. Thus our experiments have demonstrated a broader recognition specificity than previously reported. About 70% of the SH3 domains belonging this specificity class have not been characterized previously.

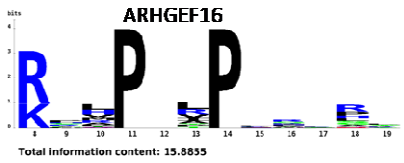
The N-terminal SH3 from Grb2 and the SH3 from Lck and Ncf1 (SH3-2) did not bind to peptides matching the consensi already described in the literature but they selected Class I and Class II peptides.

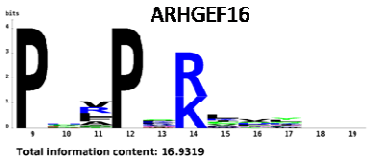

ARGBP1
Total information content: 15.9258

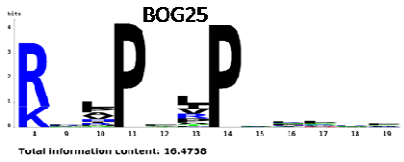

ARGBP1
Total information content: 17.3855


ARGBP2a SH3-3
Total information content: 16.1199

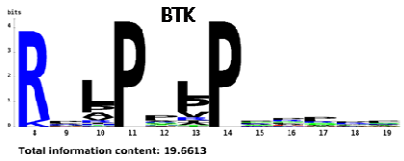

ARGBP2a SH3-3
Total information content: 19.9844

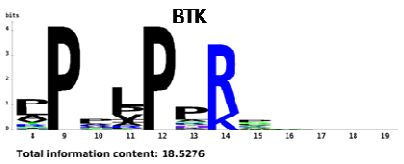

ARHGEF16
Total information content: 15.8855

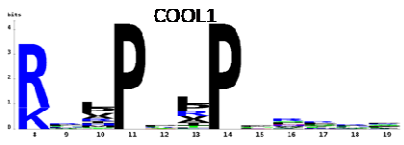

ARHGEF16
Total information content: 16.9319

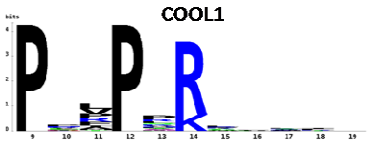

BOG25
Total information content: 16.4736

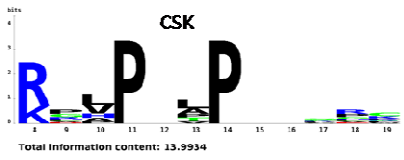

BOG25
Total information content: 16.7575

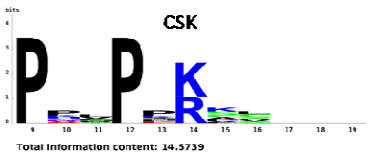

BTK
Total information content: 19.5613

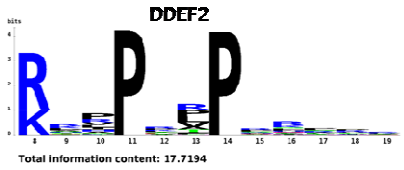

BTK
Total information content: 18.5276

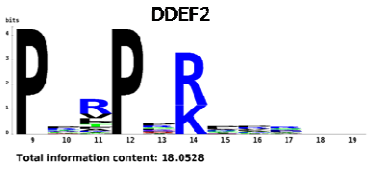

COOL1
Total information content: 17.0462

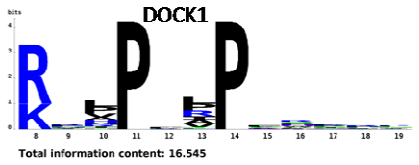

COOL1
Total information content: 17.1437

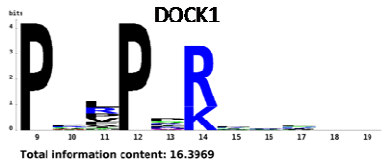

CSK
Total information content: 13.9934

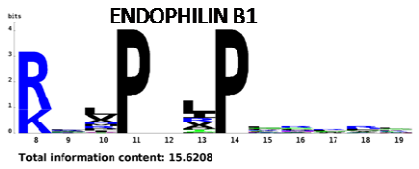

CSK
Total information content: 14.5739

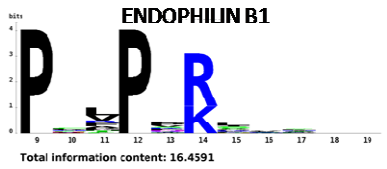

DDEF2
Total information content: 17.7194

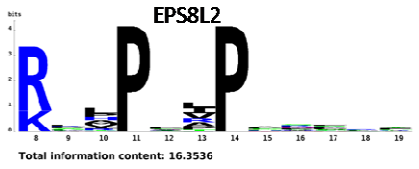

DDEF2
Total information content: 18.0528

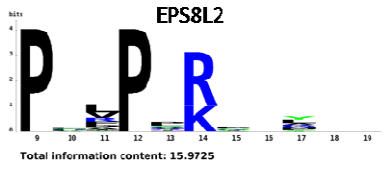

DOCK1
Total information content: 16.545

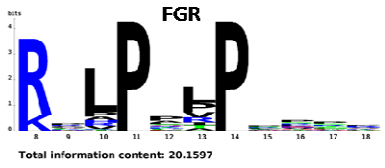

DOCK1
Total information content: 16.3969

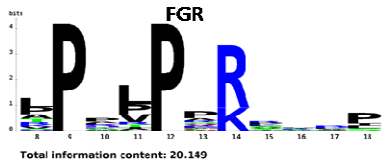

ENDOPHILIN B1
Total information content: 15.6208

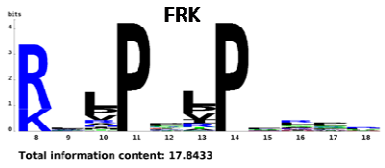

ENDOPHILIN B1
Total information content: 16.4591

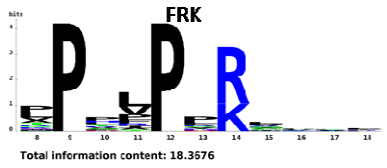

EPS8L2
Total information content: 16.3536

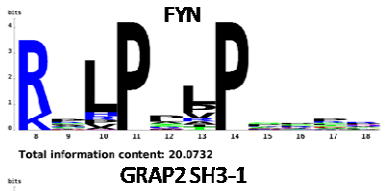

EPS8L2
Total information content: 15.9725

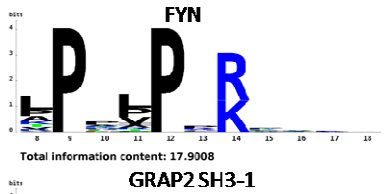

FGR
Total information content: 20.1597

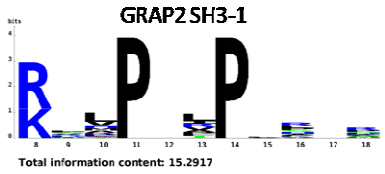

FGR
Total information content: 20.149

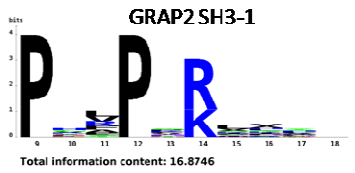

FRK
Total information content: 17.8433

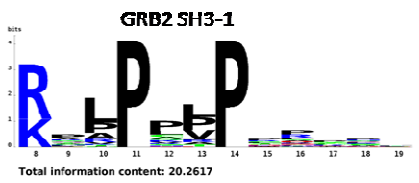

FRK
Total information content: 18.3576

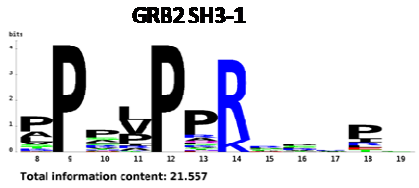

FYN
Total information content: 20.0732

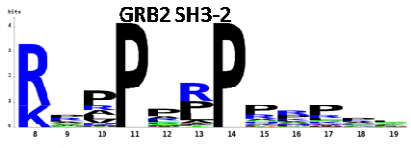

FYN
Total information content: 17.9008

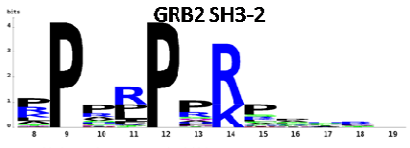

GRAP2 SH3-1
Total information content: 15.2917

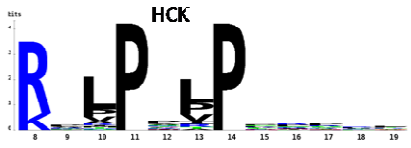

GRAP2 SH3-1
Total information content: 16.8746

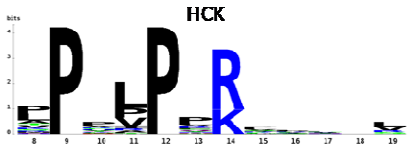

GRB2 SH3-1
Total information content: 20.2617

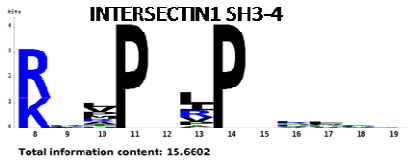

GRB2 SH3-1
Total information content: 21.557

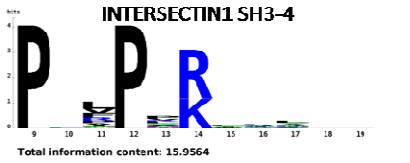

GRB2 SH3-2
Total information content: 21.1552

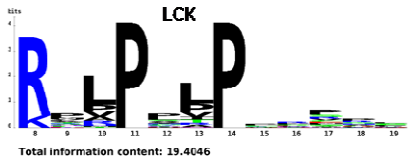

GRB2 SH3-2
Total information content: 21.3686

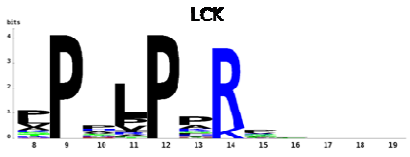

HCK
Total information content: 19.5357

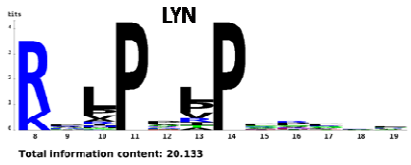

HCK
Total information content: 19.8241

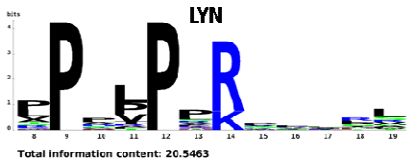

INTERSECTIN1 SH3-4
Total information content: 15.6602

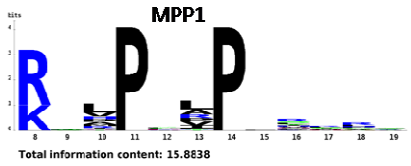

INTERSECTIN1 SH3-4
Total information content: 15.9564

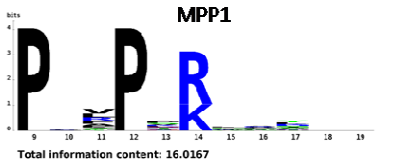

LCK
Total information content: 19.4046

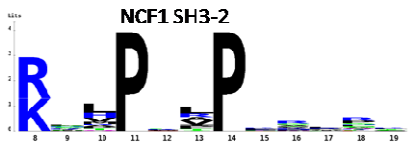

LCK
Total information content: 19.1022

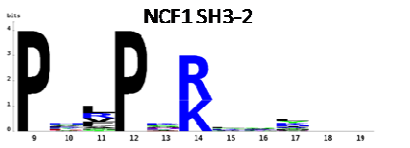

LYN
Total information content: 20.133

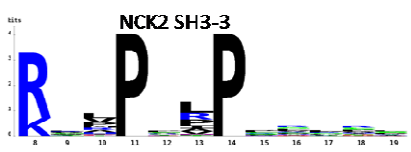

LYN
Total information content: 20.5463

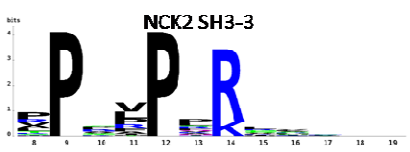

MPP1
Total information content: 15.8838

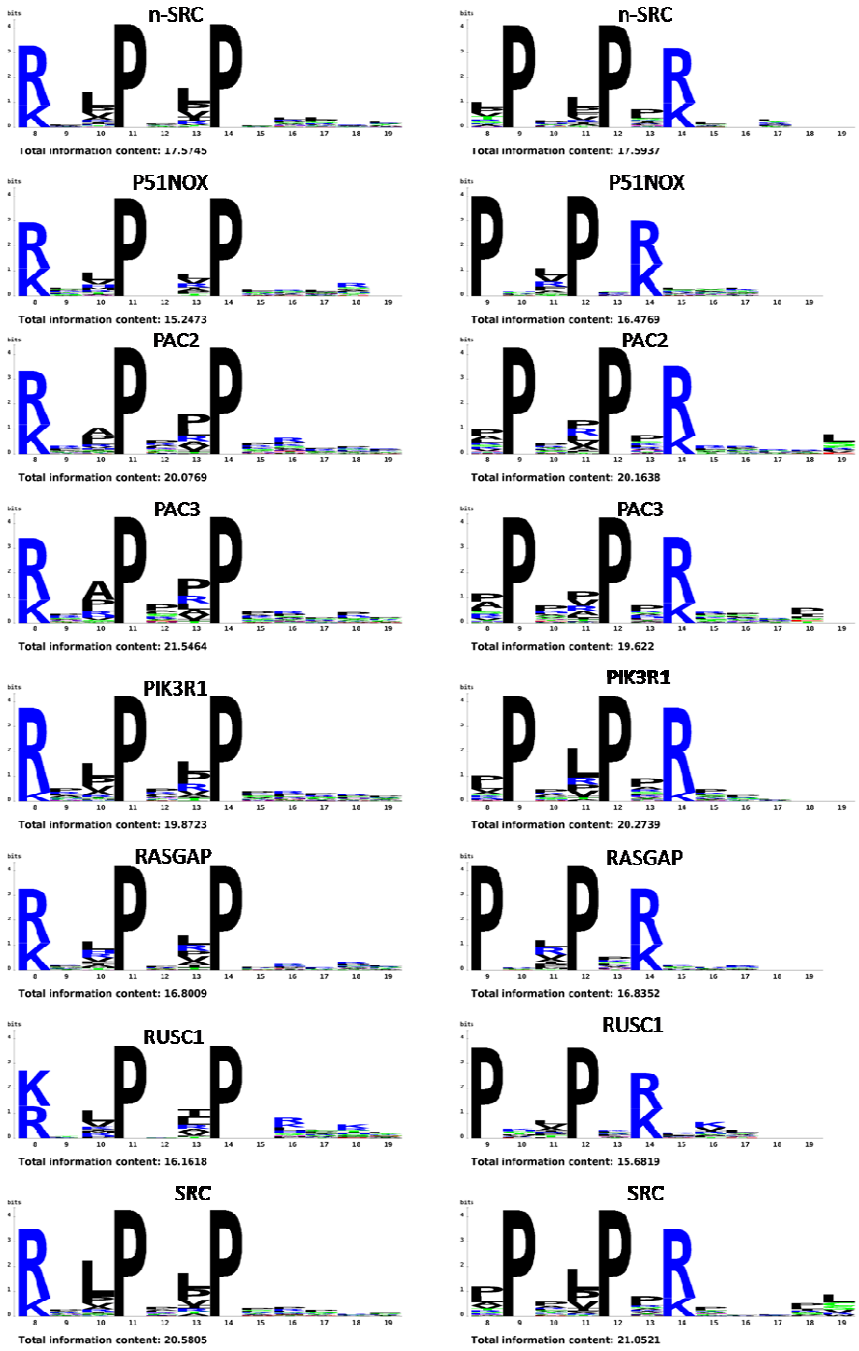

MPP1
Total information content: 16.0167


NCF1 SH3-2
Total information content: 15.6288


NCF1 SH3-2
Total information content: 16.5934


NCK2 SH3-3
Total information content: 16.4108


NCK2 SH3-3
Total information content: 17.9708


n-SRC
Total information content: 17.5745
n-SRC
Total information content: 17.5937
P51NOX
Total information content: 15.2473
P51NOX
Total information content: 16.4769
PAC2
Total information content: 20.0769
PAC2
Total information content: 20.1638
PAC3
Total information content: 21.5464
PAC3
Total information content: 19.622
PIK3R1
Total information content: 19.8723
PIK3R1
Total information content: 20.2739
RASGAP
Total information content: 16.8009
RASGAP
Total information content: 16.8352
RUSC1
Total information content: 16.1618
RUSC1
Total information content: 15.6819
SRC
Total information content: 20.5805
SRC
Total information content: 21.0521

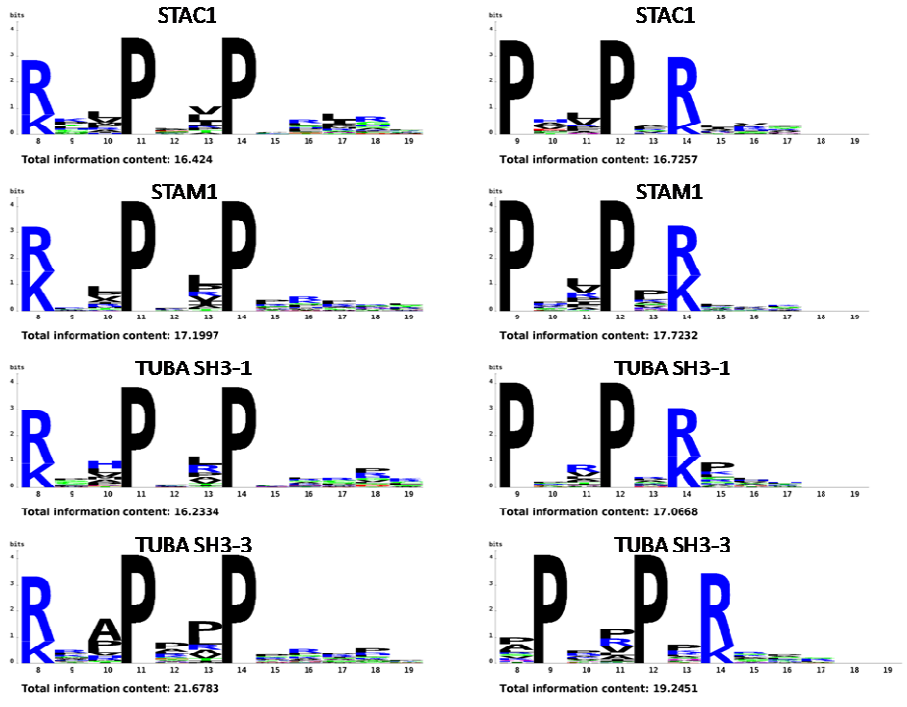


Figure 11. Class I and II peptide motifs. The recognition specificity of 36 SH3 domains that recognize both Class I and II peptides are shown in the left and right respectively in the figure.

While the majority of domains in this SH3 group preferred arginine as positively charged residue at position P-3 and P6 in Class I and Class II peptides respectively, but few of them, such as Tuba SH3-1 and Rusc1, displayed an at least equal preference for lysine.

Class I peptides

The peculiarity of Class I peptides is the presence of a positive charge at position -3, with respect to the first proline in the core PxxP motif. In our experiment we identified five SH3 domains that bind to Class I peptides with high specificity (Figure 12).

The SH3 domain from NPHP1 selected peptide with high frequency of leucine in position P-1 and a less conserved hydrophobic amino acid at position P2 (preferentially Leucine and Proline). This recognition pattern is shared also by the second and third SH3 domains from the Rimb1 protein.

The members of the SH3 domains of the Sorting Nexin protein family bind to Class I peptides with preference for alanine or proline at position P-1 and hydrophobic amino acids at position P2 and P3, as shown in Figure 12 for the SH3 domains of SNX9 and SNX18.

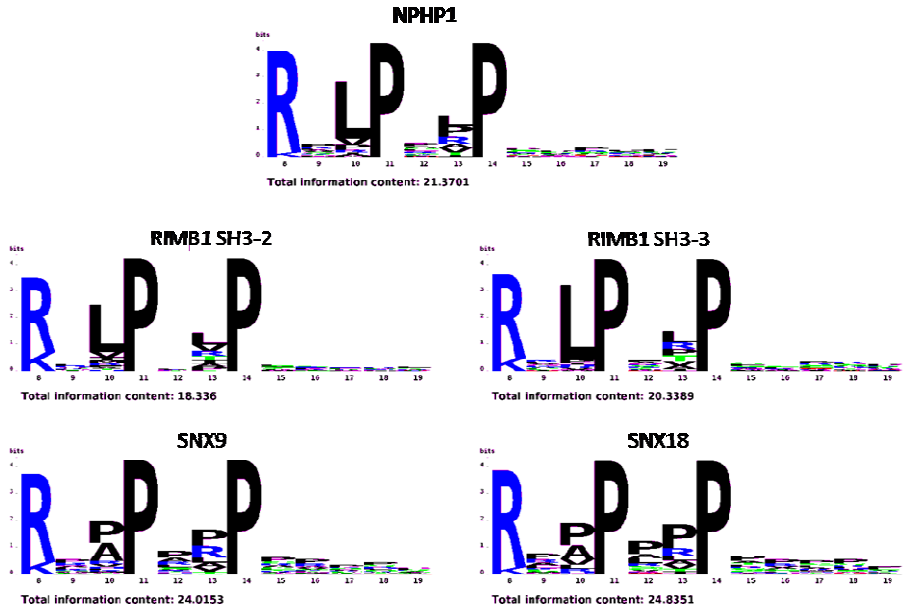


Figure 12 Class I peptide motifs. The recognition specificity of 5 SH3 domains that recognize Class II peptides are shown.

Class II peptides

Class II peptides are characterized by a positively charged residue at position P5; the consensus sequence is PxxPx(R/K). The basic residue is often arginine but also lysine is acceptable. The preference for arginine or

lysine modulates the recognition specificity within this class.

According to our results, nine SH3 domains in our collection bind preferentially to Class II peptides (Figure 12). The first SH3 domain of Intersectin 1 recognized Class II peptides, as already described in other reports using phage display as core technology (Tong et al., 2000). Its target peptides showed the presence of a conserved hydrophobic residue in position P+2 and a less conserved hydrophobic residues in position P-1 and P+4. Arginine is the preferred positively charged residue at position P+5 with lysine being present in only 20% of the cases. In our screening the third and fifth SH3 domains from Intersectin 1 displays a specificity that is very similar to the one of the first SH3 domain. A similar class II consensus also describes the preferred peptide ligands of the SH3 domains from Amphiphysin 2 and Mlk3 and OSFT1.

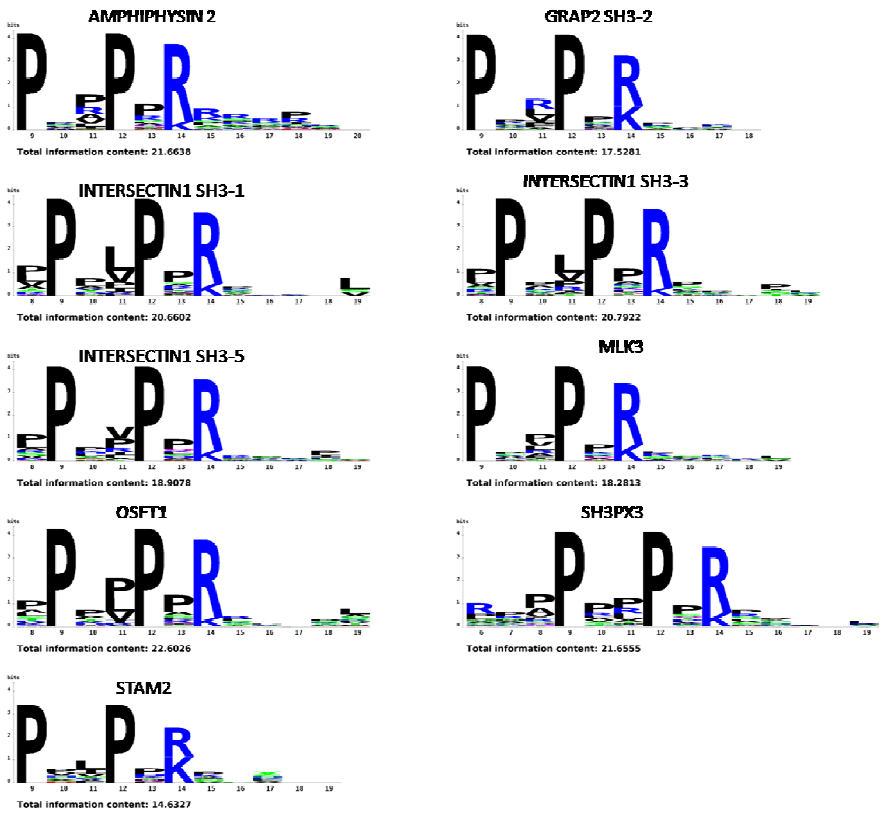


Figure 13 Class II peptide motifs. The recognition specificity of 9 SH3 domains that recognize Class II peptides are shown.

Cin85/CMS peptides

A Target Assisted Iterative Screening approach (TAIS) was used to identify the consensus sequences of SH3 from Cin85 protein family (Kurakin et al., 2003). These authors identified in Cin85 ligands an unconventional consensus characterized by the motif PxxxxR, resembling a Class II motif in which the Proline at position P+3 is substituted by any amino acid. The first and the second SH3 domains have the same sequence target, while the third SH3 domain is more specific in the position before the positive charge. Our experiments confirmed the well-known consensus for the first and third SH3 domains from Cin85.

This consensus was described as peculiar of Cin85 SH3 domains but in our experiments, among the peptides bound by the Cool1 SH3, the fifth SH3 of Intersectin1 and the MLK3 SH3 domains, 20, 40 and 33 per cent, respectively, bind to peptides matching this unconventional consensus in addition to Class I and/or II sequences (Figure 13).

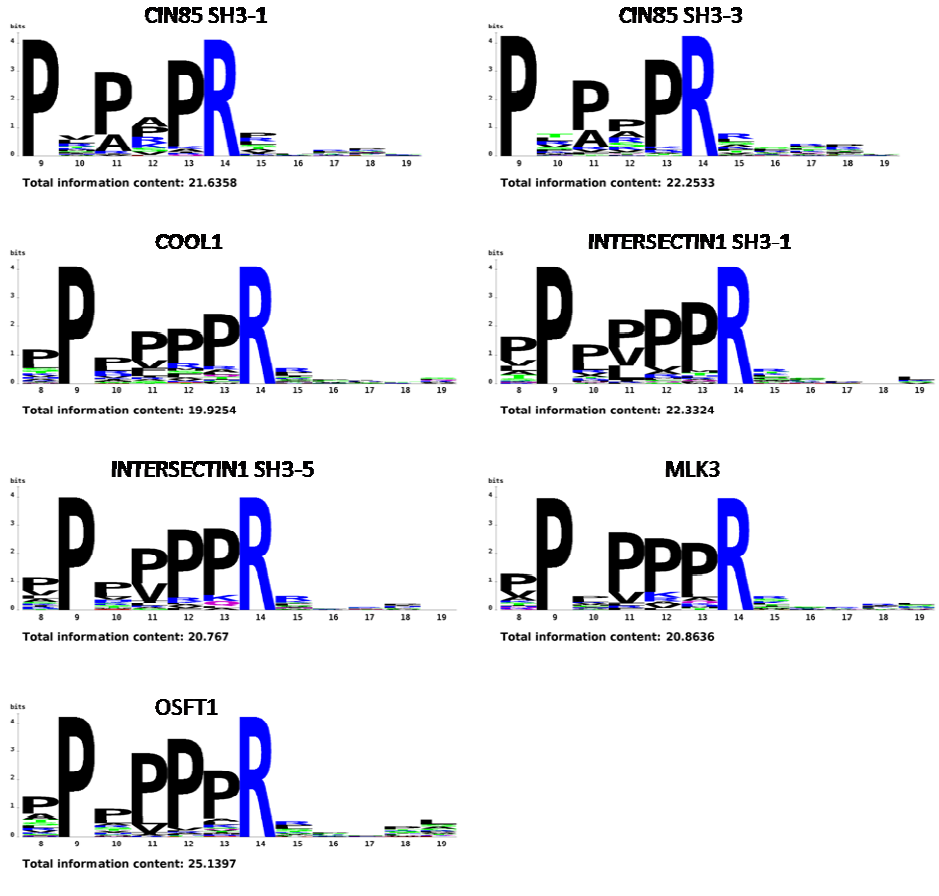


Figure 14 RxxxxR peptide motifs. The recognition specificity of 7 SH3 domains that recognize RxxxxR peptides are shown.

Uncharged proline rich motifs.

The first evidence of consensus sequence of the peptide ligands of the SH3 domain from Abl was obatained by a Phage Display screening (Rickles et al., 1994) in which this SH3 domain was shown to bind PPPYPPPP(I/V)PXX. A number of additional authors contributed to define this consensus; Sparks and collaborators in an independent phage display

screening proposed the consensus PPXΩXPPPΨP (Sparks et al., 1996). The peculiarity of this consensus is a high frequency of Proline and other hydrophobic residues in several positions of the consensus. This feature makes the alignment of the peptide sequences bound in the peptide chip approach a difficult task since many different alignments are possible. Thus the alignments shown in Figure 14 are affected by a certain degree of arbitrariness. Nine additional SH3 domains selected, similarly to Abl, uncharged proline rich peptides.

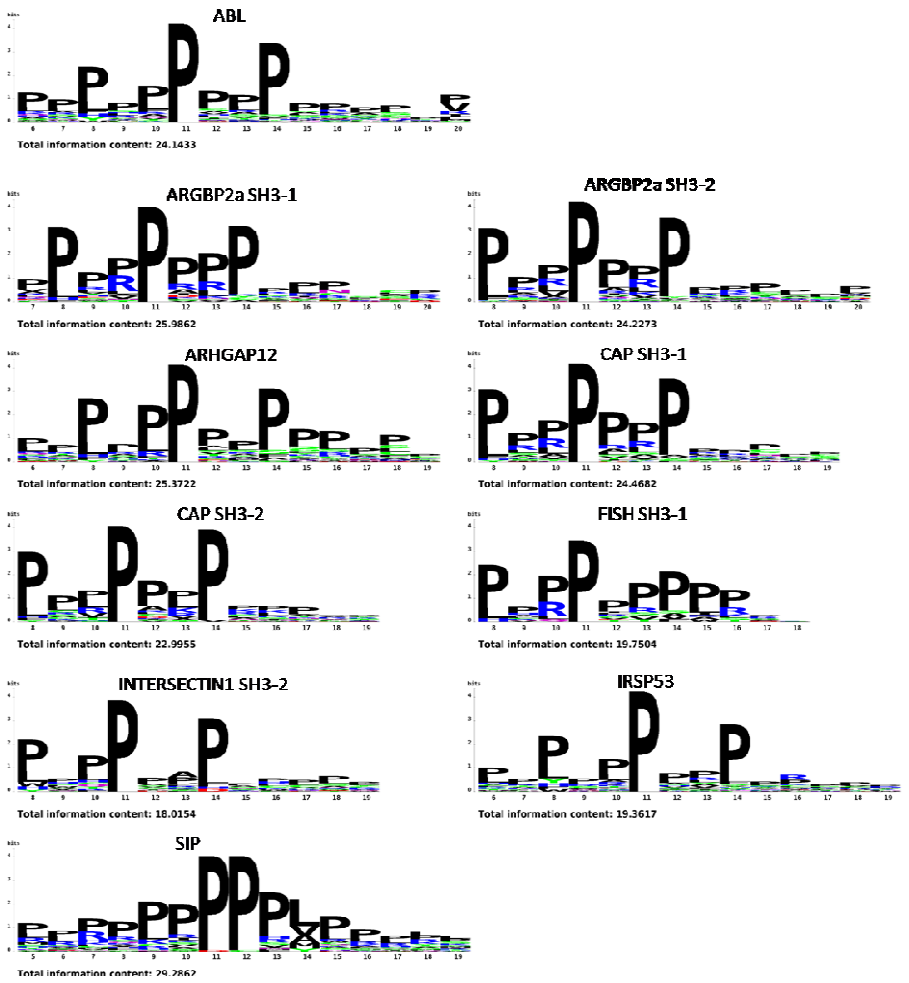


Figure 15 Uncharged proline rich peptide motifs. The recognition specificity of 10 SH3 domains that recognize uncharged proline rich peptides are shown.

PxxDY domains

Over the past recent years it became clear that a few SH3 domains can recognize atypical sequence motifs such as for instance the SH3 domain from the Eps8 family. The consensus sequence for this domain does not contain a PxxP core but requires the presence of a PxxDY motifs characterized by a single conserved proline followed by the conserved dipeptide DY (Mongiovi et al., 1999). Recently it was shown that another SH3 domain (first SH3 domain from Nck1) can recruit PxxDY motifs (Kesti et al., 2007) but from the results we have obtaind we can conclude that only the SH3 from Eps8 can bind peptides containing this motif (Figure 15).

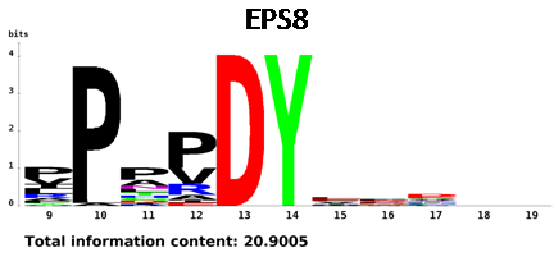


Figure 16 PxxDY peptide motifs. The sequence logo characterizing the ligands of the Eps8 SH3 domain is shown.

Amphyphisin-like motifs

Few years ago our group identified an atypical SH3 consensus by phage display approach (Cestra et al., 1999). This amphyphisin-like motif consists of a proline-rich consensus with a conserved arginine residue at position P2 between the two proline in the core PxxP. In the case of the amphyphisin SH3 domain recognition consensus, as determined by the chip technology, the first proline is substituted by hydrophobic residues and an additional positive residue of a typical class II position.

In our experiments nine SH3 have recognized this motifs (Figure 16). The third SH3 domains of intersectin1 bound to these peptides as well as to Class II peptides. A small percentage of peptides selected by P51Nox, Pik3R1 and Grb2 C-terminal SH3 domains can also be matched to this

consensus.

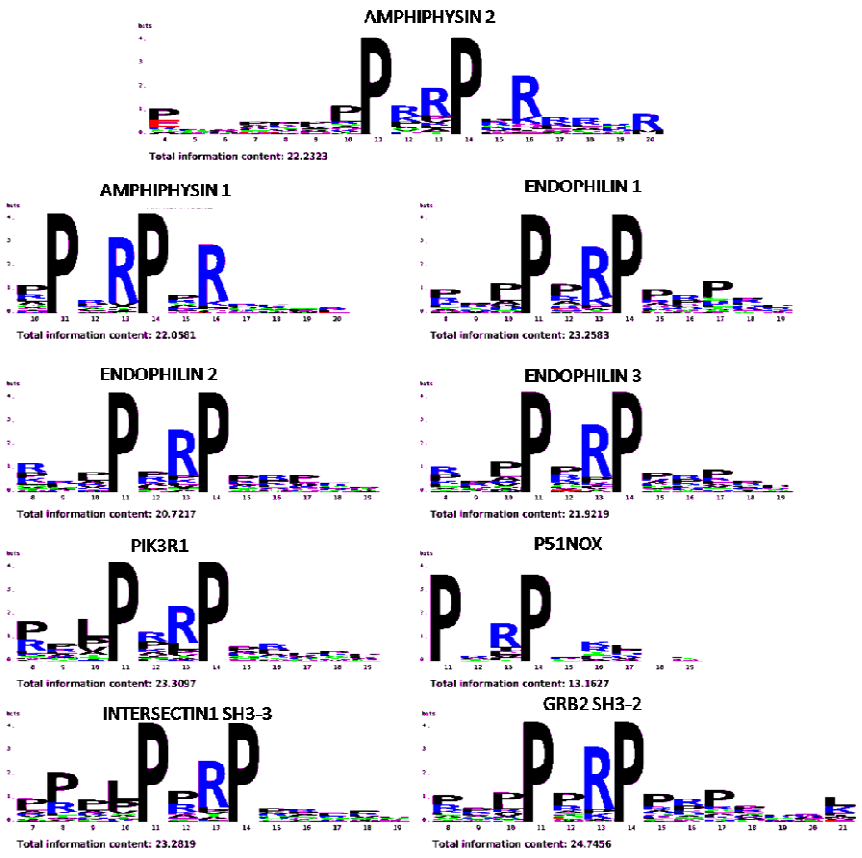


Figure 17 Amphyphisin-like peptide motifs. The nine logos of this particular class of recognition specificity, as determined from chip experiments, are shown.

RxxK motifs.

The members of the Mona/GAD protein family were described to bind to peptides containing the RxxK motif via their C terminal SH3 domains (Lewitzky et al., 2004). In our experiments we confirmed that the C terminal SH3 domains of Grap2 and Stam1 recognize this motif although most peptide ligands were typical Class I and Class II peptides. We didn't

discover any new SH3 domain binding to this unconventional motif. (Figure 17).

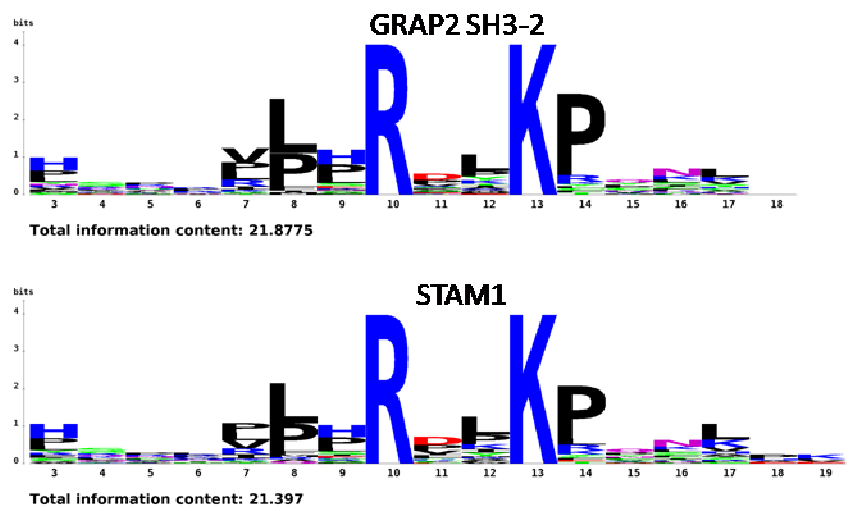


Figure 18 RxxK peptide motifs. The only two logos for this class are shown.

Phage Display experiments.

Since a few SH3 domains didn't provide an interpretable binding pattern in our chip experiments, we decide to evaluate their binding activity with an independent method, phage display.

To this end we selected 21 SH3 domains, 17 had given good results in chip experiments and 5 had not recognized any peptide on glass chip.

Protein	**Domain**	**Peptide Chips Result**
ARGBP1	SH3	Positive
ARGBP2a	SH3-2	Positive
ARHGAP12	SH3	Positive
CAP	SH3-2	Positive
DDF2	SH3	Positive
DLG1	SH3	Positive

EndophilinB1	SH3	Positive
GRB2	SH3-1	Positive
HCK	SH3	Positive
INTERSECTIN1	SH3-1	Positive
INTERSECTIN1	SH3-4	Positive
JIP1	SH3	Negative
MPP6	SH3	Negative
NCK1	SH3-1	Negative
NCK1	SH3-2	Negative
NCK2	SH3-3	Positive
NPHP1	SH3	Positive
SNX9	SH3	Positive
P51NOX	SH3	Positive
PAC3	SH3	Positive
Vav	SH3-1	Negative
GRAP2	SH3-1	Positive

Table 1 SH3 domains probed by phage display experiments. In the third column, positive and negative indicate good and bad peptide chip results, respectively.

We used the GST-SH3 fusions as baits to select phages displaying binding peptides. After three selection cyles, we tested the binding ability of each enriched phage by an ELISA assay.

Clones that turned out to be positive in ELISA were sequenced and analyzed.

For seven of the tested SH3 domains (ArgBP2-2, Nck2-3, Dlg1, Vav-1, Grap2-1, Jip1, P51nox) I couldn't obtain any enrichment of phages displaying binding peptides. The SH3 domain from Mpp6 enriched phages in the third cycle of selection but no enriched phage was positive in ELISA assay. The SH3 domains from AhrGap12, ArgBp1, Intersectin1 (SH3-4) and EndophlinB1 selected peptide with no consensus.

Eight SH3 domains have selected specific phage clones with characteristic amino acid residues. In Figure 18 I represented the alignments of the peptide sequences exposed by the specifically selected phages.

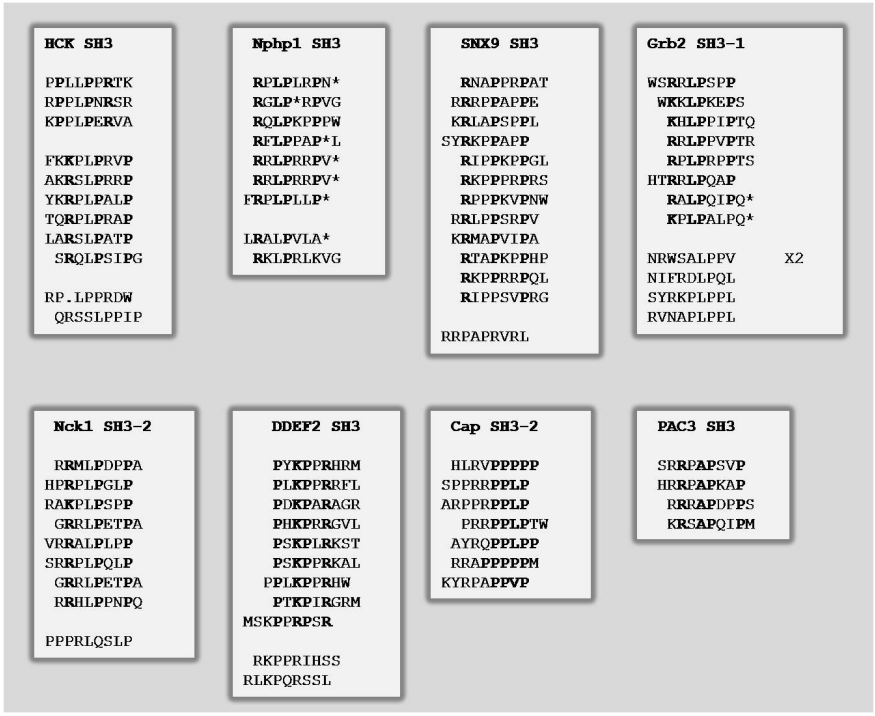


Figure 19. Sequence alignments of selected peptides in phage display experiments. The conserved residues are in bold, while the * asterisk indicate a stop codon.

We propose two consensus sequences recognized by the SH3 domain from Nphp1 (Nephrocystin1). The motif RxLPxpP is in agreement with the results obtained by peptide chips. This phage display selection highlighted a particular feature: most of the peptides selected by the NphpI-SH3 had a stop codon in their sequence (Figure 18, asterisk). This could suggest preference for peptides at the C-terminus.

The phage display selections using the SH3 domains of Hck and SNX9, confirm the peptide chip results.

The first SH3 domain from Intersectin1 protein selected only one single class II peptide PVVPPRGSR, in agreement with the chip experiment (not shown in the figure).

The peptides selected by the SH3 of DDEF2 match the Class II consensus sequence adding the preference for a lysine at position P2.

These phage display contributed to confirm and expand the binding

specificities obtained by peptide chip.

Spot synthesis experiments.

In order to further verify the data obtained by the new high density peptide chip approach, we decided to synthesize a selection of peptides, at low density, on cellulose membranes and challenge these with a selection of SH3 domains (Table 2). To this aim we synthesized peptides that were classified as positive interactors by peptide chip and we added, for each chosen SH3 domain, some ligands described in the literature. Some of them are peptides derived from phage display (Cestra et al., 1999) or spot synthesis screening (Wu et al., 2007).

Name	**Domain name**	**Peptide Chips Results**
ARHGAP12	SH3	Positive
CAP	SH3-2	Positive
COOL-1	SH3	Positive
DDF2	SH3	Positive
FYN	SH3	Positive
GRB2	SH3-1	Positive
HCK	SH3	Positive
NCK1	SH3-1	Negative
NCK1	SH3-2	Negative
NPHP1	SH3	Positive
PAC3	SH3	Positive
PLCg	SH3	Positive

Table 2 SH3 domains probed by Spot Synthesis experiments. In the third column, positive and negative indicate good and bad peptide chip results, respectively.

All twelve experiments gave interpretable results. It is worth to note that, under this experimental condition, also the Nck1 SH3 domains specifically recognized some peptides (Figure 19).

Wu and collaborators reported a screening with 12 SH3 domains from

eight human proteins (Src, PlcG1, P85A, Nck1, Grb2, Fyn, Crk) of a peptide array composed of 1536 potential ligands, which led to the identification of 921 binary interactions between these proteins and 284 targets. In our screening we re-synthesize some of their positive peptides to compare the two datasets. The probed membranes confirmed that the peptides that were positive in the high density chip experiments are strongly bound by SH3 domains also in this different experimental setup.

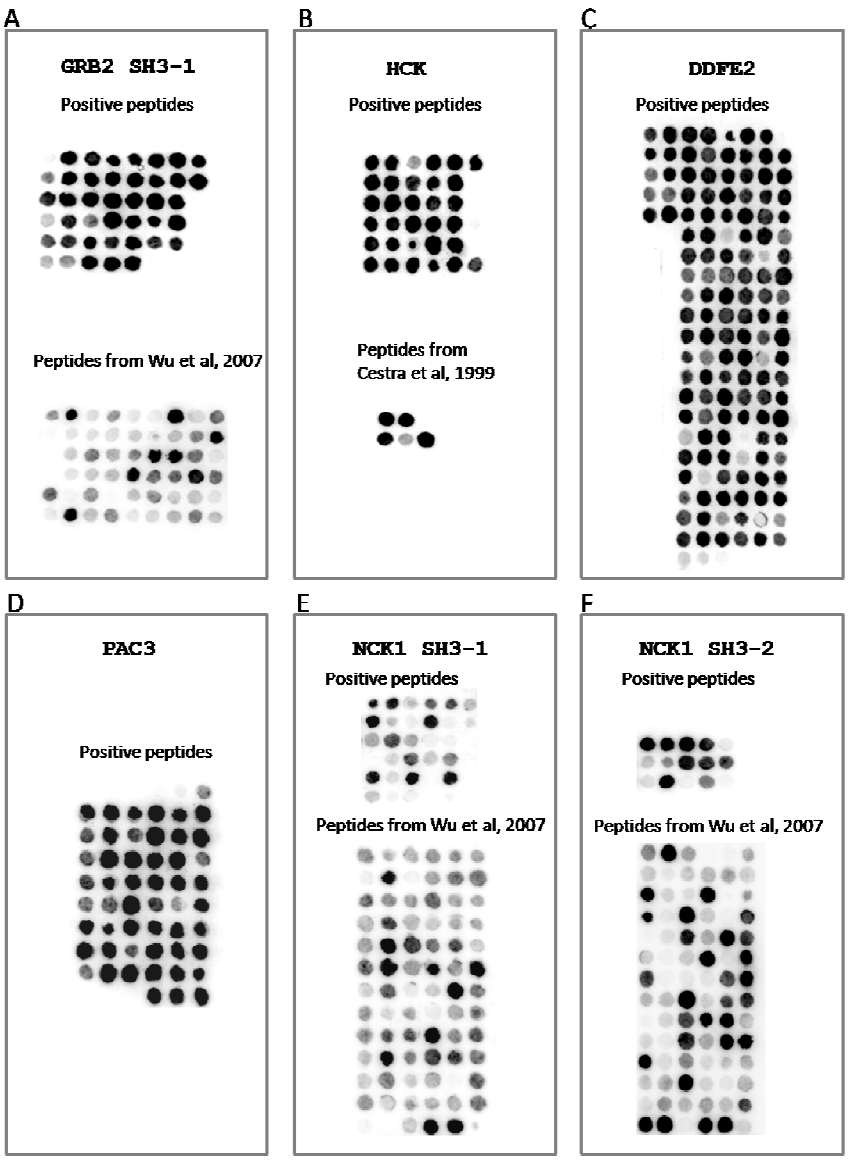


Figure 20. Spot synthesis experiments. The positive peptides selected in chip experiments by GRB2 SH3-1 (A), HCK SH3 (B), DDEF2 SH3 (C), PAC3 SH3 (D), NCK1 SH3-1 (E) and SH3-2 (F) domains were synthesized on cellulose membranes and probed with their respective GST-SH3 fusion proteins. Some peptides found in literature were also synthesized to compare our results with other technologies.

We estimate that more than 95% of the peptide ligand discovered by the peptide chip approach might be confirmed by experiments based on low density SPOT synthesis on cellulose membranes.

Identification of new potential binding partners of Cbl-b and Gab1 proteins.

The peptide chip approach we performed only addresses the ability of any given bait domain to bind a peptide outside its protein context, but we would like to assess how many of the 23000 inferred interactions could occur in the context of a natively folded protein. To this aim, we used pull down assays from cell extracts. The pull down approach cannot be used in a high throughput fashion, thus we set out to validate a small subset of the inferred interactions by focusing on the protein pairs involving two proteins, Gab1 and Cbl-b that were already under investigation in our group.

Cbl-b (Casitas B-lineage lymphoma proto-oncogene) is a ubiquitin E3 ligase involved in the ubiquitination and internalization of the EGF receptor as well as other tyrosine kinase receptors (Rao et al., 2002).
Cbl-b contains an N-terminal tyrosine kinase-binding (TKB) domain, which mediates binding to specific phosphotyrosine motifs; a Zn-coordinating RING finger domain, that allow the interaction with Ubiquitin-conjugating E2 enzymes; a C-terminal proline-rich region; and a C-terminal leucine zipper-like region (Lupher et al., 1999).

The C-terminal proline-rich region of consists of 449 amino acid residues, from residue 479 to 925, and contains 17 potential SH3 binding sites, including the PPVPPR sequence recognized by the Grb2 SH3 N-terminal domain, the poly-proline sequence of PPPPPP(E/D)R, and the atypical PX(P/A)XXR motif, bound by SH3 domains of the CIN85 (Cbl-interacting protein of 85 kDa)/RUK (regulator of ubiquitous kinase)/CD2AP (C2-associated protein) family of proteins that promotes ligand induced internalization of the EGF (Soubeyran et al., 2002) and Met (Petrelli et al., 2002) receptors by recruiting endophilin via CIN85.

Gab1 (Grb2 associated binder 1) is a scaffolding adaptor protein acting downstream to tyrosine kinase receptors, in response to various stimuli. Over the past few years, several reports demonstrated the involvement of Gab1 in

many signalling events. The phosphorilation on tyrosine of Gab1, in response to epidermal growth factor (EGF) and insulin stimulations (Holgado-Madruga et al., 1996) allows its binding to the Met receptor (Weidner et al., 1996). Furthermore, it participates in signalling initiated by interleukins (ILs), interferons (IFNs), erythropoietin (EPO) and thrombopoietin (TPO) (reviewed in (Liu and Rohrschneider, 2002).

The Gab1 protein structure consists of an amino-terminal PH domain, a 92 amino acid long proline rich region, from residue 449 to 540, region and several potential SH2 binding sites.

A canonical PxxPxR and an atypical Px(V/I)(D/N)RxxKP motifs serve as docking sites for the Grb2 SH3 domains (Holgado-Madruga et al., 1996; Schaeper et al., 2000). The atypical PxxxR region of Gab1 binds also to the SH3 domains of Grb2-like adapter proteins Grap2 (Gad family) (Lock et al., 2000).

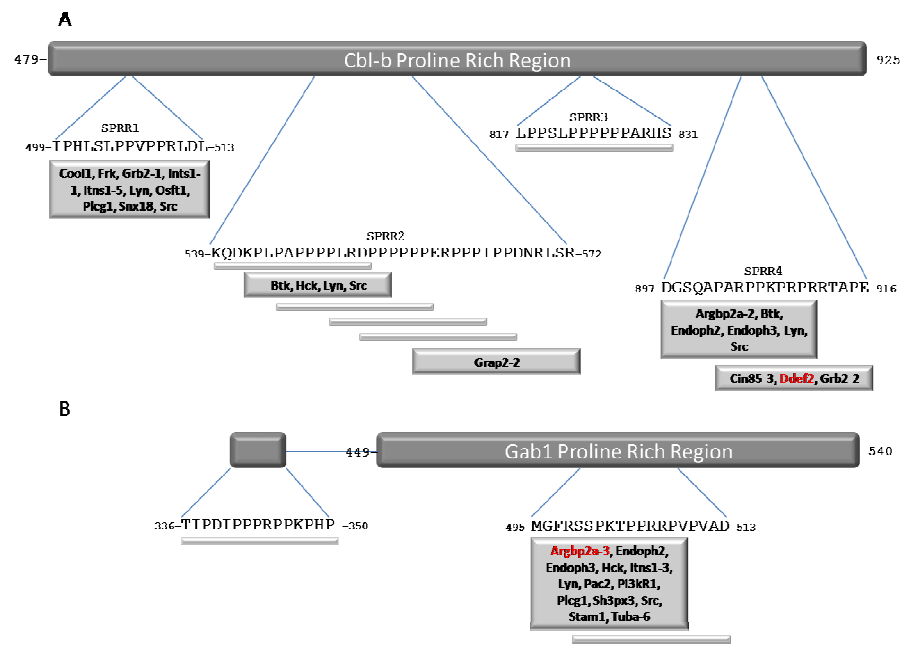


Figure 21. Schematic representation of the Cbl-b (A) and Gab1 (B) proline rich regions. Several peptides (grey bars) were tested in our peptide chip experiments. The box reports the SH3 domains bound in those particular regions. In red the predicted SH3 domains in peptide chips but not confirmed in pull down experiments.

The Gab1 and Cbl-b peptides present in the array we tested are shown in Figure 21.

Ten peptides of the proline rich region of Cbl-b were tested in our binding assay with our collection of 90 SH3 domains. We named Sub-Proline Rich Region 1 to 4 (SPRR1-4) the four Cbl-b regions that are potential SH3 targets (Figure 20A).

The SPRR1 contains peptide IPHLSLPPVPPRLDL, which showed a signal above threshold with the SH3 domains of Cool1, Frk, Grb2 (SH3-1), Intersectin1 (SH3-1 and SH3-5), Lyn, Osft, Plcg1, Snx18 and Src. SPRR2 on the other hand is represented on our chip by six partially overlapping peptides. Among these six peptides only peptide 2 and 6 showed a signal above threshold with any SH3 domain. Peptide 2 (KPLPAPPPPLRDPPP) was bound by members of the Src Tyrosine Kinase family such as Btk, Hck, Lyn and Src, while the sixth peptide (PPERPPPIPPDNRLS) bind to the second SH3 domain of the scaffolding adaptor protein Grap2.

Peptide LPPSLPPPPPPARHS, the SPRR3 proline rich peptide, did not bind to any SH3 domain.

The SPRR4 is represented on the chip as two staggered peptides that overlap in ten of their fifteen residues. DGSQA**P**AR**PP****K**PRPR binds to the SH3 domain of Btk, Lyn, Src, Endophilin2, Endophilin3 and the second SH3 of ArgBP2a. Interestingly the PARP**P**K**P**R**PR**RTAPE peptide has a completely different specificity and binds the third SH3 domain from CIN85, the second SH3 domain from Grb2 and the SH3 domain from DDEF2. The two putative motifs recognized by the two classes of SH3 domains are outlined in bold.

The peptide collection present in our microarray contains 3 Gab1 peptides (Fig. 21B). The first peptide (TIPDIPPPRPPKPHP), which maps 113 residues aminoterminal of the Gab1 proline rich region, as identified by UniProtKB, didn't recognize any SH3 domain in our experiments. By contrast the peptide MGFRSSPKTPPRRPV, in the proline rich region, proved to be very promiscuous and bound Argbp2a (SH3-3), Endophilin2, Endophilin3, Hck, Intersectin1 (SH3-3), Lyn, Pac2, Pi3kR1, Plcg1, Sh3px3, Src, Stam1, Tuba (SH3-6). This peptide contains both a class 1 RSSPKTP and a class 2 motif PKTPPR. However, removal of the four amino-terminal residues MGFR, which disrupts motif 1, as in peptide SSPKTPPRRPVPVAD, also disrupts binding, suggesting to a prominent role of the Class I motif in binding this collection of SH3 domain.

The SH3 domains were purified from E. coli as GST fusion and immobilized on glutathione Sepharose beads. 293T or 293 Phoenix cell lines and mouse brain were used to pull down Cbl-b and Gab1 proteins in protein

extracts.

All the interactions described above were confirmed by pull down assays with the exception of the interaction between ArgBp2a (SH3-3) and Gab1 and between SH3 from Ddef2 and Cbl-b (Figure 22).

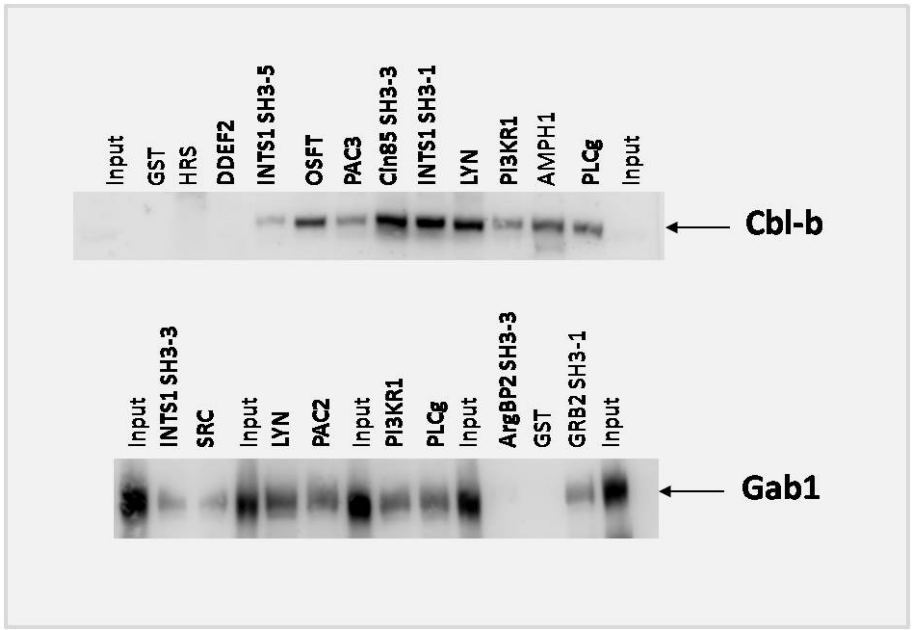


Figure 22. Pull down experiments of Cbl-b (upper panel) and Gab1 (lower panel). The predicted SH3 domains were used as GST fusions to pull down the Cbl- and Gab1 proteins. Only 2 SH3 domains (DDEF2 and ArgBP2a SH3-3) weren't able to precipitate the proteins of interest as predicted in chip experiments. GST and HRS lanes are negative controls (HRS correspond to the VHS domain region of HRS protein).

Some interactions, such as the one between Cbl-b and Cin85 have been described previously (Kurakin et al., 2003).

The validation of such a high proportion of interaction inferences based on array screening, demonstrates the potential of our peptide-based array approach for the discovery of new interactions mediated by SH3 domains in the human proteome.

The functions of the different binding proteins are suggestive of novel functional tasks mediated by the interacting protein. For instance the proteins whose SH3 domains reacted with the Cbl-b peptide DGSQAPARPPKPRPR

are either involved in growth factor signaling like the Src kinase family members (Src, Lyn, Btk), in diverse signaling (ArgBP2a) or in endocytosis and vesicular trafficking (Endophilin, CIN85) or in both (Grb2, Intersectin1, Stam1, Tuba, etc.).

Additional pull down assays were performed to validate interactions mediated by proline rich peptides from Itch, Abl1, Pob1 and Amphyphisyn1 and their predicted SH3 domains (Figure 23). Taken together the pull down assays confirmed 96% of the inferred interactions.

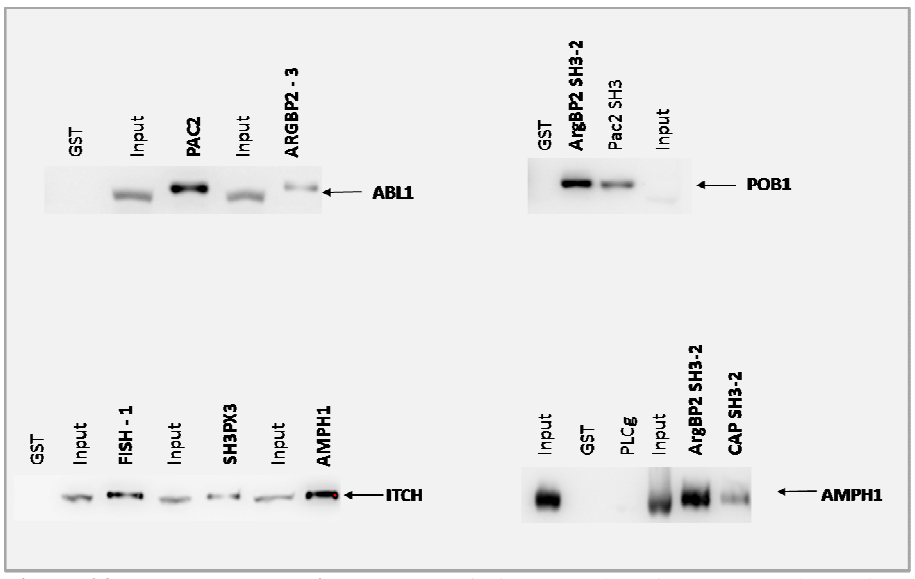


Figure 23. Pull down experiments. GST fusion SH3 domains were used as bait to pull down Abl1 (A), Pob1 (B), Itch (C) and Amphyphisyn1 (D) proteins. The GST lane is the negative control.

Identification of potential targets of SH3 domains in viral proteins.

The WISE approach consists in the definition of a number of regular expressions covering all the motifs recognized by the members of a given domain family (obtained, for instance, by phage display experiments or from the literature) followed by an informatic scanning of the whole proteome in search of peptides matching the regular expressions. Finally, the putative

ligands are experimentally validated by probing peptides synthesized by spatially addressed synthesis on a cellulose membrane.

In particular, we used a list of 21 regular expressions covering most of the SH3 recognition consensi described till now. This list was compiled by combining the results of published and unpublished experiments mostly based on phage display (Table 3).

Relaxed Cosensus	**Corresponding Classes or Proteins**
4 consecutive proline in 12 residues	
xxx(R/K)x(A/V/L/I/M/R/T/H/P)Px(V/L/A/I/P/R/T)Pxxxxx	Class1
xxxxx(P/V/S/L/R/K/H/E/A/Q/T/M)Px(I/L/M/V/P/Y/A/F/T/R)P(A/P/V/L/R/S/K/H/F/N/Q/D/I/M/T)(R/K)xxx	Class2
xxx(K/R)(K/R)(A/H/L/P/K/R/G/R)PPPxxxxxx	Yhr114
xxx(K/R)(K/R)PPPxxxxxxx	Yhr114
xxx(K/R)(K/R)PP(V/L/P)P	Yhr114
xx(K/R)xxxPxxPxKPxx	Abp1
xx(K/R)xxxPPxxxKPxx	Abp1
x(L/F/P)x(Y/L/W/F/P)x(A/L/M/V/F/P)P(A/L/M/V/F/P)(L/S/A/P/I/V/R)Pxxxxx	Myo3 Abl
xxxxPxR(L/A/P)APxxxxx	Nbp2
xxxxRP(A/S)xxxxY(T/R/I)xx	Yhl002
xxxxRP(A/S)xxx(K/T/H/N)Yxxx	Yhl002
xxxxxPPx(I/V)xP(F/Y)xxx	Bem1
xxxxxPxR(N/S)P(R/A/M/G/N/S)R(I/L/V/A/T/P)xx	Boi
xxxxxxPR(N/S/D/H)P(R/A/M/G/N/S)R(I/L/V/A/T/P)xx	Boi
xxx(Y/F/R/K)xR(F/L/Y/E/P)x(A/V/L/Y/I/M/F/H/W/R/T/P)P(I/P/L)Pxxx	Ypr154
xxx(E/D)P(L/S/T/V/K/R/L/I/M)(A/P)(P/V/I/F/S/A/R/K/M)PR(F/P/T/R/I/V/K/A/S/L)xxxx	Cin85
xxxx(I/L/V/A/W/F/Y/R/P)P(P/T/A/D/L/R/I/N/A/H/G/Q)(P/M/A/R/I/V)DYxxxxx	Eps8
xxxxx(L/S/A/P/I/V)P(K/R)RPPxxxxx	Endophilin
xxxx(P/I/V/R/A/L)(P/L/A)(P/E/P/K/A/L/H)RxxKPxxx	Stam2/GAD
xxx(P/I/V/R/A/L)x(V/I)(D/N)RxxKPxxx	Stam2/GAD

Table 3 Regular expressions mapping the viral proline-rich peptides used in the

screening. The "x" correspond to any amino acid residues.

We next searched matching peptides in the HBV, SV40, Adenovirus and HPV viral proteomes. Only adenovirus A type 12 and papilloma type 16 viral proteomes were found to contain matching proline-rich peptides (Table 4).

VIRUS	PROTEIN	UNIPROT	SEQUENCE	RANGE
Human adenovirus A type 12	E1A; transcription activation/early	P03259	FLPEPPVLSPVCEPI	97-111
	E1B protein, small T-antigen	P04492	ARPLTMPPLPTLQEE	137-151
	maturation protein (protein IVA2)	P12540	EGQNPRPFGCSPPRS	49-63
	DNA polymerase	P06538	ALTHPMPWGPPLNPY	557-571
			PMPWGPPLNPYERAL	561-575
			EDQGPAPFYSPPPEN	781-795
			PAPFYSPPPENCEHV	785-799
			FRQILYRAEKPQRTH	494-508
	DNA terminal protein	P12541	GWPLYREPPPHFLVG	37-51
			RRRRRQPPPPVVEEE	325-339
			RQPPPPVVEEEIMEV	329-343
			PERQADIPLPPLPAG	577-591
			ADIPLPPLPAGPEPP	581-595
			LPPLPAGPEPPLPPG	585-599
			PAGPEPPLPPGARPR	589-603
	Peripentonal hexon-associated protein	P36712	ILNPHWLPPPGFYTG	393-407
			WEERQPRPVRPPRQR	521-535
	penton protein	P36716	TVAFPETPPPSYETV	9-23
			ETVMAAAPPYVPPRY	21-35
			AAAPPYVPPRYLGPT	25-39

			VMAAAPPYVPPRYLG	23-37
	protein VI	P35988	ENVPMTLELPPLPEP	189-203
			MTLELPPLPEPTIAD	193-207
			LPPLPEPTIADPVGS	197-211
	early E2A DNA-binding protein	P04498	RERTPDRSAQPPPPK	9-23
			PDRSAQPPPPKMGRY	13-27
			AQPPPPKMGRYFLDS	17-31
			EELEAVPLPPKKKVK	35-49
	late 100 KD protein	P36714	DFVPIYFRECPPPLW	505-519
			ERNLFSPRVPPKRQA	113-127
	hexon associated protein	P36713	QENPPPTTVLLPRDA	81-95
	E3B 14.5 KD protein	P36707	DLPEYPNPQDDLPLN	73-87
			PLNIVFPEPPRPPSV	85-99
			VFPEPPRPPSVVSYF	89-103
	fiber protein	P36711	YDPLTLWTTPDPPPN	405-419
			TLWTTPDPPPNCSLI	409-423
	E4 13 KD protein	P36709	MPLPCIPPPPVSRDT	1-15
			CIPPPPVSRDTAACI	5-19
Human papillomavirus type 16	E4 ORF from 3332 to 3619; putative	P06922	STWPTTPPRPIPKPS	21-35
			TTPPRPIPKPSPWAP	25-39
			RPIPKPSPWAPKKHR	29-43
	minor capsid protein	P03107	TLAPVRPPLTVDPVG	85-99
			VRPPLTVDPVGPSDP	89-103
			AGAPTSVPSIPPDVS	117-131
	major capsid protein	P03101	ACQKHTPPAPKEDDP	452-466
			LLAVGHPYFPIKKPN	68-82

Table 4 Viral proline-rich peptides. 46 viral peptides synthesized on cellulose membrane, belonging to Human Adenovirus A type 12 and Human Papilloma type 16 proteome (column 1), viral protein name (column 2), UniProtKB accession number (column 3), peptide sequences (column 4) and peptide ranges (column 5) are shown.

In human adenovirus type 12 proteome, 14 proteins have proline rich regions: the E1A, the E1B, the maturation proteins and the hexon-associated protein (one peptide each); the Peripentonal hexon-associated protein, the late 100 KD protein, the fiber protein and the E3 13kDa protein (two peptides each); the E3B 14.5 KD protein, the protein VI (three peptides each); the penton protein and the early E2A DNA-binding protein (four peptides each); the DNA polymerase (five partially overlapping peptides); the DNA terminal protein (seven partially overlapping peptides). The proteome of Papilloma type 16 contains only 3 proteins with proline-rich regions: the major capsid protein (two peptides); the E4 (three partially overlapping peptides) and the minor capside protein.

These 46 viral proline-rich peptides were synthesized on cellulose membranes, by the SPOT synthesis technique, and probed with GST fusions of 15 human SH3 domains, representative of discrete specificity classes (Table 5). The proteins containging those domains showed very different molecular functions: ArhGAP12 is a Rho-GTPase activator; ArgBP2a is an adapter protein that plays a role in the assembling of signaling complexes, being a link between Abl kinases and actin cytoskeleton (Soubeyran et al., 2003; Yuan et al., 2005); CAP, plays a role in tyrosine phosphorylation of Cbl by linking Cbl to the insulin receptor (Ribon et al., 1998); Ddef2 is a small GTPase activator (Andreev et al., 1999); Grb2 is an adapter protein that provides a critical link between cell surface growth factor receptors and the Ras signaling pathway (Lowenstein et al., 1992); Hck is a protein tyrosine kinase that may interact (via SH3 domain) with HIV-1 Nef and Vif (Briggs et al., 1997; Hassaine et al., 2001); Nck1 is an adapter protein which associates with tyrosine-phosphorylated growth factor receptors or their cellular substrates (Latreille and Larose, 2006); Nephrocystin-1 (Nphp1) through its SH3 domain binds to Cas and together may play a role in the control of epithelial cell polarity (Donaldson et al., 2000); Pacsin3 (Pacn3) bind to endocytic proteins and inhibit endocytosis (Modregger et al., 2000); PLCg1 is a phospholipase involved in host-virus interactions (Korkaya et al., 2001); Rimb1 is a binding partners of the presynaptic active zone proteins RIMs and of voltage-gated Ca2+-channels (Mittelstaedt and Schoch, 2007). Table 5 lists the SH3 domains used in the experiment on the corresponding amino acid ranges as well as their multiple sequence alignment. Some of these proteins contain multiple SH3 domains.

PROTEIN	UNIPROT	RANGE	SH3
ARGBP2A	O94875	1041-1100	3/3
ARHGAP12	Q8IWW6	12-74	1/1
CAP	Q9BX66	867-928	2/3
COOL1	Q14155	184-243	1/1
DDEF2	O43150	944-1006	1/1
FYN	P06241	82-143	1/1
GRB2	P62993	1-58	1/2
HCK	P08631	78-138	1/1
IRSP53	Q9UQB8	374-437	1/1
NCK1	P16333	1-61	1/3
NCK1	P16333	115-165	2/3
NPHP1	O15259	152-212	1/1
PAC3	Q9UKS6	363-424	1/1
PLCG1	P19174	791-851	1/1
RIMB1	O95153	1764-1831	3/3

Table 5 List of 15 SH3 domains used in the screening. First column contains protein name of the 14 SH3 containing proteins; the second one their Uniprot accession number; the third one the amino acid range of each SH3 domain and the last one the numbering N to C-term of SH3 multiple SH3 domains of those proteins, e. g. the Nck1 protein contains 3 SH3 domains and in this work we probed the binding properties of the first two domains, referred in the table as 1/3 and 2/3.

Identification of viral binding partners of SH3 domains.

Identification of the domain binding peptides was obtained by an overlay assays. After binding of the GST-SH3 fusions, the membranes were treated with a goat anti-GST antibody and finally with a peroxidase labeled anti-goat antibody. The chemiluminescent signal intensity of each spot was quantified. We considered as positive interactions the spots (peptides) with signal intensity higher than the signal average of each membrane plus one standard deviation. By this approach we identified 114 new putative

interactions between the 15 SH3 domains and the proline rich regions of the two viral proteomes (Figure 24 and Table 6).

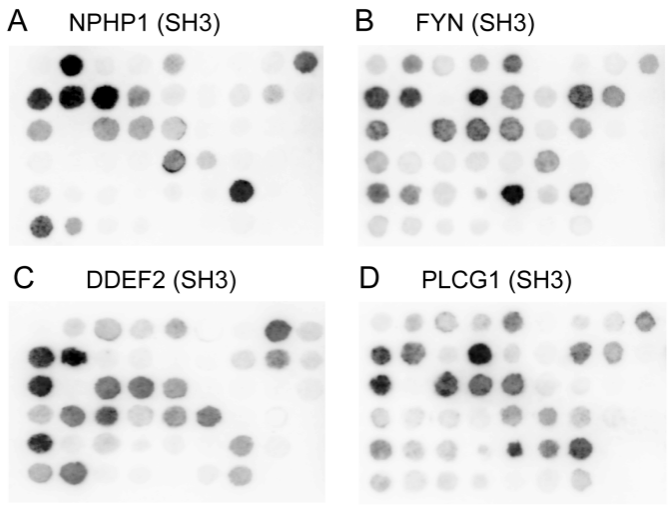


Figure 24. Identification of SH3 viral binding partner by the Spot method. The viral peptides were synthesized on cellulose membranes and then probed with different GST-SH3 fusion proteins. The binding was revealed by anti-GST antibodies as described; the binding intensity was evaluated in arbitrary chemioluminescence units. Each array (9 x 6 spots) contains 46 viral Pro-rich peptides (Table 4), a peptide (LAKDLIVPRRPEWNA) recognized by anti-GST antibody (column 9 row 1) and four peptides of the viral Adenovirus A type 12 DNA Poliymerase protein, lacking any proline residue and used as negative controls. Their sequences are: QDFKYQYLK (column 8 row 5), DKEEYLLNQ (column 8 row 6), SHELYDYVR (column 9 lane 2), ELYDYVRAS (column 9 lane 3). Four empty spots (column 9 row 3, 4, 5, 6) were included for evaluation of the signal background intensity. The membranes were probed with the SH3 domains of the NPHP1 (A) FYN (B) DDEF2 (C) and PLCG1 (D) proteins respectively.

Human Protein	Viral Protein	Viral PRS
ARHGAP12 (1/1)	DNA polymerase	PMPWGPPLNPYERAL
ARHGAP12 (1/1)	Early E3B 12.7 kDa	PLNIVFPEPPRPPSV
ARHGAP12 (1/1)	Late 100 kDa	DFVPIYFRECPPPLW

ARHGAP12 (1/1)	DNA terminal	GWPLYREPPPHFLVG
ARHGAP12 (1/1)	DNA terminal	PAGPEPPLPPGARPR
ARHGAP12 (1/1)	early E4 13 kDa	MPLPCIPPPPVSRDT
ARHGAP12 (1/1)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
ARHGAP12 (1/1)	Fiber	YDPLTLWTTPDPPPN
ARHGAP12 (1/1)	DNA terminal	PERQADIPLPPLPAG
CAP (2/3)	early E4 13 kDa	MPLPCIPPPPVSRDT
CAP (2/3)	DNA terminal	GWPLYREPPPHFLVG
CAP (2/3)	Early E3B 12.7 kDa	VFPEPPRPPSVVSYF
CAP (2/3)	DNA polymerase	PMPWGPPLNPYERAL
CAP (2/3)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
CAP (2/3)	Early E3B 12.7 kDa	PLNIVFPEPPRPPSV
CAP (2/3)	DNA terminal	PERQADIPLPPLPAG
CAP (2/3)	E4	STWPTTPPRPIPKPS
CAP (2/3)	Peripentonal hexon-associated	ILNPHWLPPPGFYTG
COOL1 (1/1)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
COOL1 (1/1)	Penton	VMAAAPPYVPPRYLG
COOL1 (1/1)	Penton	AAAPPYVPPRYLGPT
COOL1 (1/1)	DNA terminal	RRRRRQPPPPVVEEE
COOL1 (1/1)	E4	STWPTTPPRPIPKPS
COOL1 (1/1)	Peripentonal hexon-associated	ILNPHWLPPPGFYTG
COOL1 (1/1)	DNA terminal	PAGPEPPLPPGARPR
COOL1 (1/1)	Penton	ETVMAAAPPYVPPRY
COOL1 (1/1)	Late 100 kDa	ERNLFSPRVPPKRQA
FYN (1/1)	early E4 13 kDa	MPLPCIPPPPVSRDT
FYN (1/1)	DNA terminal	GWPLYREPPPHFLVG
FYN (1/1)	DNA terminal	PAGPEPPLPPGARPR
FYN (1/1)	DNA terminal	PERQADIPLPPLPAG
FYN (1/1)	Penton	AAAPPYVPPRYLGPT
FYN (1/1)	DNA terminal	RRRRRQPPPPVVEEE
FYN (1/1)	Penton	ETVMAAAPPYVPPRY
FYN (1/1)	Penton	VMAAAPPYVPPRYLG
DDEF2 (1/1)	DNA terminal	RRRRRQPPPPVVEEE
DDEF2 (1/1)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
DDEF2 (1/1)	DNA terminal	GWPLYREPPPHFLVG
DDEF2 (1/1)	Early E3B 12.7 kDa	PLNIVFPEPPRPPSV
DDEF2 (1/1)	DNA polymerase	FRQILYRAEKPQRTH
DDEF2 (1/1)	Early E2A DNA-binding	AQPPPPKMGRYFLDS
DDEF2 (1/1)	Penton	AAAPPYVPPRYLGPT
DDEF2 (1/1)	Penton	ETVMAAAPPYVPPRY
DDEF2 (1/1)	E4	RPIPKPSPWAPKKHR
GRB2 (1/2)	DNA terminal	PERQADIPLPPLPAG
GRB2 (1/2)	DNA terminal	RRRRRQPPPPVVEEE

GRB2 (1/2)	Penton	AAAPPYVPPRYLGPT
GRB2 (1/2)	Penton	ETVMAAAPPYVPPRY
GRB2 (1/2)	DNA terminal	GWPLYREPPPHFLVG
GRB2 (1/2)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
GRB2 (1/2)	Penton	VMAAAPPYVPPRYLG
HCK (1/1)	DNA terminal	ADIPLPPLPAGPEPP
HCK (1/1)	DNA terminal	PERQADIPLPPLPAG
HCK (1/1)	Minor capsid 6	MTLELPPLPEPTIAD
HCK (1/1)	Early E1A 29.5 kDa	FLPEPPVLSPVCEPI
HCK (1/1)	DNA terminal	PAGPEPPLPPGARPR
HCK (1/1)	Minor capsid 6	ENVPMTLELPPLPEP
HCK (1/1)	early E4 13 kDa	MPLPCIPPPPVSRDT
HCK (1/1)	DNA terminal	LPPLPAGPEPPLPPG
NCK1 (1/3)	DNA terminal	RRRRRQPPPPVVEEE
NCK1 (1/3)	DNA terminal	GWPLYREPPPHFLVG
NCK1 (1/3)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
NCK1 (1/3)	DNA polymerase	FRQILYRAEKPQRTH
NCK1 (1/3)	Late 100 kDa	ERNLFSPRVPPKRQA
NCK1 (1/3)	Peripentonal hexon-associated	ILNPHWLPPPGFYTG
NCK1 (1/3)	Early E2A DNA-binding	RERTPDRSAQPPPPK
NCK1 (1/3)	Late 100 kDa	DFVPIYFRECPPPLW
NCK1 (2/3)	DNA terminal	PERQADIPLPPLPAG
NCK1 (2/3)	DNA terminal	RRRRRQPPPPVVEEE
NCK1 (2/3)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
NCK1 (2/3)	Penton	ETVMAAAPPYVPPRY
NCK1 (2/3)	Penton	AAAPPYVPPRYLGPT
NCK1 (2/3)	Penton	VMAAAPPYVPPRYLG
NCK1 (2/3)	Late 100 kDa	ERNLFSPRVPPKRQA
NPHP1 (1/1)	DNA terminal	RQPPPPVVEEEIMEV
NPHP1 (1/1)	E1B, small T-antigen	ARPLTMPPLPTLQEE
NPHP1 (1/1)	DNA terminal	RRRRRQPPPPVVEEE
NPHP1 (1/1)	E4	STWPTTPPRPIPKPS
NPHP1 (1/1)	DNA terminal	GWPLYREPPPHFLVG
NPHP1 (1/1)	E4	TTPPRPIPKPSPWAP
PAC3 (1/1)	DNA terminal	RRRRRQPPPPVVEEE
PAC3 (1/1)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
PAC3 (1/1)	Penton	ETVMAAAPPYVPPRY
PAC3 (1/1)	Penton	AAAPPYVPPRYLGPT
PAC3 (1/1)	DNA terminal	GWPLYREPPPHFLVG
PAC3 (1/1)	Penton	VMAAAPPYVPPRYLG
PAC3 (1/1)	DNA terminal	PAGPEPPLPPGARPR
PLCG1 (1/1)	DNA terminal	PERQADIPLPPLPAG
PLCG1 (1/1)	Peripentonal hexon-associated	WEERQPRPVRPPRQR

PLCG1 (1/1)	Penton	ETVMAAAPPYVPPRY
PLCG1 (1/1)	E4	STWPTTPPRPIPKPS
PLCG1 (1/1)	Penton	AAAPPYVPPRYLGPT
PLCG1 (1/1)	DNA terminal	GWPLYREPPPHFLVG
PLCG1 (1/1)	Penton	VMAAAPPYVPPRYLG
PLCG1 (1/1)	DNA terminal	RRRRRQPPPPVVEEE
ARGBP2A (3/3)	Penton	ETVMAAAPPYVPPRY
ARGBP2A (3/3)	Penton	VMAAAPPYVPPRYLG
ARGBP2A (3/3)	DNA terminal	GWPLYREPPPHFLVG
ARGBP2A (3/3)	Peripentonal hexon-associated	WEERQPRPVRPPRQR
ARGBP2A (3/3)	DNA polymerase	PMPWGPPLNPYERAL
ARGBP2A (3/3)	DNA terminal	RRRRRQPPPPVVEEE
IRSP53 (1/1)	DNA polymerase	PMPWGPPLNPYERAL
IRSP53 (1/1)	DNA terminal	GWPLYREPPPHFLVG
IRSP53 (1/1)	Late 100 kDa	DFVPIYFRECPPPLW
IRSP53 (1/1)	Early E3B 12.7 kDa	VFPEPPRPPSVVSYF
IRSP53 (1/1)	DNA polymerase	ALTHPMPWGPPLNPY
IRSP53 (1/1)	Penton	ETVMAAAPPYVPPRY
IRSP53 (1/1)	Early E3B 12.7 kDa	PLNIVFPEPPRPPSV
IRSP53 (1/1)	Peripentonal hexon-associated	ILNPHWLPPPGFYTG
RIMB1 (3/3)	DNA terminal	RRRRRQPPPPVVEEE
RIMB1 (3/3)	DNA terminal	GWPLYREPPPHFLVG
RIMB1 (3/3)	DNA polymerase	PMPWGPPLNPYERAL
RIMB1 (3/3)	Late 100 kDa	DFVPIYFRECPPPLW

Table 6 Host-virus domain-peptide interactions. The 213 physical interaction found in this screening are listed. Column 1 and 2 show the name of Human and Viral proteins. The third column contains viral proline-rich sequence peptides.

The SH3 domains bind with high specificity to few proline-rich sequences while others recognize several proline-rich sequences, displaying a lower selectivity. Figure 25 shows the VirusMINT protein network including the interactions described here. Only the SH3 of Hck and Fyn were already described in published papers as ligands of viral proteins (Hiyoshi et al., 2008; Shelton and Harris, 2008; Trible et al., 2006).

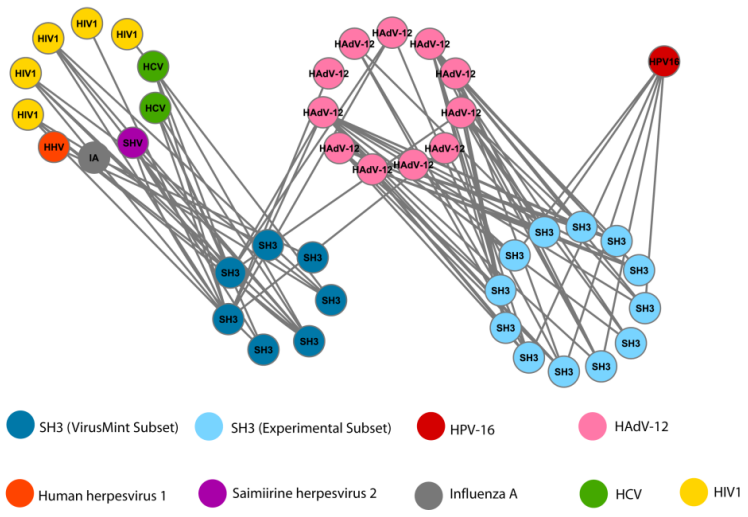


Figure 25. Enrichment of SH3 mediated interactions in the VirusMINT Database obtained by SPOT Synthesis experiments. The virus strains and proteins containing SH3 domains present in the network are indicated (upper panel). The blue nodes are the SH3 domain containing proteins that were already present in the VirusMINT database, while the light blue represents the SH3 used as probes in our screening. Two of the SH3 containing proteins (Hck and Fyn) are shared by the two subsets. Integrated virus-host interactions in MINT database (lower panel). The network has been visualized by the cytoscape software (Shannon et al., 2003).

Discussion

The SH3 domain is a small module involved in many signal transduction events that regulate important cellular processes. Often they have a key role in the regulation of those events and take part in formation of multi protein complexes.

Many strategies have been used in order to describe the binding partners and the recognition specificity of this domain's family.

During the past 20 years phage display experiments were carried out to indentify the consensus sequences recognized by several different protein domains. More recently, other approaches have been proposed to understand the intricate protein interaction network mediated by short peptide and domain: among these the WISE approach (Landgraf et al., 2004).

Our work represents the largest screening of naturally occurring human peptides for SH3 binders. The peptide chip screening reported herein was highly reproducible due to the optimization of the spotting process and to the ability of printing several copies of pre-made slides. The uniformity of the chip used in the screens also ensured that the generated data for different SH3 domains were comparable.

Similarly to other high throughput technologies, our approach might be affected by a high percentage of false positive and false negative results. Nevertheless, we are able to validate many predicted interactions by independent methods, like phage display assays, SPOT synthesis experiments and pull down assays.

We have been able to identify new recognition specificities for previously uncharacterized SH3 domains, such as Dddef2 or Nphp1, and to widen our understanding of the recognition mechanisms of well-known SH3domains, such as Fyn. Most importantly we described novel putative protein-protein interactions that could become help in formulating new hypothesis.

In a parallel approach, we collected all the information about peptide-SH3 domain interaction present in the literature, in order to verify and complete our predicted protein interaction network. To this aim we developed a simple text mining tool able to identify all the papers present in Pubmed reporting an interaction between a proline-rich peptide and an SH3 domain. All information was collected in a structured format in the MINT (http://mint.bio.uniroma2.it/mint/) the Molecular INTeraction database, (Chatr-aryamontri et al., 2007).

The future development of our database is an integrated resource aimed

to describe all the available information about the SH3 domains: interactions, structure, tissue expression and cellular localization.

In this work I demonstrated the feasibility of combining the SPOT synthesis technology with the WISE approach in order to the identify protein targets in viral proteomes, by means of selected peptide binding domains (Carducci et al., 2009). This strategy may facilitate the identification of critical nodes in the human proteome that are targeted by viruses. I refrain here from functional speculations because the *in vitro* data we produced need extension and further processing and validation in order to identify interactions that might be physiologically relevant. We probed only less than 10% of the known human SH3 domains but we can easily extend the information presented here to the whole SH3 family, by utilizing our classification of the recognition specificity of the entire SH3 domain family (Carducci et al in preparation). In addition, our data address only the potential of a given SH3 to bind a peptide on a cellulose membrane. There is no guarantee that these partners will ever meet in an infected cell. Gene expression as well as protein levels and intracellular localization, in different cell types are likely to confer a physiological dimension to these in vitro data.

As an example whose biological significance would be worth investigating is the inferred interaction between the Cool1 SH3 domain and some viral proteins. Cool1, a Rac/Rho GEF is found in a multiprotein complex with the p21 dependent kinase Pak, that has already been described as a HIV Nef target (Renkema et al., 2001). This mechanism could offer an alternative viral strategy to disturb the Cool1-Pak protein complex that is involved in focal adhesion assembly and in the control of growth and apoptosis.

It has been reported that viral proteins have a tendency to interact with hubs, that is proteins that display a high connectivity in the protein interaction network, and bottlenecks, proteins that are key connectors of many pathways in the network (Yu et al., 2007).

For these reason, we evaluated the enrichment, if any, of human hub proteins in the VirusMINT database. We considered as hubs the proteins whose connectivity in the human MINT network exceeds the connectivity of 95% of the nodes. By this definition 310 of the 5736 proteins in MINT are hubs. 72 of these are also present in VirusMINT, that is they are partners of viral proteins. Since the VirusMINT interactome contains a total of 508 proteins, this means that the hubs are 14% of the total human proteins in VirusMINT corresponding to a threefold enrichment.

The MINT database, for each interaction, provides a score, ranging from

0 to 1, and representing the experimental evidence supporting the interaction. By establishing a threshold of 0.3 one can obtain a smaller but higher confidence interactome. A hub enrichment analysis carried out on such a network yields similar results confirming that viral proteins have a preference to target human hubs.

Interestingly, most of the SH3 proteins interacting with virus proteins detected, like GRB2, Fyn and Nck1, Hck and ARHG7 are indeed highly connected cell proteins, confirming the crucial role of these proteins in the cell regulatory circuits and, as a consequence, in the host-virus relationship (Carducci et al., 2009).

Experimental Procedures

Chip Design

Chip slides (JPT) contained 9192 proline-rich peptides encoded in the human proteome.

Each chip contains two identical replicated arrays and each sub-array contains, in addition to the poly-proline peptides, several control spots arranged in a grid of 9600 dots. The control spots include Phosphate-Buffered Saline (PBS, as negative control), IgG or IgM antibodies (to check the binding of the secondary antibody), glutathione S-transferase protein (used as positive control for the primary antibody when a GST fusion protein is used as a probe) and the triple FLAG-epitope peptide (used as negative control). Finally, Tetramethyl-5-Carboxyrhodamine (TAMRA) and Cyanine 3 (Cy3) dyes are spotted to facilitate grid orientation.

Recombinant proteins production and overlay protocol

Escherichia coli BL21 Rosetta cells were transformed with pGEX-4TK recombinant plasmids directing the expression of specific GST-SH3 proteins. A single colony was grown in 25 ml of LB with ampicillin and chloramphenicol at 37 °C until an OD of 0.6 was reached. Protein production was performed at 30 °C for 5 hours after the addition of IPTG at a final concentration of 0,5 mM. The bacteria were pelleted and resuspended in lysis buffer (50 mM Tris, pH 8.0, 5 mM EDTA, 0.1% Triton X-100, 150 mM NaCl) plus a proteinase inhibitor mixture (Roche Applied Science) and lysed by a treatment with lysozyme (200 µg/ml final) for 1 hour on ice followed by three rounds of sonication. The lysate supernatant was incubated 1 hour with 100 µl of a 50% solution of glutathione-Sepharose beads (Amersham Biosciences) at 4 °C. Finally, beads were extensively washed in PBS, and the SH3 domains were eluted in 50 mM Tris, pH 8.0, with 10 mM glutathione. The amount of proteins produced was determined with a Bio-Rad protein assay. Proteins were dyalized in PBS over night.

The chips were blocked in PBS BSA 5% for 3 hours, then 20 µg/ml the GST fusion protein was added and the chips were incubated overnight. Chip

was washed in PBS and incubated 1:1000 with anti-GST Cy-5 conjugated antibody (Amersham-Pharmacia) avoiding light exposure. The chip was extensively washed in PBS and fluorescence intensity was revealed with ScanArray Gx Plus (Perkin Elmer).

SPOT signal evaluation

To identify the spots, the scanner software, prior to quantification, superimposes to the scanned image a grid where each spot occupies a circular area of fixed size and whose x-y coordinates have been previously defined. The average (mean) intensity of the pixels laying within the area delimited by the circle gives the foreground signal of the spot; the average (mean) intensity of the pixels falling outside of the circle, but still within a small distance from the circle's borderline, gives the background signal of the spot. The net signal intensity of a spot is computed as a function of the foreground and background signal, either by 1) subtracting background from foreground intensity (FG – BG); 2) taking the logarithm of the ratio of foreground to background intensity, which is equivalent to subtracting the intensity values in logarithmic rather than linear scale (log(FG/BG)). We considered both approaches and, in the end, opted for the latter, because it is more commonly employed in microarray analysis.

The chip readout may be affected by spatial biases and random noise. To cope with these sources of error, careful post-processing of the data was achieved.

The post-processing workflow involves several steps:

Assignment of 'high foreground' and 'low background' flags

In a few cases, due to a wrong positioning of the grid delimiting spot areas, the machine reads an unduly high background intensity value: in such situations, there is a risk of missing good binding candidates, namely spots with considerably high foreground intensity that would also have high FG-BG or log(FG/BG) values, if a mistake in background estimation had not occurred. To detect likely instances of this problem, we introduced two flags, 'high foreground' (fg_flag) and 'low background' (bg_flag): spots whose foreground intensity value is greater than two times the median foreground intensity of the experiment have their fg_flag set to 'GOOD'; conversely, spots whose background intensity value is greater than two times

the median background intensity of the experiment have their bg_flag set to 'BAD'. Thus, problematic spots may be identified by looking for spots with fg_flag set to 'GOOD and bg_flag set to 'BAD'.

Rescue of spots with unduly high background

Since it was impossible to discern automatically whether an unusually high background signal indicated a real faulty condition (e.g. dirty chip) or some recoverable artefact (e.g. error in grid positioning), we resorted to visual inspection to rescue some of the spots affected by the high-backgound problem and identified them with manually associated flag (BAD for spot with high background and GOOD for spot with low background).

Spatial smoothing

Microarray data is often affected by "spatial bias", i.e. certain areas of a chip look brighter or darker than others simply due to their position and not because of biologically relevant reasons. To cope with this problem, we applied a "smoothing" filter (Workman et al., 2002) that successfully removed some of the bias and mitigated random noise, as is shown by the increased reproducibility of the experiments (higher correlation between array replicates).

Collapse replicates

Since experiments were performed in duplicate, the final foreground and background intensity values for each spot were computed by taking the mean of the two replicated measures.

GST – Pull down assay

HEK293 cells were grown in DMEM (Invitrogen) supplemented with 10% fetal bovine serum, at 37 C in 5% CO2 in a humid environment. Cells were lysed in 1% Triton X-100, 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 50 mM β-glycerophosphate, 20 mM sodium pyrophosphate, 30 mM NaF, 1

mM benzamidine, 2 mM EGTA, 200 μM NaVO4, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 10 μg/ml pepstatin, and 1 μg/ml microcystin-LR. The amount of cell lysate produced was determined with a Bio-Rad protein assay. GST-SH3 protein (100 μg) and GST (100 μg) slurry were incubated each with 1 mg of protein lysate for 2 h, rocking, at 4 C. The beads were centrifuged and washed in lysis buffer. 30 μl of reducing SDS sample buffer was added; tubes were boiled for 5 min and the 30% of contents run on 10% SDS-PAGE for 2.5 h, at 65 mA. Separated proteins were transferred to nitrocellulose membranes for 1 h at 100 V. After 1 h of incubation in blocking buffer (5% non-fat milk or 5% bovine serum albumin in 20 mM Tris–HCl (pH 7.5), 137 mM NaCl, and 0.1% Tween 20) the membranes were probed with anti Cbl-b and anti GAB1 (Santa Cruz Biotechnology) and to reveal the amount of immunoprecipitated proteins, the membrane were probed with horseradish peroxidase-conjugated anti-rabbit secondary antibody (Amersham Pharmacia Biotech) and visualized by enhanced chemiluminescence (Amersham Pharmacia Biotech) on the LAS-3000 imaging system (Fujifilm Life Science).

Spot synthesis

Cellulose membrane-bound peptides were automatically synthesized according to standard SPOT synthesis protocols (Kramer and Schneider-Mergener, 1998) using a Spot synthesizer MultiPep Spotter (Intavis AG, Germany). To limit background signals, all cysteines were replaced by serines. The generated arrays of peptides were synthesized on Amino-PEG500-UC540 Sheet (acid hardened) (Intavis AG, Germany). The membrane, was activated in Ethanol (3 times for 10 minutes), washed 3 times in PBS, blocked in PBS/BSA 5% and then incubated with Glutathione S-Transferase (GST) fusion proteins (10 micrograms/ml) in PBS/BSA 5%. Incubation with GST alone at the same concentration in the same buffer is used as negative control. The membrane was then treated with anti GST antibody and finally with peroxidase labeled anti goat antibody. Chemiluminescent signals were acquired with a LAS3000 instrument (FujiFilm) and analyzed with the software AIDA.

Phage panning

For panning experiments we have utilized a pVIII phagemid library that displays nonapeptide of random sequence after the AEGEF amino-terminal peptide of the major capsid protein (Felici et al., 1991). Every panning cycle was carried out with 10 µg of bait protein and 10^{10} transducing units (~2 pmol of random nonamers) in 100 µl phosphate-buffered saline (PBS), 1% bovine serum albumin (BSA) for 1 h at 4°C. The GST–SH3 domain fusion protein was immobilized by binding to 20 µl of gluthatione–Sepharose matrix. After two or three panning cycles, the selected phage clones were tested in ELISA against their bait proteins. Single-stranded DNA was prepared as described previously (Dente et al., 1997) and sequenced.

Phage ELISA

Each well was loaded with 0.2 µg of the appropriate GST-fusion protein in 100 µl PBS and was incubated O.N. at 4°C; after blocking 1h 37°C in 150 µl each well with PBS, dry milk 5%, and washing, 10^{10} transducing units of the selected phage was applied to each well in 50 µl of PBS, 5% dry milk, and incubated for 2h R.T. After washing, detection was performed using a 1:5000 dilution of a HRP anti-M13 conjugated antibody (GE Healthcare) pre-adsorbed for 2h with E.Coli cell extract; diluition was carried out in PBS, 5% dry milk and was added 50 µl to each well incubating for 1 h at room temperature. The chromogenic reaction was developed by adding 50 µl of ABTS substrate solution to each well. Reading was performed at 405 nm.

References

Adzhubei, A.A., and Sternberg, M.J. (1993). Left-handed polyproline II helices commonly occur in globular proteins. J Mol Biol *229*, 472-493.

Andreev, J., Simon, J.P., Sabatini, D.D., Kam, J., Plowman, G., Randazzo, P.A., and Schlessinger, J. (1999). Identification of a new Pyk2 target protein with Arf-GAP activity. Mol Cell Biol *19*, 2338-2350.

Briggs, S.D., Sharkey, M., Stevenson, M., and Smithgall, T.E. (1997). SH3-mediated Hck tyrosine kinase activation and fibroblast transformation by the Nef protein of HIV-1. J Biol Chem *272*, 17899-17902.

Carducci, M., Licata, L., Peluso, D., Castagnoli, L., and Cesareni, G. (2009). Enriching the viral-host interactomes with interactions mediated by SH3 domains. Amino Acids.

Ceol, A., Chatr-aryamontri, A., Santonico, E., Sacco, R., Castagnoli, L., and Cesareni, G. (2007). DOMINO: a database of domain-peptide interactions. Nucleic Acids Res *35*, D557-560.

Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A., Migone, N., Hoffmuller, U., Schneider-Mergener, J., and Cesareni, G. (1999). The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. J Biol Chem *274*, 32001-32007.

Chan, B., Lanyi, A., Song, H.K., Griesbach, J., Simarro-Grande, M., Poy, F., Howie, D., Sumegi, J., Terhorst, C., and Eck, M.J. (2003). SAP couples Fyn to SLAM immune receptors. Nat Cell Biol *5*, 155-160.

Chatr-aryamontri, A., Ceol, A., Palazzi, L.M., Nardelli, G., Schneider, M.V., Castagnoli, L., and Cesareni, G. (2007). MINT: the Molecular INTeraction database. Nucleic Acids Res *35*, D572-574.

Chatr-aryamontri, A., Ceol, A., Peluso, D., Nardozza, A., Panni, S., Sacco, F., Tinti, M., Smolyar, A., Castagnoli, L., Vidal, M*., et al.* (2009). VirusMINT: a viral protein interaction database. Nucleic Acids Res *37*, D669-673.

Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G.H., French, S., Howk, R., Ricca, G.A., and Jaye, M. (1994). Identification of a Src SH3 domain binding motif by screening a random phage display library. J Biol Chem *269*, 24034-24039.

Dente, L., Vetriani, C., Zucconi, A., Pelicci, G., Lanfrancone, L., Pelicci, P.G., and Cesareni, G. (1997). Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. J Mol Biol *269*, 694-703.

Donaldson, J.C., Dempsey, P.J., Reddy, S., Bouton, A.H., Coffey, R.J., and Hanks, S.K. (2000). Crk-associated substrate p130(Cas) interacts with nephrocystin and both proteins localize to cell-cell contacts of polarized epithelial cells. Exp Cell Res *256*, 168-178.
Duke-Cohan, J.S., Kang, H., Liu, H., and Rudd, C.E. (2006). Regulation and function of SKAP-55 non-canonical motif binding to the SH3c domain of adhesion and degranulation-promoting adaptor protein. J Biol Chem *281*, 13743-13750.
Dyson, N., Guida, P., Munger, K., and Harlow, E. (1992). Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. J Virol *66*, 6893-6902.
Eckert, B., Martin, A., Balbach, J., and Schmid, F.X. (2005). Prolyl isomerization as a molecular timer in phage infection. Nat Struct Mol Biol *12*, 619-623.
Felici, F., Castagnoli, L., Musacchio, A., Jappelli, R., and Cesareni, G. (1991). Selection of antibody ligands from a large library of oligopeptides expressed on a multivalent exposition vector. J Mol Biol *222*, 301-310.
Feng, S., Chen, J.K., Yu, H., Simon, J.A., and Schreiber, S.L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241-1247.
Feng, S., Kasahara, C., Rickles, R.J., and Schreiber, S.L. (1995). Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. Proc Natl Acad Sci U S A *92*, 12408-12415.
Groemping, Y., Lapouge, K., Smerdon, S.J., and Rittinger, K. (2003). Molecular basis of phosphorylation-induced activation of the NADPH oxidase. Cell *113*, 343-355.
Gu, W., and Helms, V. (2005). Dynamical binding of proline-rich peptides to their recognition domains. Biochim Biophys Acta *1754*, 232-238.
Hashimoto, S., Hirose, M., Hashimoto, A., Morishige, M., Yamada, A., Hosaka, H., Akagi, K., Ogawa, E., Oneyama, C., Agatsuma, T.*, et al.* (2006). Targeting AMAP1 and cortactin binding bearing an atypical src homology 3/proline interface for prevention of breast cancer invasion and metastasis. Proc Natl Acad Sci U S A *103*, 7036-7041.
Hassaine, G., Courcoul, M., Bessou, G., Barthalay, Y., Picard, C., Olive, D., Collette, Y., Vigne, R., and Decroly, E. (2001). The tyrosine kinase Hck is an inhibitor of HIV-1 replication counteracted by the viral vif protein. J Biol Chem *276*, 16885-16893.
Hiyoshi, M., Suzu, S., Yoshidomi, Y., Hassan, R., Harada, H., Sakashita, N., Akari, H., Motoyoshi, K., and Okada, S. (2008). Interaction between Hck

and HIV-1 Nef negatively regulates cell surface expression of M-CSF receptor. Blood *111*, 243-250.
Holgado-Madruga, M., Emlet, D.R., Moscatello, D.K., Godwin, A.K., and Wong, A.J. (1996). A Grb2-associated docking protein in EGF- and insulin-receptor signalling. Nature *379*, 560-564.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.
Janz, J.M., Sakmar, T.P., and Min, K.C. (2007). A novel interaction between atrophin-interacting protein 4 and beta-p21-activated kinase-interactive exchange factor is mediated by an SH3 domain. J Biol Chem *282*, 28893-28903.
Jozic, D., Cardenes, N., Deribe, Y.L., Moncalian, G., Hoeller, D., Groemping, Y., Dikic, I., Rittinger, K., and Bravo, J. (2005). Cbl promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol *12*, 972-979.
Kaneko, T., Li, L., and Li, S.S. (2008). The SH3 domain--a family of versatile peptide- and protein-recognition module. Front Biosci *13*, 4938-4952.
Kang, H., Freund, C., Duke-Cohan, J.S., Musacchio, A., Wagner, G., and Rudd, C.E. (2000). SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55. EMBO J *19*, 2889-2899.
Karkkainen, S., Hiipakka, M., Wang, J.H., Kleino, I., Vaha-Jaakkola, M., Renkema, G.H., Liss, M., Wagner, R., and Saksela, K. (2006). Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. EMBO Rep *7*, 186-191.
Kato, M., Miyazawa, K., and Kitamura, N. (2000). A deubiquitinating enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP. J Biol Chem *275*, 37481-37487.
Kesti, T., Ruppelt, A., Wang, J.H., Liss, M., Wagner, R., Tasken, K., and Saksela, K. (2007). Reciprocal regulation of SH3 and SH2 domain binding via tyrosine phosphorylation of a common site in CD3epsilon. J Immunol *179*, 878-885.
Korkaya, H., Jameel, S., Gupta, D., Tyagi, S., Kumar, R., Zafrullah, M., Mazumdar, M., Lal, S.K., Xiaofang, L., Sehgal, D.*, et al.* (2001). The ORF3 protein of hepatitis E virus binds to Src homology 3 domains and activates MAPK. J Biol Chem *276*, 42389-42400.
Kowanetz, K., Husnjak, K., Holler, D., Kowanetz, M., Soubeyran, P., Hirsch, D., Schmidt, M.H., Pavelic, K., De Camilli, P., Randazzo, P.A.*, et*

al. (2004). CIN85 associates with multiple effectors controlling intracellular trafficking of epidermal growth factor receptors. Mol Biol Cell *15*, 3155-3166.
Kramer, A., and Schneider-Mergener, J. (1998). Synthesis and screening of peptide libraries on continuous cellulose membrane supports. Methods Mol Biol *87*, 25-39.
Kurakin, A.V., Wu, S., and Bredesen, D.E. (2003). Atypical recognition consensus of CIN85/SETA/Ruk SH3 domains revealed by target-assisted iterative screening. J Biol Chem *278*, 34102-34109.
Landgraf, C., Panni, S., Montecchi-Palazzi, L., Castagnoli, L., Schneider-Mergener, J., Volkmer-Engert, R., and Cesareni, G. (2004). Protein interaction networks by proteome peptide scanning. PLoS Biol *2*, E14.
Latreille, M., and Larose, L. (2006). Nck in a complex containing the catalytic subunit of protein phosphatase 1 regulates eukaryotic initiation factor 2alpha signaling and cell survival to endoplasmic reticulum stress. J Biol Chem *281*, 26633-26644.
Lazzi, S., Bellan, C., De Falco, G., Cinti, C., Ferrari, F., Nyongo, A., Claudio, P.P., Tosi, G.M., Vatti, R., Gloghini, A*., et al.* (2002). Expression of RB2/p130 tumor-suppressor gene in AIDS-related non-Hodgkin's lymphomas: implications for disease pathogenesis. Hum Pathol *33*, 723-731.
Lechner, M.S., and Laimins, L.A. (1994). Inhibition of p53 DNA binding by human papillomavirus E6 proteins. J Virol *68*, 4262-4273.
Levine, A.J., Momand, J., and Finlay, C.A. (1991). The p53 tumour suppressor gene. Nature *351*, 453-456.
Lewitzky, M., Harkiolaki, M., Domart, M.C., Jones, E.Y., and Feller, S.M. (2004). Mona/Gads SH3C binding to hematopoietic progenitor kinase 1 (HPK1) combines an atypical SH3 binding motif, R/KXXK, with a classical PXXP motif embedded in a polyproline type II (PPII) helix. J Biol Chem *279*, 28724-28732.
Li, S.S. (2005). Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. Biochem J *390*, 641-653.
Lim, W.A., Richards, F.M., and Fox, R.O. (1994). Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375-379.
Liu, J., Li, M., Ran, X., Fan, J.S., and Song, J. (2006). Structural insight into the binding diversity between the human Nck2 SH3 domains and proline-rich proteins. Biochemistry *45*, 7171-7184.
Liu, Q., Berry, D., Nash, P., Pawson, T., McGlade, C.J., and Li, S.S. (2003). Structural basis for specific binding of the Gads SH3 domain to an RxxK

motif-containing SLP-76 peptide: a novel mode of peptide recognition. Mol Cell *11*, 471-481.
Liu, Y., and Rohrschneider, L.R. (2002). The gift of Gab. FEBS Lett *515*, 1-7.
Lock, L.S., Royal, I., Naujokas, M.A., and Park, M. (2000). Identification of an atypical Grb2 carboxyl-terminal SH3 domain binding site in Gab docking proteins reveals Grb2-dependent and -independent recruitment of Gab1 to receptor tyrosine kinases. J Biol Chem *275*, 31536-31545.
Lowenstein, E.J., Daly, R.J., Batzer, A.G., Li, W., Margolis, B., Lammers, R., Ullrich, A., Skolnik, E.Y., Bar-Sagi, D., and Schlessinger, J. (1992). The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. Cell *70*, 431-442.
Lupher, M.L., Jr., Rao, N., Eck, M.J., and Band, H. (1999). The Cbl protooncoprotein: a negative regulator of immune receptor signal transduction. Immunol Today *20*, 375-382.
Macias, M.J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., and Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature *382*, 646-649.
Massenet, C., Chenavas, S., Cohen-Addad, C., Dagher, M.C., Brandolin, G., Pebay-Peyroula, E., and Fieschi, F. (2005). Effects of p47phox C terminus phosphorylations on binding interactions with p40phox and p67phox. Structural and functional comparison of p40phox and p67phox SH3 domains. J Biol Chem *280*, 13752-13761.
Mayer, B.J. (2001). SH3 domains: complexity in moderation. J Cell Sci *114*, 1253-1263.
Mayer, B.J., Hamaguchi, M., and Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. Nature *332*, 272-275.
Mittelstaedt, T., and Schoch, S. (2007). Structure and evolution of RIM-BP genes: identification of a novel family member. Gene *403*, 70-79.
Modregger, J., Ritter, B., Witter, B., Paulsson, M., and Plomann, M. (2000). All three PACSIN isoforms bind to endocytic proteins and inhibit endocytosis. J Cell Sci *113 Pt 24*, 4511-4521.
Moncalian, G., Cardenes, N., Deribe, Y.L., Spinola-Amilibia, M., Dikic, I., and Bravo, J. (2006). Atypical polyproline recognition by the CMS N-terminal Src homology 3 domain. J Biol Chem *281*, 38845-38853.
Mongiovi, A.M., Romano, P.R., Panni, S., Mendoza, M., Wong, W.T., Musacchio, A., Cesareni, G., and Di Fiore, P.P. (1999). A novel peptide-SH3 interaction. EMBO J *18*, 5300-5309.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. Nature *359*, 851-855.
Nishizawa, K., Freund, C., Li, J., Wagner, G., and Reinherz, E.L. (1998). Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. Proc Natl Acad Sci U S A *95*, 14897-14902.
Persico, M., Ceol, A., Gavrila, C., Hoffmann, R., Florio, A., and Cesareni, G. (2005). HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. BMC Bioinformatics *6 Suppl 4*, S21.
Petrelli, A., Gilestro, G.F., Lanzardo, S., Comoglio, P.M., Migone, N., and Giordano, S. (2002). The endophilin-CIN85-Cbl complex mediates ligand-dependent downregulation of c-Met. Nature *416*, 187-190.
Pornillos, O., Alam, S.L., Rich, R.L., Myszka, D.G., Davis, D.R., and Sundquist, W.I. (2002). Structure and functional interactions of the Tsg101 UEV domain. EMBO J *21*, 2397-2406.
Rao, N., Dodge, I., and Band, H. (2002). The Cbl family of ubiquitin ligases: critical negative regulators of tyrosine kinase signaling in the immune system. J Leukoc Biol *71*, 753-763.
Reinhard, M., Giehl, K., Abel, K., Haffner, C., Jarchau, T., Hoppe, V., Jockusch, B.M., and Walter, U. (1995). The proline-rich focal adhesion and microfilament protein VASP is a ligand for profilins. EMBO J *14*, 1583-1589.
Renkema, G.H., Manninen, A., and Saksela, K. (2001). Human immunodeficiency virus type 1 Nef selectively associates with a catalytically active subpopulation of p21-activated kinase 2 (PAK2) independently of PAK2 binding to Nck or beta-PIX. J Virol *75*, 2154-2160.
Ribon, V., Printen, J.A., Hoffman, N.G., Kay, B.K., and Saltiel, A.R. (1998). A novel, multifuntional c-Cbl binding protein in insulin receptor signaling in 3T3-L1 adipocytes. Mol Cell Biol *18*, 872-879.
Rickles, R.J., Botfield, M.C., Weng, Z., Taylor, J.A., Green, O.M., Brugge, J.S., and Zoller, M.J. (1994). Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. EMBO J *13*, 5598-5604.
Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W.*, et al.* (2000). Comparative genomics of the eukaryotes. Science *287*, 2204-2215.
Schaeper, U., Gehring, N.H., Fuchs, K.P., Sachs, M., Kempkes, B., and Birchmeier, W. (2000). Coupling of Gab1 to c-Met, Grb2, and Shp2 mediates biological responses. J Cell Biol *149*, 1419-1432.
Schneider-Mergener, J. (2001). Synthetic Peptide and Protein Domain

Arrays Prepared by the SPOT Technology. Comp Funct Genomics *2*, 307-309.
Schutt, C.E., Myslik, J.C., Rozycki, M.D., Goonesekere, N.C., and Lindberg, U. (1993). The structure of crystalline profilin-beta-actin. Nature *365*, 810-816.
Seet, B.T., Berry, D.M., Maltzman, J.S., Shabason, J., Raina, M., Koretzky, G.A., McGlade, C.J., and Pawson, T. (2007). Efficient T-cell receptor signaling requires a high-affinity interaction between the Gads C-SH3 domain and the SLP-76 RxxK motif. EMBO J *26*, 678-689.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Shelton, H., and Harris, M. (2008). Hepatitis C virus NS5A protein binds the SH3 domain of the Fyn tyrosine kinase with high affinity: mutagenic analysis of residues within the SH3 domain that contribute to the interaction. Virol J *5*, 24.
Soubeyran, P., Barac, A., Szymkiewicz, I., and Dikic, I. (2003). Cbl-ArgBP2 complex mediates ubiquitination and degradation of c-Abl. Biochem J *370*, 29-34.
Soubeyran, P., Kowanetz, K., Szymkiewicz, I., Langdon, W.Y., and Dikic, I. (2002). Cbl-CIN85-endophilin complex mediates ligand-induced downregulation of EGF receptors. Nature *416*, 183-187.
Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M., and Kay, B.K. (1996). Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins. Nat Biotechnol *14*, 741-744.
Stahl, M.L., Ferenz, C.R., Kelleher, K.L., Kriz, R.W., and Knopf, J.L. (1988). Sequence similarity of phospholipase C with the non-catalytic region of src. Nature *332*, 269-272.
Stamenova, S.D., French, M.E., He, Y., Francis, S.A., Kramer, Z.B., and Hicke, L. (2007). Ubiquitin binds to and regulates a subset of SH3 domains. Mol Cell *25*, 273-284.
Stapley, B.J., and Creamer, T.P. (1999). A survey of left-handed polyproline II helices. Protein Sci *8*, 587-595.
Tian, L., Chen, L., McClafferty, H., Sailer, C.A., Ruth, P., Knaus, H.G., and Shipston, M.J. (2006). A noncanonical SH3 domain binding motif links BK channels to the actin cytoskeleton via the SH3 adapter cortactin. FASEB J *20*, 2588-2590.
Tong, A.H., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., *et al.* (2002). A

combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science *295*, 321-324.
Tong, X.K., Hussain, N.K., de Heuvel, E., Kurakin, A., Abi-Jaoude, E., Quinn, C.C., Olson, M.F., Marais, R., Baranes, D., Kay, B.K., *et al.* (2000). The endocytic protein intersectin is a major binding partner for the Ras exchange factor mSos1 in rat brain. EMBO J *19*, 1263-1271.
Trible, R.P., Emert-Sedlak, L., and Smithgall, T.E. (2006). HIV-1 Nef selectively activates Src family kinases Hck, Lyn, and c-Src through direct SH3 domain interaction. J Biol Chem *281*, 27029-27038.
Vaynberg, J., Fukuda, T., Chen, K., Vinogradova, O., Velyvis, A., Tu, Y., Ng, L., Wu, C., and Qin, J. (2005). Structure of an ultraweak protein-protein complex and its crucial role in regulation of cell morphology and motility. Mol Cell *17*, 513-523.
Weidner, K.M., Di Cesare, S., Sachs, M., Brinkmann, V., Behrens, J., and Birchmeier, W. (1996). Interaction between Gab1 and the c-Met receptor tyrosine kinase is responsible for epithelial morphogenesis. Nature *384*, 173-176.
Whisstock, J.C., and Lesk, A.M. (1999). SH3 domains in prokaryotes. Trends Biochem Sci *24*, 132-133.
Williamson, M.P. (1994). The structure and function of proline-rich regions in proteins. Biochem J *297 (Pt 2)*, 249-260.
Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol *3*, research0048.
Wu, C., Ma, M.H., Brown, K.R., Geisler, M., Li, L., Tzeng, E., Jia, C.Y., Jurisica, I., and Li, S.S. (2007). Systematic identification of SH3 domain-mediated human protein-protein interactions by peptide array target screening. Proteomics *7*, 1775-1785.
Wu, X., Knudsen, B., Feller, S.M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure *3*, 215-226.
Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W., and Schreiber, S.L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. Cell *76*, 933-945.
Yu, H., Kim, P.M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol *3*, e59.
Yu, H., Rosen, M.K., Shin, T.B., Seidel-Dugan, C., Brugge, J.S., and

Schreiber, S.L. (1992). Solution structure of the SH3 domain of Src and identification of its ligand-binding site. Science *258*, 1665-1668.
Yuan, Z.Q., Kim, D., Kaneko, S., Sussman, M., Bokoch, G.M., Kruh, G.D., Nicosia, S.V., Testa, J.R., and Cheng, J.Q. (2005). ArgBP2gamma interacts with Akt and p21-activated kinase-1 and promotes cell survival. J Biol Chem *280*, 21483-21490.
Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., and Cesareni, G. (2002). MINT: a Molecular INTeraction database. FEBS Lett *513*, 135-140.
Zarrinpar, A., Bhattacharyya, R.P., and Lim, W.A. (2003). The structure and function of proline recognition domains. Sci STKE *2003*, RE8.