Send Orders for Reprints to reprints@benthamscience.ae

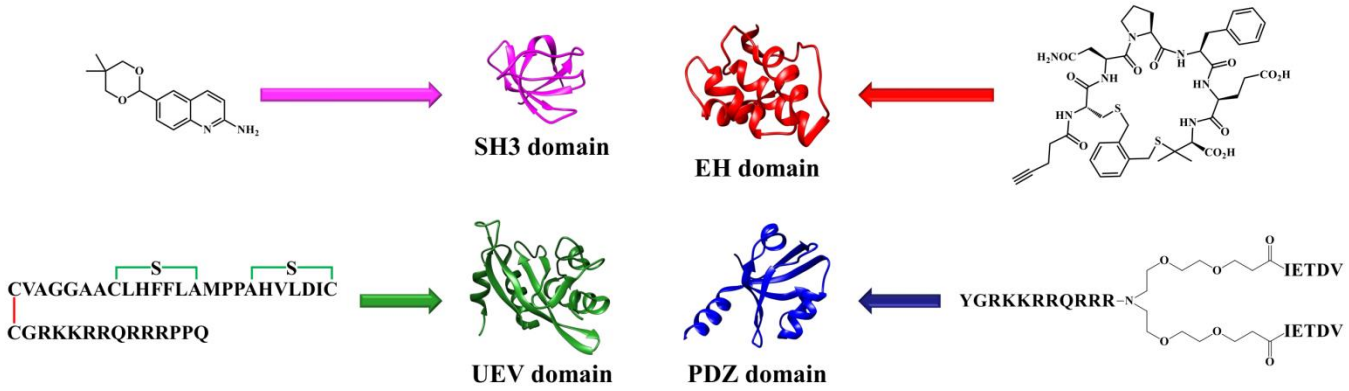
Current Medicinal Chemistry, **2019**, *Volume, Page Enation*

Protein Interaction Domains: structural features and drug discovery applications (part 2).

Marian Vincenzi, Flavia Anna Mercurio, Marilisa Leone*

Institute of Biostructures and Bioimaging, National Research Council (CNR), Via Mezzocannone 16, 80134 Naples, Italy

Graphical abstract



This review focuses on Protein Interaction Domains able to bind peptide motifs; structural features and potential approaches to target related protein-protein interactions and discover novel therapeutics are discussed.

0929-8673/19 $58.00+.00

ARTICLE TYPE

Protein Interaction Domains: structural features and drug discovery applications (part 2).

Marian Vincenzi[a], Flavia Anna Mercurio[a], Marilisa Leone*[a]

[a] *Institute of Biostructures and Bioimaging, National Research Council (CNR), Via Mezzocannone 16, 80134 Naples, Italy*

[*] *Corresponding author: Marilisa Leone, Institute of Biostructures and Bioimaging, National Research Council (CNR), Via Mezzocannone 16, 80134 Naples, Italy, email: marilisa.leone@cnr.it*

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract

Background: Proteins present a modular organization made up of several domains. Apart from domains playing catalytic functions, many others are crucial to recruit interactors. The latter domains can be defined "PIDs" (Protein Interaction Domains) and are responsible for pivotal outcomes in signal transduction and a certain array of normal physiological and disease-related pathways. Targeting such PIDs with small molecules and peptides able to modulate their interaction networks, may represent a valuable route to discover novel therapeutics. ***Objective:*** This work represents a continuation of a very recent review describing PIDs able to recognize post-translationally modified peptide segments. On the contrary, this second part concerns with PIDs that interact with simple peptide sequences provided with standard amino acids. ***Method:*** Crucial structural information on different domain subfamilies and their interactomes was gained by a wide search in different online available databases (including the PDB (Protein Data Bank), the Pfam (Protein family), and the SMART (Simple Modular Architecture Research Tool)). Pubmed was searched as well to explore the most recent literature related to the topic. ***Results and Conclusion:*** PIDs are multifaceted: they have all diverse structural features and can recognize several consensus sequences. PIDs can be linked to different diseases onset and progression, like cancer or viral infections and find applications in the personalized medicine field. Many efforts have been centered on peptide/peptidomimetic inhibitors of PIDs mediated interactions but much more work needs to be conducted to improve drug-likeness and interaction affinities of identified compounds.

Keywords: protein interaction domains, protein-protein interactions, structure, drug discovery, ligands, inhibitors, peptides, small molecules.

1. INTRODUCTION

Regulatory proteins are generally made up of several modules/domains that can be simply involved in molecular interactions or enzymatic activities. In this context protein interaction domains (PIDs) act as structural elements able to confer specific binding properties to the proteins they belong to [1]. PIDs were initially related to phosphotyrosine signaling pathways, due to the capacity of Src homology 2 (SH2) domains of certain cytoplasmic proteins to bind specific peptide segments containing phosphotyrosines from activated tyrosine kinase receptors [2]. However, to date, it is well established that PIDs play crucial functions in signaling from many different cell surface receptors, and intervene in many cellular processes including but not limited to cell cycle, DNA repair, vesicle trafficking, gene expression, cytoskeleton control and protein degradation [2, 3]. In details, PIDs may work to confine proteins to defined subcellular locations, represent "readers" of protein post-translational modifications, mediate formation of large multi-protein signaling complexes, and regulate properties of enzymes including conformation, activity, and substrate specificity [4].

Several proteins in the human genome contain modular PIDs able to bind different peptide motifs. There are interaction domains like SH2, Chromo, Bromo, VHL (Von Hippel-Lindau) domains able to specifically recognize post-translationally modified amino acid sequences [5]. Other PIDs such as SH3 (Src Homology 3) and PDZ (Post synaptic Density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and Zonula Occludens-1 protein (ZO-1)) domains recognize unmodified peptide segments [5].

We have very recently reported on PIDs interacting with post-translationally modified peptide sequences. This second part represents an extension of the first one and will be

centered instead on PIDs able to recognize simple peptides lacking post-translational modifications.

SH3, EVH1 (Enabled/VASP Homology 1), GYF (Glycine-Tyrosine-Phenylalanine) and UEV (Ubiquitin E2 Variant) are protein modules that can bind proline rich segments, and participate to many cellular events such as cell growth, cytoskeletal rearrangements, transcription and postsynaptic signaling [6]. Nevertheless, EH (Eps15 Homology) domains can bind peptides with a -NPF- (Asn-Pro-Phe) amino acid sequence and act in various processes including DNA repair and neuronal functioning at different key levels [7, 8].

Several modular domains have the ability to form dimers and/or oligomers by homo- or hetero-typic interactions such as ~~the~~ SAM (Sterile Alpha Motif) domains [2]. PDZ domains recognize instead C-terminal ends of multiple proteins and guarantee formation of protein assemblies which, in turn, are again involved in regulation of signal transduction, and modulation of trafficking of membrane proteins [9]. Other considerable examples of PIDs are given by LIM (Lin11, Isl-1 & Mec-3) domains [10], which mediate protein interactions between transcription factors, cytoskeletal and signaling proteins, and CH (Calponin Homology) domains [11], with a major regulatory role in muscle contraction.

1.1. PIDs containing proteins: a short overview

PIDs can be found in a vast array of proteins playing different functions. For instance, kinases, lipases and GTPases contain SH3 domains [12] whereas, WASP (Wiskott-Aldrich Syndrome Protein), Homer/Vesl and Spred (Sprouty-related protein) are among the protein families provided with EVH1 domains [13, 14]. The GYF domain was identified in the human CD2BP2 (CD2 binding protein 2) protein but is also contained in other proteins from eukaryotes with unclear functions [15]. Proteins like Vps23 (yeast Vacuolar protein sorting-associated protein 23) and its homologue Mammalian tumor susceptibility gene TSG101 protein contain the ubiquitin binding module UEV; Mms2 (Methyl MethaneSulfonate sensitivity 2) represents as well a UEV domain containing protein that, through hetero-dimerization with Ubc13 (Ubiquitin-conjugating enzyme E2 13), favors the assembly of polyubiquitin chains [16]. LIM domains are located in different proteins of nucleus and cytoplasm that largely work as docking sites where to assemble multi-protein complexes (See paragraph 3.2 for a detailed list of proteins provided with LIM domains) [17]. In addition, there are several cytoskeletal and signal transduction proteins possessing the actin binding CH domain (many of which are mentioned in paragraph 4.1) [18]. EH domains are present in a large number of proteins from Haemosporida Fungi, Plants, Nematodes, Artropoda, Amphibia and Mammalia (for a list of proteins containing EH domains see Table 1 in reference 19) [19]. PDZ domains can be found in a variety of protein families including but not limited to MAGKs (Membrane Associated Guanilate Kinase) and tyrosine phosphatases (See paragraph 6.1 for the different classifications proposed for the PDZ domain containing proteins). Finally SAM domains are instead present in Tyr and Ser/Thr kinases (like Eph (Erythropoietin-producing hepatocellular carcinoma cell) receptors), lipid kinases (like diacylglycerol kinase), scaffolding proteins (like Shank (SH3 and multiple ankyrin repeat domains)), RNA binding proteins (such as Smaug) and transcription factors (like TEL (Translocation-Ets-Leukemia), Yan etc..) [20].

1.2. Exploiting PIDs as therapeutic targets

PIDs hold a great interest as potential therapeutic targets in drug discovery for different pathologies including cancer.

Cancer largely derives by genetic modifications leading to abnormal signaling proteins producing aberrant cell growth, survival, and metastasis [21]. In the field of personalized medicine new strategies could be envisioned by better comprehend the way through which protein modular domains and their interactomes contribute to assembly certain cancer machineries [21]. For example it has been shown that within the human interaction network, certain protein modules could be exploited to classify populations of breast cancer patients with different prognosis [21, 22].

In a similar contest interesting studies have been conducted on the SH2 (Src Homology 2) protein binding modules that recognize pTyr (phosphoTyrosine) provided ligands [23].

High EGFR (Epidermal Growth Factor Receptor) activity is observed in lung cancer and blockage of EGFR by means of the TKI (Tyrosine Kinase Inhibitor) erlotinib can favor survival in patients affected by lung cancer at advanced stadium after chemotherapy failure [24]. Indeed, there are known activating somatic mutations in EGFR that induce sensitivity towards TKIs having EGFR as target [24]. Patients affected by lung cancer, that present these EGFR mutations, respond well to TKIs and have an improved survival perspective than those treated simply with cytotoxic agents [24]. However, drug resistance can arise from mechanisms like MET (Mesenchymal to Epithelial Transition) amplifications or secondary mutations in EGFR [24].

Certain studies reported on the potential impact that SH2 domain profiling could have in personalized medicine due to the ability of this technique to provide detailed knowledge about global tyrosine kinase signaling in lung cancers [21, 24]. SH2 domain profiling relies on a phosphoproteomic approach that allows quantitative detection of the phosphorylated binding loci for SH2 domains, that cells utilize to react to modifications in tyrosine phosphorylation signaling [21, 24]. SH2 profiles were obtained by implementing a set of purified SH2 domains on various lung cancer cell lines, presenting several oncogene (EGFR, KRAS (Kirsten rat sarcoma)) mutations and diverging in drug sensitivity to EGFR TKIs (Tyrosine Kinase Inhibitors) [21, 24]. By analyzing SH2 profiling data and according to SH2 binding patterns, cancer cells could be subdivided in classes with a few groups associated with EGFR or KRAS mutation status [21, 24]. In addition prominent dissimilarities in global tyrosine kinase signaling patterns could be identified among cells having the same histology and characterized by activating EGFR mutations, thus suggesting the relevance of SH2 domains in determining the overall tyrosine kinase signaling in cells directed by the same oncogene. Interactions with specific SH2 domains (for examples activators of RAS pathway like Grb2 (Growth-factor-receptor-bound protein 2) and ShcA (SH2-containing collagen-related transforming protein A)) correlate with EGFR mutation and sensitivity to the EGFR inhibitor

erlotinib; SH2 interaction patterns mirrored activation of MET as well [24].

Inspired by this study on SH2 domains, as already pointed out in a review by Haura [21], it would be fascinating to classify certain cancer lines derived directly from patients according to results from quantitative profiles of different binding modules and find a correlation with clinical outcomes [21]. Being able to set up strategies to recognize networks of PIDs mediated interactions associated with clinically prominent factors characteristic of specific types of cancers, including oncogenes mutations, pathological features, histology, tumor phase and survival could lead to personalized therapies and consequently have a tremendous impact on human health [21]. It could be appealing to generate, starting directly from patients based material, through both experimental methods and data taken from databases, in situ protein-protein interaction networks thus allowing 1) a finest sub-classification of a certain cancer type regardless of the common genotype and 2) identification of key interactions that need to be targeted by medicinal chemistry efforts [21, 24]. In light of this, small-molecule, peptide and peptidomimetic inhibitors or stabilizers of crucial protein–protein interactions mediated by PIDs may hold prominent therapeutic properties [25].

Structural features and interaction properties of different PIDs, that are essential to design compounds targeting specific interaction patterns, will be described in detail in the next paragraphs. In addition, PID potential druggability and capacity to act as therapeutic targets for specific cancers and several pathological conditions will be exhaustively analyzed below.

2. DOMAINS RECOGNIZING PROLINE RICH SEGMENTS

One of the largest groups of protein interaction domains is formed by those modules which recognize Proline-Rich Sequences (PRSs) [6]. These domains, also indicated as PRS-Recognition Domains (or PRDs), play a pivotal role in the formation and regulation of signaling complexes involved in different biological processes [13, 26-28]. Examples of PRDs are represented by SH3 (Src-homology 3), WW (name due to the presence of two conserved tryptophan residues), GYF, and EVH1 (Enabled (Ena)/Vasodilator-stimulated phosphoprotein (VASP) Homology 1) domains [25].

The mechanism by which PRDs interact with PRSs depends on the combination of different features characterizing proline residues such as the uncommon form of the pyrrolidine ring, the cyclic side-chain with the consequent conformational constraints on the dihedral angles and on the secondary structural preferences, the substituted amide nitrogen, and the CIS-TRANS isomerization with partial stability of the CIS isomer of the Xaa-Pro peptide groups [6]. The common feature shared by the bound states of motifs recognized by PRDs, is the tendency to assume a polyproline type II (PPII) helical structure which consists of an extended left-handed organization characterized by three residues per turn and precise values of the backbone angles (Φ = -78° and Ψ = +146°) [6]. PPII presents two characteristics which make it an advantageous structural element for the interaction with proteins [6]. In fact, the combined effect of the exposure in solution at regular interval of both the side-chains and the backbone carbonyls together with the lack of a backbone H-bond donor on the proline residue, significantly contribute to the absence of intramolecular H-bonds in the structure and thus favor an easier "read" by target proteins which recognize PRSs [6]. The interaction between proteins and PRSs is also enhanced by the structural restriction which characterizes the PPII structural arrangement [29]. Indeed, the backbone conformation presents restrains which lower the entropic cost associated with the binding [29]. On the other side, the interaction pockets of a considerable number of PRDs are characterized by aromatic residues whose planar side-chains fit with high complementarities the crests and grooves located on the PPII helix surface [6]**.** Furthermore, PPII helices possess an axial symmetry which allows the definition of two different types of consensus motifs, one bearing the PXXP sequence (typical of EVH1 and SH3 domains; where X= any residue) and the other including the XPPX sequence (canonical for WW and GYF domains) [25].

Each PRDs will be described more in detail and from different points of views (i.e., biological function, structural details and, if possible, drug discovery applications) in the following subparagraphs. WW domains being also able to bind sequences with certain phosphorylated amino acids (pSer and pThr) have already been described in our previous review [23].

2.1. SH3 (Src Homology 3) domains

SH3 domains are small protein modules composed by 55-70 residues and firstly identified as homology domains located in the tyrosine kinase product of the v-Src oncogene [12].

SH3 domains can be found in different proteins such as kinases, lipases, GTPases, adapter and structural proteins as well as viral regulatory proteins [12]. Furthermore, the presence of SH3 modules may be accompanied by other protein–protein interaction domains such as SH2 or PH (Pleckstrin Homology) domains but also by catalytic modules [12]. In this way, SH3 domains and their interactions with PRSs represent key elements in different biological processes by regulating enzymes through intramolecular interactions of reasonable affinity and selectivity, enhancing the local concentration or altering subcellular localization of specific members of diverse signaling routes (related but not limited to cell proliferation, migration and cytoskeletal modifications), mediating the raise of large multiprotein complexes [30, 31].

2.1.1 Structural features and interaction properties

SH3 domain structure consists of five β-strands (βA, βB, βC, βD and βE) and a single turn of 3_{10} helix (Fig. 1) [12, 32]. Furthermore, SH3 domains structures are characterized by a variable RT-loop (name due to the presence of conserved arginine and threonine residues) with a discrete β-sheet nature, that is positioned between the first two β-strands; a n-Src loop (name related to the substantial larger length found in Src from neuronal tissue) located in between βB and βC; a distal loop separating βC from βD and, a single 3_{10} helical turn between βD and βE (Fig. 1) [12, 32]. These structural elements are organized into two three-stranded antiparallel β-sheets (one formed by βE, βA and βB and the other by βB, βC, βD) which assume a perpendicular orientation with respect to each other and form the hydrophobic pocket by

which SH3 domains recognize PRSs with an affinity which may vary from being moderate to high (Fig. 1) [12, 29, 32, 33]. In addition, the N-terminal and C-terminal ends of SH3 domains are positioned close to each other but on opposite sides with respect to the binding cleft (Fig. 1) [12].

The binding surface of SH3 domains can be subdivided into three sub-pockets [12]. Two of them are hydrophobic clefts composed mainly by aromatic amino acids (W, Y and F), which are conserved among different SH3 domains (Fig. 1) [12, 32-34]. Furthermore, SH3 domains have also a third negatively charged cavity which is named "specificity pocket" and can be composed by elements belonging to the RT- and n-Src loops [34-36].

A PXXP (P=Proline, X=any amino acid) exposed motif on a protein binds SH3 domains with a PPII conformation, as mentioned in the previous paragraph, but, it possesses a pseudo-symmetry and thus it can assume two different orientations (known as "plus" and "minus" orientations) in the binding pocket [32]. An aliphatic amino acid (θ) positioned at specific positions within the consensus motif influences somehow the polarity of the PPII helical arrangement and ligand orientation within the SH3 binding cleft [32]. Furthermore, a considerable contribution to the polarity and ligand orientational preferences is given also by a basic amino acid (arginine or lysine) located in the -3 position with respect to the first N-terminal proline of the consensus sequence PXXP or +2 with respect to the last C-terminal proline [32, 37]. Thus, recognition of ligands by SH3 domains depends on the basic residue (R/K) which interacts with an acidic amino acid present in the RT-loop (specificity pocket) and on the XP dipeptides of the consensus motif which fit the two hydrophobic clefts [38].
In major details, SH3 ligands can be divided into two groups, one termed "Class I" and characterized by the motif "(R/K)XθPXXP" and the other named "Class II" and possessing the "PXθPX(R/K)" sequence (Fig. 1) [33, 35, 36]. The presence of (R/K) in these two different positions of the sequence induces two different orientations of the consensus sequence with respect to the binding surface of the SH3 domain (Fig. 1) [33, 35, 36].
Interestingly, there are also cases in which SH3 domains recognize "non-canonical" consensus sequences. For instance, CIN85 (Cbl INteracting protein of 85 kDa) and CMS (Cas ligand with Multiple SH3 domains) are two proteins characterized by three SH3 domains which bind a proline-arginine motif (i.e., PXXXPR) [38, 39]. In addition, the SH3 domain of cortactin, an actin-binding protein, interacts with the RXXPXXXP motifs which have some similarity to those of "Class I" ligands, with the only difference that there is an additional residue between the two conserved prolines [38, 40]. Instead, the C-SH3 domain of ADAP/SLAP (Adaptor adhesion and Degranulation promoting Adaptor Protein / or SLP-76-Associated Protein) interacts with a consensus sequence which is characterized by the absence of prolines (i.e., RKXXY) [38]. Another example of unusual consensus motif is given by PX(V/I)(D/N)RXXKP which is recognized by the SH3 domains of STAM2 (Signal Transducing Adapter Molecule 2) and Gads (Grb2-related adaptor downstream of Shc) [38, 41].

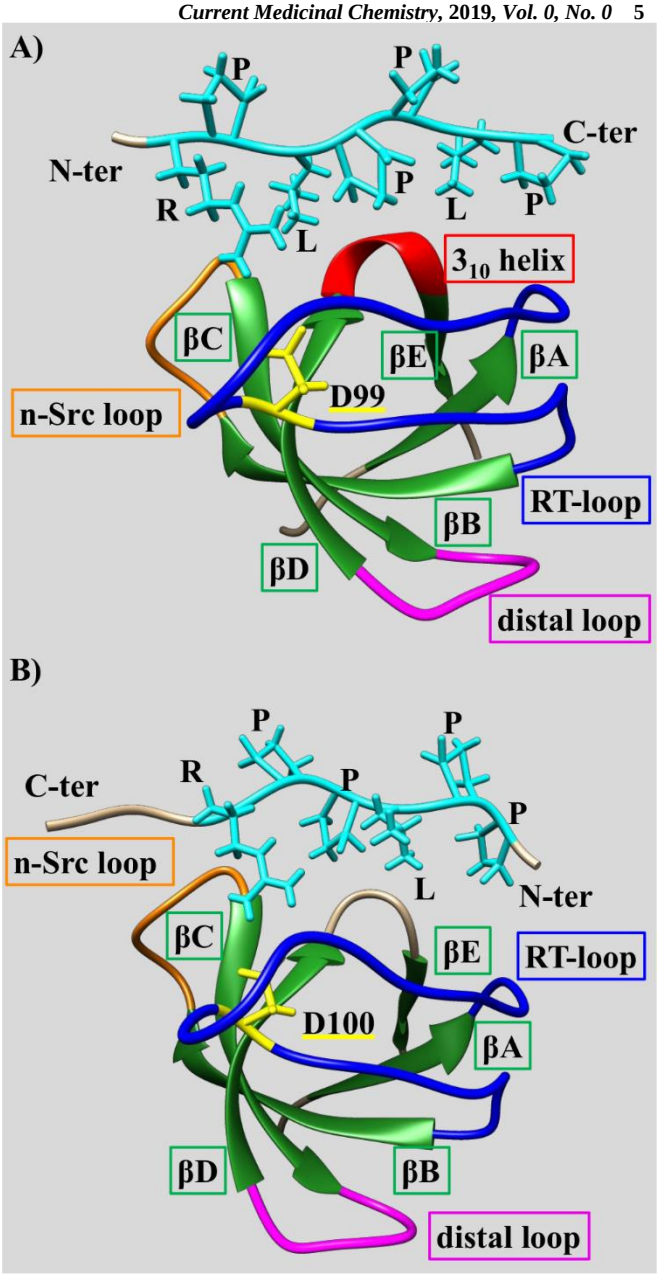


Fig. (**1**). A) Crystal structure the c-Src-SH3 domain in complex with a peptide (VSL12) belonging to the "Class I" group (VSLARRPLPPLP, PDB code: 4RTZ, Table S1). The residues of the RXθPXXP consensus sequence are colored in cyan. The β-strands forming the two three-stranded antiparallel β-sheets (one composed by βA (T85-A88), βB (R107-N112) and βE (V137-P139) and the other by βB (R107-N112), βC (W118-S123), βD (T129-P133)) are indicated in green. The 3_{10} helical turn between βD and βE is shown in red. RT-loop (Y92-E106), n-Src loop (N113-D117) and distal loop (L124-Q128) are reported in blue, orange and magenta, respectively. The side chain of D99 in the specificity pocket on the RT loop is providing a crucial interaction with the arginine in the peptide ligand. B) Crystal structure of the Fyn SH3 domain in complex with a peptide (APP12) belonging to the "Class II" group (APPLPPRNRPRL, PDB code: 4ZNX [42], chains A and E, Table S1). The residues of the PXθPXR consensus sequence are colored in cyan. The β-strands forming the two three-stranded antiparallel β-sheets (βA (L86-A89), βB (K108-N113), βC (W119-S124), βD (T130-P134) and βE (V138-P140)) are colored green. RT-loop (L90-E107), n-Src loop (S114-D118) and distal loop (L125-E129) are colored blue, orange and magenta, respectively. D100 side chain is positioned on the RT loop in the specificity pocket and is involved in a crucial interaction with the arginine in the peptide ligand.

2.1.2 Drug discovery approaches: Peptides/ Peptidomimetics

The attention on SH3 domains is due to their involvement in different biological processes which assume relevance in pathological conditions including cancer, osteoporosis, periodontitis, multiple myeloma and inflammation [43, 44].

For instance, it is well known that proliferative signals may be initiated by tyrosine kinases and transmitted through protein-protein interactions involving SH2 and SH3 domains [45]. Therefore, it has been proposed that the design of peptidomimetics able to inhibit SH3 domains interactions with specific ligands may be considered an interesting route for the treatment of proliferative diseases in which constitutively activated tyrosine kinases play a critical role [45]. Examples are given by chronic myelogenous and acute lymphocytic leukemias which depend on BCR (Breakpoint Cluster Region protein)/ABL (ABelson murine Leukemia viral oncogene homolog 1) and by ovarian and breast cancers linked to HER2 (Human Epidermal growth factor Receptor 2) [45].

HOSF (human Osteoclast-Stimulating Factor), is a protein engaged in cellular signal transduction cascades, which lead to the formation of osteoclasts and the resorption of bones, and is provided with a SH3 domain which mediates its interaction with biological interactors [44]. Therefore, it has been proposed that targeting protein-protein interactions related to the SH3 domain of hOSF may represent a valuable therapeutic route for the treatment of diseases affecting bones like osteoporosis, periodontitis and multiple myeloma [44]. The peptide region 425-APPARPVK-432 from the protein Sam68 (Src-associated in mitosis 68 kDa protein) binds hOSF-SH3 domain with a dissociation constant K_d= 3.2 μM [44]. In a recent study, this sequence has been considered as starting point to develop novel improved ligand of the hOSF-SH3 domain [44]. More in detail, the two key prolines (P427 and P430) of Sam68 peptide have been mutated in different N-substituted amino-acids and the resulting peptide-peptoid hybrids have been evaluated in fluorescence spectroscopy-based binding assays and structural studies (Fig. 2A) [44]. The results revealed that the first proline (P427) residue in the PXXP core of the starting peptide is crucial to guarantee the presence of the polyproline II helix conformation thus supporting the recognition by the SH3 domain [44]. Instead, substitution of P430 with the N-Clp (where Clp= Cyclo-propylamine side chain) favors a considerable increase of affinity (K_d= 0.87 μM) whereas, replacement of the same proline with N-Ffa (where N-Ffa= tetrahydroFurfurylamine moiety) increases only slightly the affinity for the SH3 domain (K_d= 2.9 μM) (Fig. 2A) [44].

One of the major issues when dealing with the design of inhibitors of SH3 domains mediated interactions is to achieve a decent level of selectivity towards the targets. SH3 domains interact with PXXP motifs and selectivity is provided by flanking amino acids resulting in a limited number of possibilities. Peptoids generated by replacing prolines with N-substituted side chains introduce a higher level of variability and may enhance selectivity. In this context, an interesting study was centered on the SH3 domains of the proto-oncoproteins Crk (CT10 (Chicken. Tumor virus n. 10) regulator of kinase), Src and N-Grb2 (amino terminal domain of the Growth factor receptor-bound protein 2). SH3 domains from these proteins can bind consensus sequence - like PXXPXR/K - and interact for example with the peptide -YEVPPPVPPRRR- from the Sos (Son of sevenless) protein with similar affinities (dissociation constants K_d= 6 μM, 25 μM and 5 μM, for Crk-SH3, Src-SH3 and N-Grb2-SH3, respectively) [43]. The purpose of the study was to evaluate how the introduction of modifications in the consensus sequence could affect the affinity and specificity of binding [43]. Thus, the Sos peptide sequence was modified and twelve hybrid peptide–peptoid molecules were generated and tested in interaction assays based on perturbation of tryptophan fluorescence. The results pointed out that the N-Grb2-SH3 domain strongly prefers bulky aromatic side-chains in the consensus sequence [43]. Indeed, the best ligands obtained for N-Grb2 were those in which the proline -PXX**P**- was replaced by a N-dmb (dimethoxybenzyl) and N-dfb (difluorobenzyl) side chains (Fig. 2B) [43]. Instead, the presence of positively charged side-chains in the peptide/peptoids is strongly unfavoured, as indicated by the high K_d values which could be retrieved for binding to N-Grb2-SH3 [43]. Nevertheless, as concerning Crk-SH3, the introduction of various polar side chains seems to be considerably more tolerated [43].

Interestingly, this study shows how substitution of a single key proline residue can induce an increase in affinity greater than 100-fold and enhances specificity by roughly 300-fold [43].

It has also been reported that an alternative way to significantly increase the affinity for SH3 domains consists in the formation of rhodium(II) conjugates [46]. Indeed, the presence of the rhodium(II) transition metal favors the interactions with histidines or other Lewis-basic residues located close to the binding interface in SH3 domains [46]. In this way, the transition metal can offer support to the interaction between the ligand and the SH3 domain [46]. In a recent work, a multidisciplinary approach including ITC (Isothermal Titration Calorimetry), structural studies, kinase assays and ESI-MS (Electrospray Ionization-Mass Spectrometry) allowed to identify the S2ERh (i.e., substitution of a serine with a glutamic acid followed by the coordination of rhodium(II)) as the best modification to be introduced in the sequence -VSLARRPLPPLPN- that is recognized by the SH3 domain of Lyn (Fig. 2C) [46]. Indeed, the resulting metallo-peptide is characterized by a nanomolar affinity (6 nM) and by a considerable specificity for Lyn-SH3 with respect to the SH3 domains of Yes and Fyn proteins. Nevertheless, this peptide induces also activation of Lyn kinase activity (Fig. 2C) [46]. Therefore, the use of a metallo-peptide provides a chance to carry out ligand design driven by structural data and overcomes issues usually related to traditional inhibitors, such as, the lack of specificity among members of the same homologous protein family [46].

2.1.3 Drug discovery approaches: Small molecules

Small molecule modulators of SH3 domains mediated interactions have been reported in literature too.

The compound UCS15A (Fig. 2D) represents the first example of a non-peptide inhibitor of protein-protein associations in which SH3 domains are involved [47]. However, *in vitro* studies let speculate that UCS15A does not directly interact with the SH3 domains itself but, may instead bind proline rich domains [47]. Synthetic analogs of UCS15A were designed and evaluated as well [48]. A few of

them resulted more active than UCS15A in inhibiting the interaction between proline rich ligands and SH3 domains. One of the UCS15A derivatives, indicated as 2C (Fig. 2D), blocked in vitro the interaction between Sam68 and Fyn-SH3 as well as *in vivo* association of Grb2-SH3 with Sam68 and Sos1 [48]. MEK (MAPK (Mitogen-activated protein kinase)/Erk (Extracellular signal-regulated kinase)) kinase activation was also inhibited by compound 2C that resulted less cytotoxic than UCS15A and, following cell treatment, did not produce morphological variations [48].

Other studies were centered on the Abl protein tyrosine kinase as, selective compound agonists of Abl activity could in principle be implemented as chemical tools to better investigate the relevance of Abl kinase activity in solid tumors and following genotoxic stress [49]. The tyrosine kinase Abl is linked to several types of cancers [50]. From a structural point of view, Abl presents a certain similarity with the Src-family kinases. In fact, in the inactive/down-regulated form, the reverse-side of the kinase domain is wrapped by the SH3 and SH2 domains [50]. Inhibition of kinase function is linked to intramolecular contacts between the SH3 domain and the linker region connecting the SH2 and kinase domains [50]. Modulating allosterically Abl kinase activity through small molecule interactors of the non-catalytic domain may represent a route to find selective active compounds with therapeutic potentials [49]. With this in mind a screening strategy based on FPA (Fluorescence Polarization Assays) was set up by implementing recombinant Abl protein (including Ncap-SH3-SH2-linker), a synthetic polyproline containing peptide specific for Abl and, a collection of 1200 FDA approved drugs [49]. This screening let to identify dipyridamole (Fig. 2D) as an inhibitor of the interactions mediated by the SH3 domain of Abl [49]. Dipyridamole works as an agonist of Abl function as it enhances the kinase activity of down-regulated Abl *in vitro*. Nevertheless, molecular dynamics simulations and docking studies supported the hypothesis that dipyridamole perturbs the SH3/linker interface in Abl [49].

In addition, previous studies revealed that 2-aminoquinolines may potentially be implemented for the development of potent small molecule interactors of SH3 domains [51]. A structure-based ligand design approach was set up using the SH3 domain of the mouse Tec kinase as a model system and led to the discovery of a few simple heterocyclic small molecules as selective interactors of Tec-SH3 [51]. Tec belongs to a family of intracellular tyrosine kinases that are largely expressed in hematopoietic tissues and possibly act in growth and differentiation mechanisms in immune cells like B and T lymphocytes and monocytes [51]. NMR techniques coupled to mutagenesis and structure-activity relationship studies were used to characterize the binding of these compounds at the proline-rich peptide binding site. The most potent 2-aminoquinoline derivative, interacted with Tec-SH3 with a K_d = 125 µM (compound 2 in Fig. 2D) and competed for binding with a proline-rich peptide [51]. However, 6-substitued-2-aminoquinolines (compounds 27, 33, 38 and 39 in Fig. 2D) resulted better Tec-SH3 ligands provided with largely enhanced affinities (up to 6-fold) and specificities (Fig. 2D) [51].

In summary, during the last years several efforts have been focused to find SH3 ligands able to inhibit or somehow modulate crucial interactions with polyproline rich segments which could be linked to pathological conditions.
Some success has been obtained in term of improved selectivity and enhanced affinity by implementing peptoids in which the proline is replaced by different N-substituted moieties [43, 44] as well as metallo-peptides [46]. Small non peptide compounds against SH3 domains have been also explored. It appears clear that targeting SH3 domains rather than catalytic domains is a valuable route to overcome the specificity challenge.

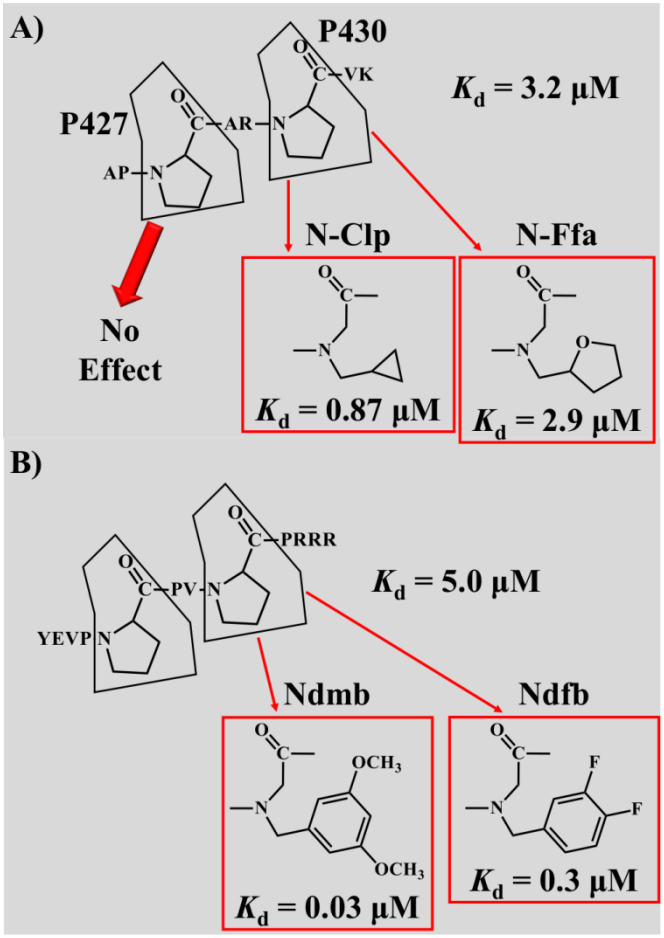


Fig. (**2**). A) Chemical structure of the peptide recognized by hOSF-SH3 domain and examples of chemical modifications that can be introduced to improve the binding affinity. The starting sequence (APPARPVK) is shown on top and the two key prolines are highlighted by black boxes. Chemical modifications, which induce increase of binding affinity, are indicated in the red rectangles [44]. B) Chemical structure of the wild-type peptide (YEVPPPVPPRRR) recognized by the SH3 domain of Grb2. The two key prolines are highlighted by black boxes; side-chains substitutions producing a significant enhancement of binding affinity are highlighted by red rectangles [43].

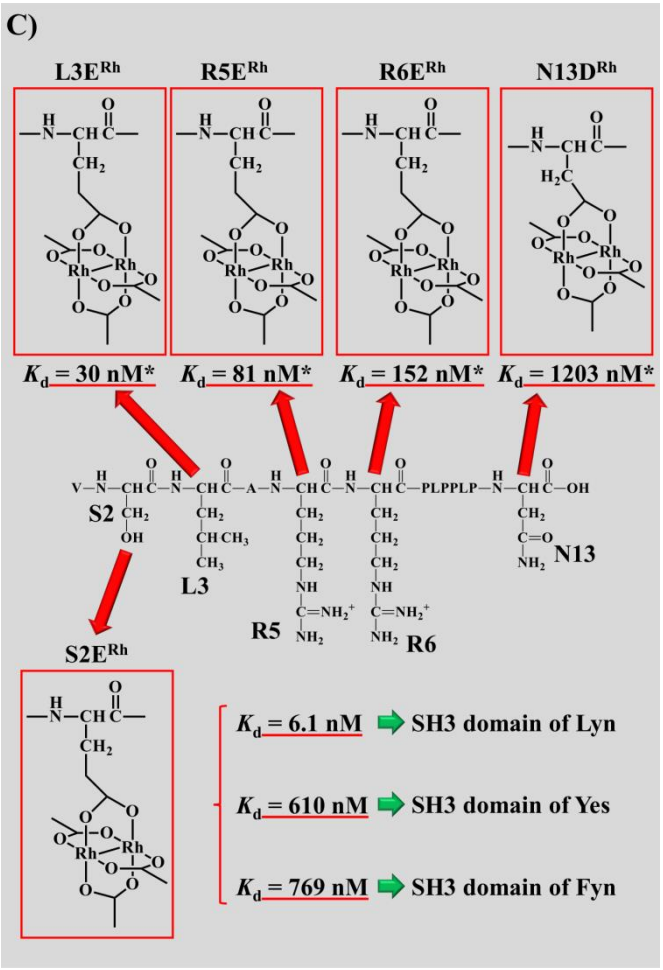


Fig. (**2**). C) Chemical structure of the starting peptide sequence (VSLARRPLPPLPN) implemented to evaluate the effect of rhodium(II) conjugates on the binding affinity against the SH3 domains of Lyn, Yes and Fyn. Introduced chemical substitutions are indicated in the red rectangles. The "*" refer to K_D values related to the Lyn-SH3 domain [46].

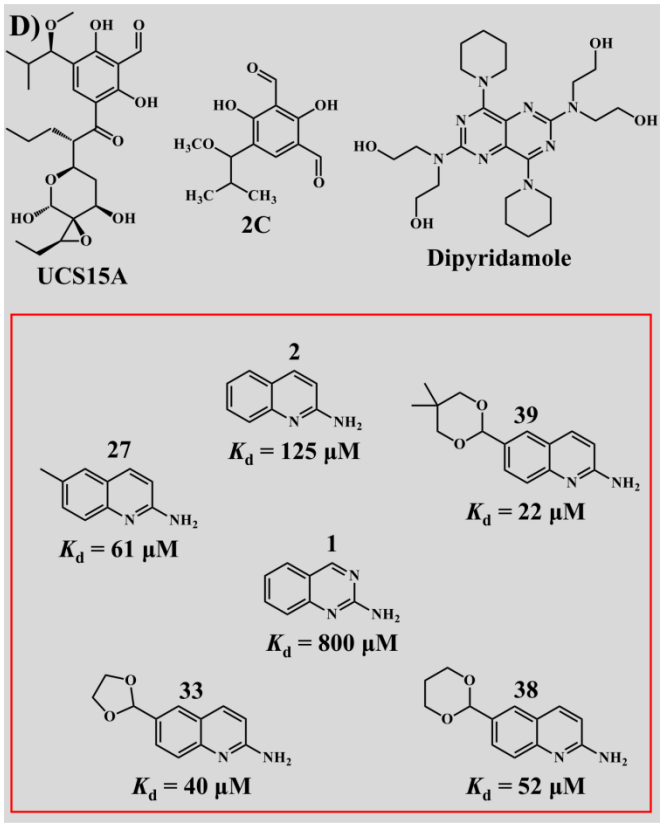


Fig. (**2**). D) Chemical structure of small molecules targeting SH3 domains/proline rich peptides interactions [47-49]. Ligands of the Tek-SH3 domain based on a 2-aminoquinoline scaffold and relative dissociation constant values are reported in the lower panel inside the red rectangle [51].

2.2. EVH1 (Enabled/VASP Homology 1) domains

EVH1 domains represent a group of modules composed by 110-115 residues which were firstly described as N-terminal homologous regions of proteins from the Ena / VASP (Drosophila Enabled / VAsodilator-Stimulated Phosphoprotein) family [13]. Nowadays it is known that EVH1 domains can be found also in other protein families (WASP, Homer/Vesl and Spred) [13, 14]. EVH1 domains play a fundamental role as regulators of signal transduction events, re-organizers of the actin cytoskeleton and modulators of actin dynamics and actin-based motility [27].

2.2.1 Structural features and interaction properties

Interestingly, all members of the four protein families - mentioned above- share the location of the EVH1 domain at the N-terminal end but, differ in the sequence and domain composition of the regions positioned C-terminally with respect to the EVH1 domain [14]. Furthermore, all EVH1 domains belong to the PH (Pleckstrin Homology) domain-like superfamily whose members are characterized by a fold which consists of two N-terminal anti-parallel β-sheets oriented perpendicularly to each other and an α-helix at the C-terminus (Fig. 3) [14].

EVH1 domains differ from the other PRDs like SH3, WW and GYF, for the shape of the binding cleft and consequently, the surface recognized on the PPII helix structure in polyproline ligands [14]. Indeed, the binding pocket of SH3, WW and GYF domains are characterized by a relative flatness whereas, that of EVH1 domains consists of

a wedge-shaped hydrophobic groove [14]. Furthermore, interactions between prism-shaped polyproline helices and SH3, WW and GYF domains occur through one face of the "prism" whereas, in EVH1 domains involve one apex of the triangular PPII helix in the ligand [14].

As introduced in the previous paragraphs, the binding cleft of a considerable number of PRDs is characterized by the presence of aromatic residues whose side chains show high complementarities with the crests and grooves located on the PPII helix surface in the ligands [14]. As concerning EVH1 domains, the faces of the binding cavity are formed by the side chains of aromatic residues and other conserved amino acids which are located in five positions distributed among the β1, β2 and β6 strands (Fig. 3) [14]. Intriguingly, the β2 strand of all EVH1 domains possesses a tryptophan residue which forms a ledge in the middle of the binding pockets and thus affects the orientation of the consensus sequence in the interaction partner (Fig. 3) [14]. Furthermore, the amino acidic composition of the other four positions changes in the different subfamilies of EVH1 domains. Indeed, Homer proteins have four conserved residues which occupy always the same positions (phenylalanine in the first position, a tyrosine in the second position, an isoleucine in the third position and a glutamine in the last position). The amino-acid type in the third position of Homer proteins is shared by all the members of the other subfamilies [14]. Instead, residues in the first, second and fourth positions of the binding pocket in Ena/VASP proteins are a methionine, a tyrosine and a glutamine, respectively [14]. The Spred family members have a binding pocket with a methionine in the first position, an arginine in the second position and a histidine in the last position [14]. As concerning WASP proteins, the first position of the binding pocket is occupied by a tyrosine whereas, the second and fourth positions are occupied by less conserved alanine and threonine residues, respectively [14].

The Ena/VASP family includes different proteins such as Ena (protein Enabled) and its counterpart in the mammals Mena (Mammalian Ena), VASP (VAsodilator-Stimulated Phosphoprotein) and Evl (Ena/VASP-like protein) [14, 27]. The members of this protein family are characterized by the presence of a N-terminal EVH1 domain along with a central proline rich region and a C-terminal EVH2 domain; different family members have proline rich and EVH2 domains of variable dimensions [14]. Indeed, the number of amino acids contained in the EVH2 domain can vary from approximately 151 residues, as in the case of VASP family, to approximately 190 residues, as in the Ena proteins [13]. It seems that Ena/VASP proteins are responsible for recruitment of polymerization competent profilin-actin complexes to their proline-rich interactors and thus for the enhancement (or acceleration) of the actin filament formation and reorganization processes [27]. More in detail, the central low complexity region of Ena/VASP proteins possesses a GPPPPP consensus sequence which is engaged by profilin-actin whereas, the cytoskeleton-associated partners involved in the actin filament formation (e.g., zyxin, vinculin and the listerial ActA protein) contain a (F/L)PPPP consensus motif which is recognized by the highly conserved N-terminal EVH1 domain of Ena/VASP proteins (Fig. 3A) [14, 27].

On the contrary Homer/Vesl proteins are not clearly connected to the actin assembly machinery, but, they are abundant in the neuronal tissue [27]. Intriguingly, the EVH1 domains of these proteins seem to act on the long-term potentiation in excitatory synapses and consequently on memory formation [27]. More in detail, these protein modules bind proline-rich sequences located in their biological partners (e.g., group I mGluRs (metabotropic Glutamate Receptors), IP3Rs (Inositol-1,4,5-trisphosphate Receptors), RyRs (Ryanodine Receptors) and the Shank family proteins) [52, 53]. For instance, IP3R, mGluRs and Shanks proteins possess three different proline-rich sequences (PPKKFRD, ALTPPSPFRD and PPXXF, respectively) which are recognized by the EVH1 domains of the Homer/Vesl proteins (Fig. 3B) [52, 54].

The three members of the WASP family (i.e., WASP, its homolog N-WASP (neuronal-WASP) and the yeast homolog Bee1p) are multi-domain proteins characterized by an EVH1 domain at the N-terminus, a basic peptide region, a GTPase binding module, a proline rich region and a verprolin-cofilin acidic motif [14]. The members of this group interact with extended consensus motifs which contain a LPPPEP core (Fig. 3C) and are longer than the consensus motifs recognized by Ena/VASP (Fig. 3A) and Homer/Vesl proteins (Fig. 3B) [14, 27].

As concerning the Spred family, its components contain an EVH1 module at the N-terminus (Fig. 3D), a c-Kit binding module in the central region and a Sprouty-like cysteine-rich domain at the C-terminus [14].

Interestingly, EVH1 domains have been divided in two classes (i.e., class I and class II) depending on the recognized polyproline consensus sequence [27]. EVH1 domains of the first class (e.g., Ena/VASP proteins) specifically interact with FPXΦP-containing sequences (where X and Φ correspond to any residue and hydrophobic amino acid, respectively) whereas those of the second class (e.g., Homer/Vesl proteins) recognize PPXXF-containing sequences [27]. Interestingly, both classes of EVH1 domains have a hydrophobic cleft formed by the side chains of two exposed conserved residues (i.e., W23 (Vasp sequence numbering) / W24 (Homer1a sequence numbering) and F74 (Homer1a sequence numbering) / F78 or F79 (Vasp sequence numbering)) which are oriented perpendicularly with respect to each other; the second C-terminal proline of the class I and class II consensus sequences inserts into this binding pocket (Fig. 3) [27, 55]. Nevertheless, class I EVH1 domains are characterized also by a second hydrophobic pocket formed by W23 and by a third aromatic residue (i.e., Y16 (Vasp sequence numbering)) which is conserved only among members of this group, the first N-terminal proline in the consensus motif of class I can be well located into this second cleft [27]. The lack of preferences for a precise amino acid type in the third position of the consensus sequences of class I rises from the fact that this residue is not able to interact with surface exposed residues of EVH1 domains due to its particular orientation in the cleft [27]. On the contrary, there's a preference for a hydrophobic residue in the fourth position as it is involved in weak contacts with the EVH1 surface [27].

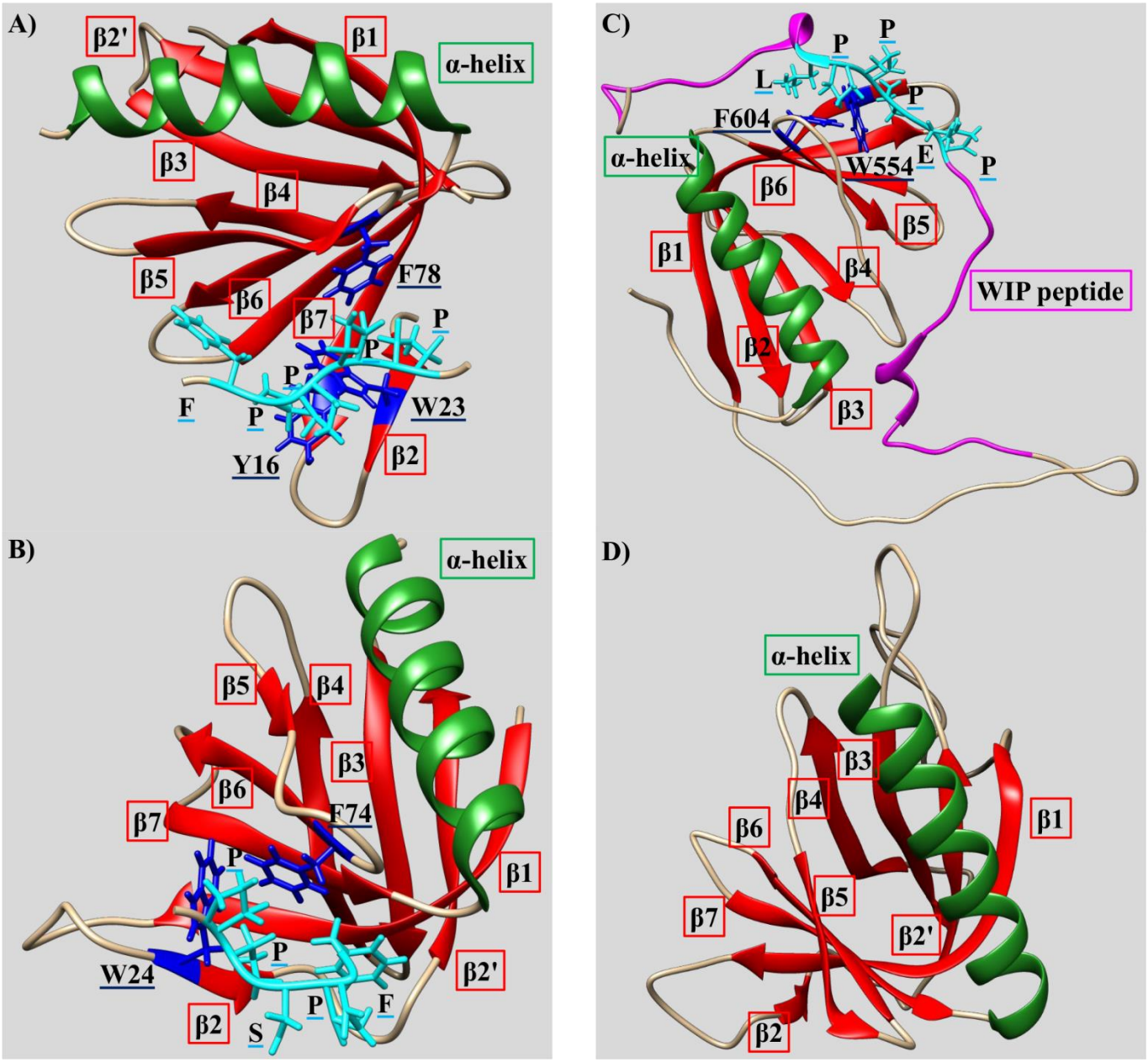


Fig. (**3**). Structures of the four subfamilies of EVH1 domains. A) X-ray structure of the EVH1 domain of ENA/VASP-like protein in complex with the ACTA peptide (FEFPPPPTDEE, PDB code: 1QC6 [55], only chains A and C are shown and only the -FPPPP- peptide fragment can be seen in the crystals, Table S2). The strands in the anti-parallel β-sheets (β1 (Q4-D17), β2 (K22-P25), β2' (S34-N41), β3 (T46-V52), β4 (V59-S64), β5 (K70-T75), β6 (F78-R82) and β7 (V87-F92)) are colored red whereas, the α-helix (S94-M112) at the C-terminus is reported in green. The residues, whose side chains form the two hydrophobic clefts (Y16, W23 and F78) are shown in blue. The residues of the consensus motif (F/L)PPPP are colored cyan. B) X-ray structure of the Homer EVH1 domain in complex with mGluR peptide (TPPSPF, PDB code: 1DDV [56]). The strands of the anti-parallel β-sheets (β1 (F7-I16), β2 (W24-P26), β2' (V32-Y38), β3 (V44-D51), β4 (K54-T60), β5 (F67-T68), β6 (F74-D79) and β7 (T84-G89)) are colored red whereas, the α-helix (S92-E108) at the C-terminus is shown in green. The residues, whose side chains form the hydrophobic cleft (W24 and F74) are shown in blue. The residues of the consensus motif PPSPF are reported in cyan.

Fig. (**3**). C) NMR structure of a fusion construct made up of the EVH1 domain of N-WASP and a WIP peptide (residues 451-485 ESRFYFHPISDLPPPEPYVYTTKSYPSKLARNESR) (PDB code: 2IFS [57], conformer n.1, Table S2). The strands of the β-sheets forming the canonical anti-parallel β-sandwich (β1 (C535-D549), β2 (M553-D567), β3 (S572-D579), β4 (W587-E590), β5 (F604-A608) and β6 (Q613-N617)) are colored red whereas, the α-helix (N620-R638) is in green. Residues colored in magenta represent the regions of the WIP peptide flanking the poly-proline motif LPPPEP, that is shown in cyan. The residues, whose side chains contribute the hydrophobic cleft (W554 and F604) are shown in blue. D) Solution structure of the EVH1 domain from the human Spred2 protein (PDB code: 2JP2 [58], conformer n.1). The anti-parallel β-sandwich (β1 (Y14-R25), β2 (F33-P34), β2' (S41-C46), β3 (F59-R65), β4 (V71-V77), β5 (Y83-N87), β6 (F90-K94) and β7 (K99-F104)) is colored red whereas, the α-helix (S106-E126) is colored green.

2.2.2 Drug Discovery approaches

The attention on the interactions between PRDs and their PRSs comes from their fundamental role in the regulation of many important signaling cascades [25].

For instance, EVH1 domains of Ena/VASP proteins are known for their regulatory function of the actin cytoskeleton, which plays a pivotal role in cell migration [59]. Cell migration in turn is critical in disease-relevant processes like tumor metastasis [59]. Indeed, Mena, a member of Ena/VASP protein subfamily, is highly up-regulated in invasive mammary tumor cells collected *in vivo*; nevertheless, in breast cancer patients high levels of Mena have been correlated to poor clinical outcomes [60]. These observations let speculate a potential therapeutic implementation of peptide or small molecule modulators of the interaction between Ena/VASP EVH1 domains and their consensus motifs [60].

One of the strategies used to target protein/protein interactions mediated by the Mena EVH1 domain (Mena-EVH1) consists in the design of peptides mimicking the discontinuous epitope characterizing the binding site of Mena for proline rich segments.

Mena-EVH1 binds the Acta surface protein from Listeria monocytogenes and in fact, a peptide derived from Acta (i.e., Ac-DFPPPPTDEEL-NH_2, where Ac= N-terminal acetylation and NH_2= C-terminal amidation) interacts with Mena-EVH1 with high affinity (dissociation constant K_d= 5 µM) with respect to a shorter form (Ac-FPPPPT-NH_2, K_d= 602 µM) [61]. The core of the consensus sequence (i.e., PPPP) is recognized by a triad of aromatic residues (Y16, W23 and F27) whereas, the phenylalanine in the peptide ligand fits in a cleft of Mena-EVH1 close to this aromatic patch [61]. Furthermore, the affinity of Mena-EVH1 for Acta depends also on a group of basic residues which are close to the aromatic triad and interact with acidic amino acids which flank the PPPP core sequence in Acta [61]. Based on these evidences, two different discontinuous segments of Mena-EVH1, containing residues involved in binding to Acta, named Peptide A (sequence: VMVYDDANKKWVPA) and Peptide B (sequence: YNQATQTFHQWR), were chosen as starting point to generate a small library of compounds modulators of Mena-EVH1 mediated interactions [61]. The two segments A and B were first synthetized in linear and cyclic versions; next ligation points (i.e., a cysteine and a BrAc (bromoacetyl) moieties in Peptide A and B, respectively) were inserted at either the N- or the C-terminal ends (Fig. 4A) [61]. In addition, three different spacer elements (i.e., Ahx (ε-Aminohexanoic acid), lysine and Dap (α,β-Diamino propionic acid)) were introduced between the ligation points and the A and B peptides in order to obtain different combinations of the two sequences together and modulate the spatial distance in between them (Fig. 4A) [61]. A small library of thirty peptides, simulating the binding cleft of Mena-EVH1, was thus assembled and implemented in competition-type binding assays. Through these assays, the capacity of each peptide to interfere with formation of the complex between GST (Glutathione S-Transferase)-Mena-EVH1 and the interactor pGolemi was evaluated [61]. The results revealed that a compound (P11) made up of a cyclic version of Peptide A with a cysteine at the C-terminus and a linear Peptide B portion -prepared with a BrAc at the N-terminus and a Ahx spacer- presented the highest ability to inhibit the binding of pGolemi to Mena-EVH1 (Fig. 4A) [61].

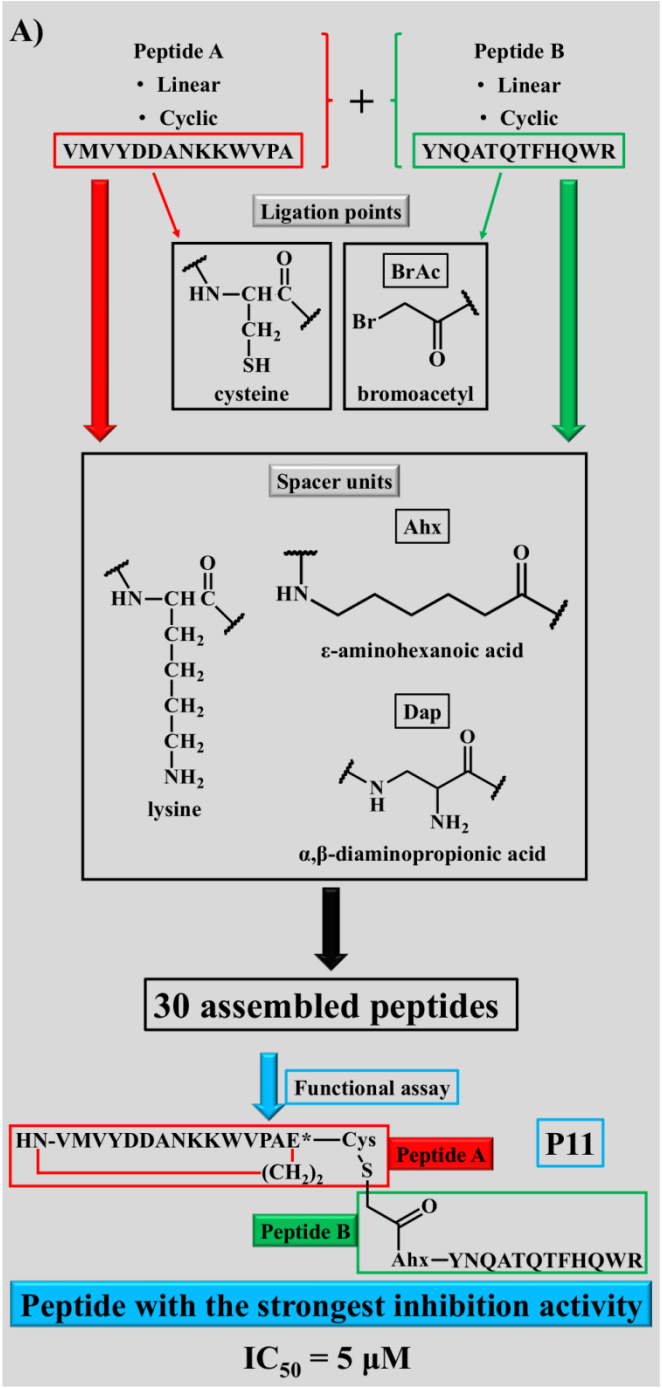


Fig. (**4**). Possible strategies to develop novel inhibitors of protein-protein interactions mediated by EVH1 domains. A) Design of compounds mimicking the discontinuous binding epitope of Mena-EVH1 for polyproline ligands [61]. "*" Cyclic sequences contain an additional glutamic acid to favor synthetic procedures.

An alternative route to design peptidomimetics able to target the interaction between EVH1 domains and their binding partners has been described in a recent work by Opitz and collaborators [25]. The key elements of this strategy are the core of the consensus sequence recognized by EVH1 domains along with conformationally restricted small-

molecule fragments mimicking somehow di-proline motifs [25]. Medicinal chemistry efforts were directed towards optimization of the peptide Ac-SFE-FPPPP-TEDEL-NH_2, (the polyproline consensus sequence is underlined, compound 1 in Fig. 4B) encompassing a fragment of the Acta protein from *Listeria monocytogenes* that is recognized by the EVH1 domains of Ena/VASP proteins contained in host cells [25]. The core motif -FPPPP- was firstly modified by introducing different hydrophobic groups in the ortho position of the phenylalanine (Fig. 4B) [25]. ITC (Isothermal Titration Calorimetry) and FT (Fluorescence Titration) studies revealed that the 2-Cl-F (2-chloro-*L*-phenylalanine) group induced the highest increase in affinity against Ena/Vasp EVH1 domain (Fig. 4B) [25]. The two couples of prolines contained in the consensus sequence were next substituted by different small molecule moieties (named ProM-1, ProM-2, ProM-3 and ProM-4) in order to evaluate the chance to further increase binding affinity for EVH1 domains but also eventually improve hydrolase resistance and maintain a certain cell permeability [25]. Interestingly, replacement of the first couple of prolines by ProM-2 and substitution of the second couple by ProM-1 led to a peptidomimetic with a considerably enhanced affinity with respect to the starting peptide (compound 4 in Fig. 4B) [25]. As already mentioned above, removal of residues flanking the core of the consensus motif favors a decrease in interaction affinity, (compounds 1a in Fig. 4B). Instead, the same truncation in the modified sequence improved ligand efficiency (LE = ΔG°/number of heavy atoms), a parameter related to the efficacy by which a molecule uses its structural features to bind targets (compound 4 in Fig. 4B) [25, 62]. X-ray studies revealed also that molecule 4 fits the binding cleft of the EVH1 domain with a structural topology almost identical to that adopted by the core of the consensus sequence -FPPPP- [25]. Therefore, the inserted chemical modifications, including incorporation of the ProM-1 and ProM-2 scaffolds, enhance the binding properties of the ligand and induce in its bound state a conformation similar to the PPII helical conformation observed for the consensus sequence [25].

Compound 4 resulted poorly cell permeable, however, this undesirable characteristic was considerably attenuated by esterification of the C-terminal COOH group (See compound **5** in Fig. 4B) [25]. Indeed, cellular assays on highly invasive breast cancer cells (i.e., MDA-MB231) showed that the resulting compound 5 was possibly influencing remodeling of the actin cytoskeleton, inducing delocalization of VASP from leading edge and from focal adhesions and thus decreasing the invasion power of the cells (Fig. 4B) [25].

In conclusion, the examples described here highlight the possibility to identify molecules, which can pass cell barriers, target the EVH1 domains of Ena/VASP proteins and block the invasion of cancer cells. Similar strategies can be adopted also to target diverse PRDs involved in diseases in order to find a lead compound with interesting drug-like properties.

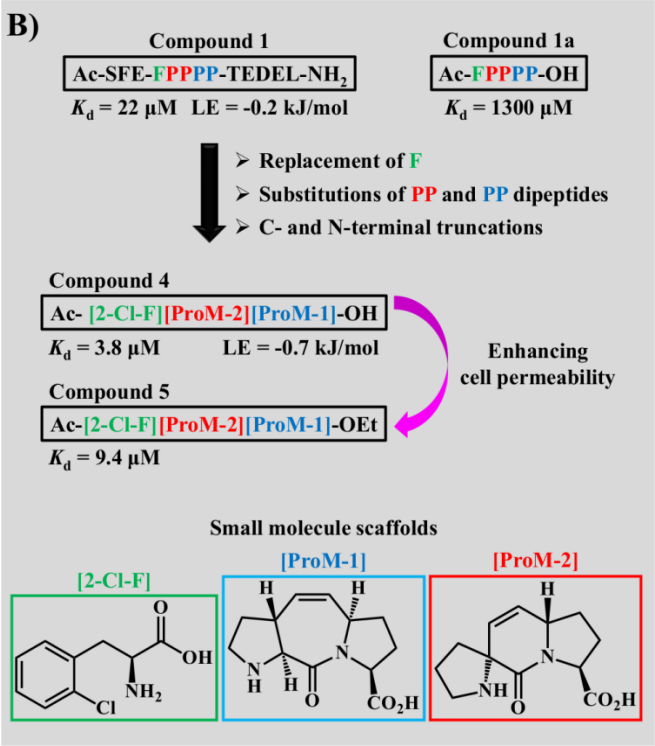


Fig. (**4**). B) Design of cell permeable peptidomimetic inhibitors of EVH1 domains starting from a proline-rich sequence of the Acta protein from *Listeria monocytogenes*. Chemical modifications introduced to improve binding affinity and ligand efficiency are indicated. An ethoxy (OEt) moiety at the C-terminus improves cell permeability (See compound 5). K_d values were measured in ITC experiments, compounds were evaluated against VASP-EVH1 [25].

2.3. GYF (Glycine-Tyrosine-Phenylalanine) domain

The GYF domain owns this name to the presence of a conserved GYF amino acid motif in the binding pocket, and it was originally found in CD2BP2 (CD2 (Cluster of Differentiation 2) antigen cytoplasmic tail-Binding Protein 2) as the module responsible for recognition of proline-rich peptide repeats (i.e., PPPGHR) contained in the T cell adhesion molecule CD2 [63].

2.3.1 Structural features and interaction properties

There are two different groups of GYF domains, whose names derive by the proteins in which they were originally discovered, CD2BP2 and SMY2 (Suppressor of MYosin 2). These two groups share the characteristic conserved sequence GPF-X_4-(M/V/I)-X_2-W-X_3-GYF, where X_2, X_3 and X_4 indicate sequences made up of 2, 3 and 4 residues of any type, respectively, organized in a bulge-helix-bulge structure, and a β1-β2 loop (Fig. 5) [64]. The components of the first group have a role in the process of mRNA splicing whereas, the components of the second group play an unclear function [65]. GYF domains from both subfamilies possess a compact fold and a core region containing the ligand binding site and in which one α-helix (CD2BP2 group) or two α-helices (SMY2 group) pack against a small β-sheet and the side chains of bulky aromatic residues locate themselves in available space between the two types of secondary structure elements (Fig. 5). Outside this core region different GYF domains present C-terminal portions with diverse structural

elements [63, 64]; in fact, CD2BP2 GYF domains are characterized by the presence of 3_{10} helical turns (Fig. 5A) whereas, SMY2 GYF domains possess an additional α-helix at the C-terminus (Fig. 5B) [63, 64, 66]. Furthermore, members of the CD2BP2 group possess a characteristic tryptophan residue working as lid in the binding pocket allowing stacking interactions with bulky residues in the ligand including Trp and/or Arg and in fact, GYF domains of this group recognize two classes of consensus motifs: PPGW and PPG-X-(R/K) [63, 64]. Members of the SMY2 family are characterized by a shorter β1-β2 loop, and an aspartate (D22) instead of the W (characteristic of GYF domains from CD2BP2 class). This aspartate is involved in a hydrogen bond network that moves it away from the binding pocket and allows interaction with hydrophobic ligands containing the consensus core PPGΦ, where Φ represents a hydrophobic amino acid with a moderate size, in fact, the relatively small binding site dimension does not favor the presence of a Trp in the recognition motif differently from the CD2BP2 GYF subgroup (Fig. 5B) [63, 64].

2.3.2 Drug Discovery approaches

Interestingly, an over-expression of GYF-containing proteins has been found in chronic fatigue syndrome and in estradiol-, 4-hydroxytamoxifen- and acolbifene-treated T47D breast cancer cells letting speculate that GYF domains may be target in drug discovery [67].
Furthermore, the p.Arg610Gly mutation in the GYF domain of GIGYF2 (GRB10 (GRowth factor receptor-Bound protein 10)-interacting GYF protein 2), a protein linked to insulin/IGF-1 (Insulin-like Growth Factor-1) signaling pathway in the central nervous system, has been associated to Parkinson's disease and is known to disrupt ligand binding ability of the GYF domain in GIGYF2 [68].
However, not many studies centered on the design of peptide or small molecule modulators of GYF domains have been reported in literature at the best of our knowledge.

Peptide interactors of the GYF domains of CD2BP2 and the consensus motif PPG(W/F/Y/M/L) were discovered by a phage display approach using a X9 and a X2PPPX3 peptide libraries [69].

Nevertheless, another interesting work reported on the identification of high affinity peptide ligands of GYF domains from CD2BP2 and PERQ2. These two domains were targeted with libraries of peptides containing a florescent probe at either their N- and C-terminal ends [70]. Focused libraries were designed starting from the peptide ligand of CD2BP2-GYF (-EFGPPPGWLGR-) and two ligands of PERQ2-GYF (-FNGSPPGLSRD and WRPGPPPPPPPGLV-). Fluorescence polarization assays were conducted to investigate binding of peptides to GYF domains along with NMR interaction assays with ^{15}N labelled GYF domains [70]. The results revealed that peptides with a fluorescein moiety attached at the N-terminal extremity had improved interaction affinity for GYF domains (Fig. 5C) [70]. Furthermore, introduction of the unnatural amino acid *p*-benzoylphenylalanine (Bpa), a fragment usually used to enhance the sensitivity of the fluorescence-based assay, between the N-terminal fluorescein moiety and the peptide sequence, led to a molecule with a low micromolar affinity for the CD2BP2 GYF domain (Fig. 5C) [70].

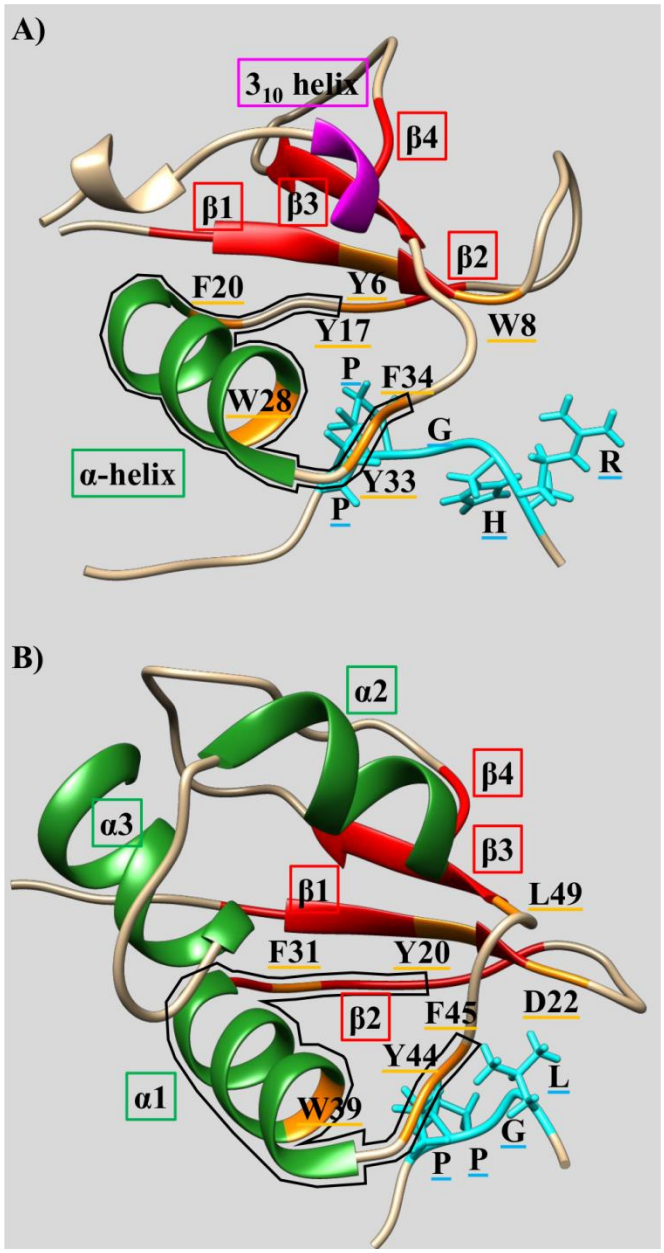


Fig. (**5**). Structures of GYF domains belonging to two different subgroups. A) NMR structure of CD2BP2-GYF in complex with a proline-rich peptide from CD2 (1-SHRPPPPGHRV-11, PDB code: 1L2Z [71], conformer n.1, Table S3). The strands in the anti-parallel β-sheet (β1 (M3-K7), β2 (L16-Y17), β3 (Y39-K42) and β4 (Y51-N52)) are colored red, the α-helix is shown in green. The bulge-helix-bulge motif is highlighted by a black box. Residues contributing the binding surface (Y6, W8, Y17, F20, W28, Y33 and F34) are highlighted in orange. Amino acids in the consensus PPGX motif of the peptide ligand are colored cyan. The peptide binds in an extended conformation with the segment 4-7 involved in a PPII helix. B) X-ray structure of SMY2-GYF in complex with a proline-rich peptide from BBP (Branchpoint Binding Protein)/ScSF1 (Saccharomyces cerevisiae Splicing Factor 1) (SSIAPPPGLSG, PDB code: 3FMA [64], only chains A and L are shown, Table S3). The strands in the anti-parallel β-sheet (β1 (S17-I21), β2 (I27-T32), β3 (Q50-R53) and β4 (I68-T69)) are reported in red and, the α-helices (α1, α2 and α3) in green. The bulge-helix-bulge structure is indicated by a black box. The residues, whose side chains form the binding surface (Y20, D22, F31, W39, Y44, F45 and L49) are shown in orange. The residues of the consensus ligand motif PPGΦ are shown in cyan.

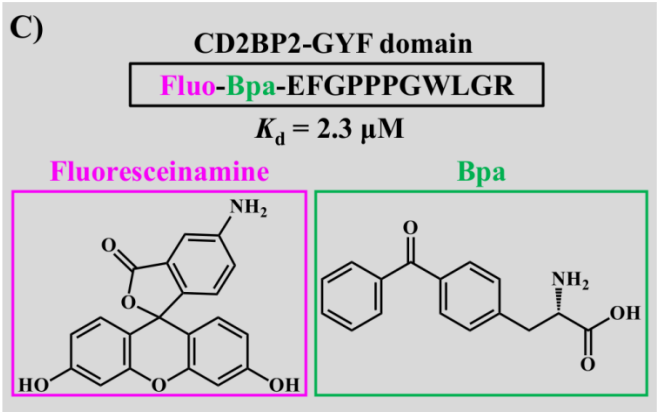


Fig. (5). C) Peptidomimetic compound targeting the GYF domain of CD2BP2 [70].

2.4. UEV (Ubiquitin E2 Variant) domain

The UEV module belongs to the group of ubiquitin binding domains and is linked to protein ubiquitination [16]. This process plays a fundamental role in the regulation of many different biological events such as protein degradation and quality control, endocytosis, vesicular trafficking, cell-cycle control, stress response, DNA repair, growth-factor signaling, transcription and gene silencing [16]. The ubiquitination process consists in conjugation of the C-terminal end of a single ubiquitin unit (or a polyubiquitin chain) to specific lysines in the target proteins and involves the action of different enzymes: E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme) and E3 (ubiquitin protein ligase) [16].

2.4.1 The UEV fold

UEV domains share some common structural features with the E2 enzymes, however, differently from E2, they are inactive, as the catalytic cysteine residue is missing thus, they are unable to catalyze ubiquitin transfer; both UEV and E2 domains interact with ubiquitin but, only UEV has been thus far described as a binder of proline rich segments [72]. UEV domains mainly share with the canonical enzymes the characteristic α/β/loop/α fold (i.e., packing of four helices towards one face of an antiparallel β-sheet composed by four strands) but a difference is represented by the very frequent absence of the two C-terminal α-helices (Fig. 6A) [72, 73].

2.4.2 Drug Discovery Approaches

A considerable attention has been focused on the human Tsg101 (Tumor susceptibility gene 101) protein because of its fundamental role in the budding process of the enveloped human immunodeficiency (HIV) and Ebola viruses [73, 74]. Interestingly, the UEV domain of Tsg101 recognizes the P(S/T)AP consensus sequences (also named "late domain") contained in the p6 region of two matrix structural proteins (Gag for HIV and Vp40 for Ebola) [73]. In detail, Tsg101-UEV binds p6 gag of HIV-1 with a dissociation constant K_d= 27 ± 5 µM, this value becomes lower (K_d= 4.3 ± 1.6 µM) in presence of low salts and acidic conditions [73]. A nine residue long peptide, spanning the PTAP motif 5-PEPTAPPEE-13, interacts with the UEV domain with the same affinity of the complete full length p6 gag HIV-1 protein [73]. Structural studies of the complex between PTAP peptide and Tsg101-UEV have been conducted [72, 75]. The "late domain" motif fits in a bifurcated cleft formed by the β2/β3 hairpin, the N-terminal third of the vestigial active site loop (name due to its analogy with the region containing the active site in the canonical E2 enzymes) and residues located at the C-terminus of the UEV domain (Fig. 6A) [72, 73]. The first proline of the consensus sequence (**P**TAP) inserts in a shallow pocket defined by P71 and the methyl groups of T58, T92 and M95 (from UEV domain) whereas, the couple AP of the consensus motif fits a distinct deeper cleft in which the alanine residue makes interactions with the methyl and methylene groups of I70, M95, V141 and S143 (from the UEV domain) (Fig. 6A). Instead, the last C-terminal proline (i.e., PTA**P**) from "late domain" motif inserts in a pocket formed by the aromatic rings of Y63 and Y68 and by P139 and V141 (Fig. 6A) [72, 73]. Intriguingly, it seems that the AP dipeptide plays the most important role in the recognition of the consensus sequence by the UEV domain and has some similarity with the x-Pro motif in the proline rich ligands of WW and SH3 domains [72, 73].

A potential valuable route to develop novel antiviral therapies consists in the inhibition of the interaction of Tsg101 UEV domain with the "late domain" motif of the HIV p6 gag protein. To achieve this goal, one of the strategies, that is reported in literature, relies on lanthipeptides, which represent a large set of RiPPs (Ribosomally-synthesized and Post-translationally modified Peptides) produced enzymatically in *Escherichia coli*, with a bicyclic organization and intramolecular thioether bridges [74]. These peptides derive first by dehydration of threonine or serine residues and next, by cyclization of cysteines into this dehydrated amino acids [74]. These lanthipeptides were meant to incorporate all amino acids except S, T, and C in order to avoid the risk of additional reactions and undesired ring closures [74]. On the contrary, negatively charged amino acids were inserted to simulate the features of the p6 region containing the "late domain" motif (-PEPTAPPEE-) [74]. The biosynthetic procedure produced molecules provided with a highly conserved prochlorosin leader peptide which is recognized by the protein ProcM, that is responsible for the dehydration and cyclization on diverse core regions (Fig. 6B) [74]. Among members of the library, one peptide named XY3-3 was identified as inhibitor of the interaction between Tsg101-UEV and HIV p6 protein. XY3-3 was evaluated *in vitro* by ELISA (Enzyme-Linked ImmunoSorbent Assay), MST (MicroScale Thermophoresis) and FP (Fluorescence Polarization) assays [74]. The dissociation constant for the XY3-3 interaction with UEV domain resulted equal to 16 µM in MST experiments and further confirmed by FPA (Fig. 6B). Furthermore, removal of the leader peptide led to a compound (named "PepClev" in Fig. 6B) with an improved binding affinity (K_d= 4 µM by MST and K_d= 5.5 µM by FP assay with a peptide containing a fluorescein-isothiocyanate (FITC) moiety at the N-terminus) even respect to the p6 peptide considered as reference (K_d= 16.6 µM) (Fig. 6B) [74]. Cell-based assays showed that "PepClev" was not able to affect the gag-mediated budding of virus-like particles, thus suggesting the lack of a proper cell permeability [74]. Replacement of the N-terminal glycine by a cysteine and the labelling of the peptide with a Tat sequence led to the analogue -named XY3-3-Tat- which can reduce the viral budding from HEK293T cells by ~65% at 100 nM concentration (Fig. 6B)

[74]. Instead, the Tat moiety alone does not induce any effect at 25 μM concentration, thus suggesting that the inhibition action of XY3-3-Tat depends exclusively on the core peptide region [74].

A similar study was conducted on HEV (Hepatitis E virus) which is responsible for a self-limiting acute viral hepatitis in normal individuals [76]. Binding of the UEV domain of Tsg101 to the PSAP motif of the viral ORF3 (Open Reading Frame 3) protein is crucial in HEV infection. Indeed, this interaction is involved in the release of genotype 3 HEV [76]. Intriguingly, molecular dynamics studies revealed that the gag-PTAP (from HIV) and ORF3-PSAP (from HEV) motifs insert in the same binding cleft in Tsg101-UEV. CP11 (cyclic peptide 11, CGWIYWNV) is the most effective molecule against HIV release as it blocks the association of Tsg101-UEV and gag-PTAP, thus, it was evaluated also as inhibitor of the budding of HEV (Fig. 6C) [76]. Further MD analyses showed that CP11 could bind the Tsg101-UEV in the same binding pocket of gag-PTAP (from HIV) and ORF3-PSAP (from HEV) motifs (Fig. 6C) [76].
Cell-based assays clearly showed that CP11 effectively blocks the interaction between the ORF3-PSAP motif and Tsg101-UEV. Interestingly, the CP11 amino acid sequence is rather different from the P(S/T)AP motifs thus pointing out that this specific sequence may be dispensable for recognition by UEV domains [76]. Interestingly, CP11 is able to lower the release of both genotype 1 and genotype 3 HEV by approximately 90% with an IC_{50} (50% inhibitory concentration) value equal to 2 μM [76].
In conclusion, CP11 can be considered a promising starting point for the development of antiviral agents specific for HEV. However, further deep studies are required to overcome the challenge of more severe conditions associated with the presence of HEV in pregnant women and immunocompromised patients [76].

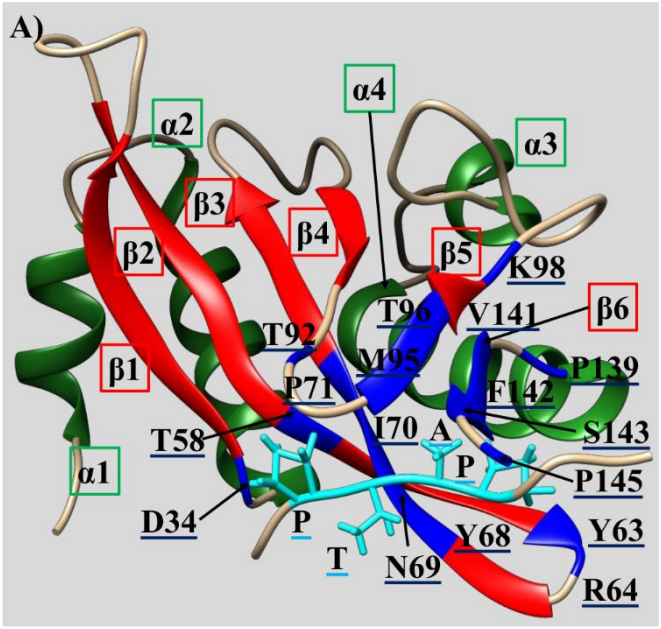


Fig. (**6**). A) X-ray structure of the Tsg101-UEV in complex with a PTAP peptide from HIV-1 (5-PEPTAPPEE-13, PDB code: 3OBU [75], Table S4). The β-strands (β1 (L35-S41), β2 (E51-Y63), β3 (N66-L76), β4 (I86-V89), β5 (M95-I97) and β6 (V141-S143)) are reported in red whereas, the α-helices (α1 (S4-V12), α2 (Y17-Y32), α3 (L111-E116) and α4 (D123-E138)) in green. The residues in blue (D34, T58, Y63, R64, Y68, N69, I70, P71, T92, M95, T96, K98**,** P139, V141, F142, S143, P145) are the ones whose side chains contribute the peptide binding cleft. The amino acids of the peptide consensus interaction motif P(S/T)AP are shown in cyan.

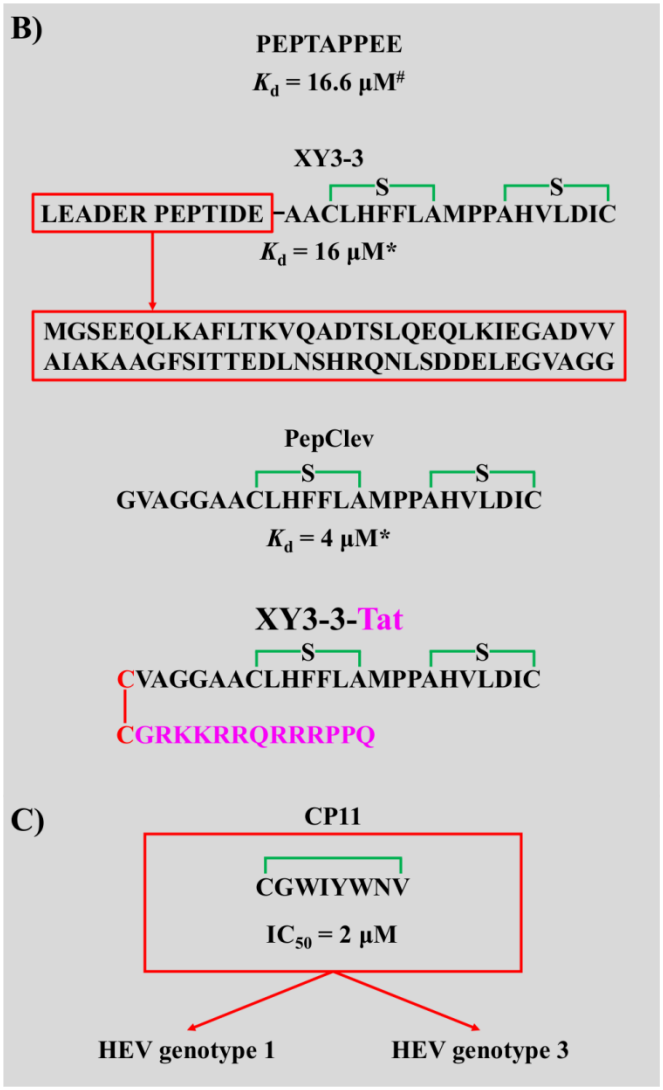


Fig. (**6**). B) Examples of lanthipeptide inhibitors of the interaction between Tsg101-UEV and the “late domain” motif of the HIV p6 Gag protein [74]. The "#" indicates a value obtained for the fluorescein tagged PTAP nonapeptide through FPA. The "*" points to MST measurements. The residues belonging to the Tat sequence (GRKKRRQRRRPPQ) are colored magenta. The cysteines, used to insert the Tat sequence, are highlighted in red. C) Example of an inhibitor of the interaction between Tsg101-UEV and the PSAP motif of the ORF3 protein from HEV [76].

.

3. LIM (LIN11, ISL-1 & MEC-3) DOMAIN

The LIM module owes its name to the initial letters of three proteins from *Caenorhabditis elegans* and rat in which the LIM homeodomain -initially described as a cysteine rich region- was firstly discovered (i.e., Lin-1 (a cell lineage protein), Isl1 (Insulin enhancer binding protein 1), and MEC-3 (MEChanosensory neuron differentiation protein 3)) [77-79]. Proteins containing LIM domains play roles in different events of nucleus and cytoplasm such as regulation of actin

structure and dynamics, neuronal pathfinding, integrin-dependent adhesion and signaling, cell-fate determination and tissue-specific gene expression [79].

3.1. Structural features and interaction properties

All LIM domains share the following sequence composed by two zinc finger units: [C-X_2-C-X_{16-23}-H-X_2-C] and [C-X_2-C-X_{16-21}-C-X_{2-3}-C/D/H], where X is any amino acid, separated by a two-residues (i.e., X_2) spacer (Fig. 7A) [77, 79]. Interestingly, the spacer results to be invariant in length and seems to be fundamental for LIM-domain function [77, 79]. Furthermore, the first zinc finger unit possesses a zinc ion coordinated by three cysteines and one histidine whereas, the second unit has a zinc ion that may be coordinated in three different ways (i.e., four cysteines, three cysteines and a histidine, three cysteines and an aspartic acid) [79, 80].

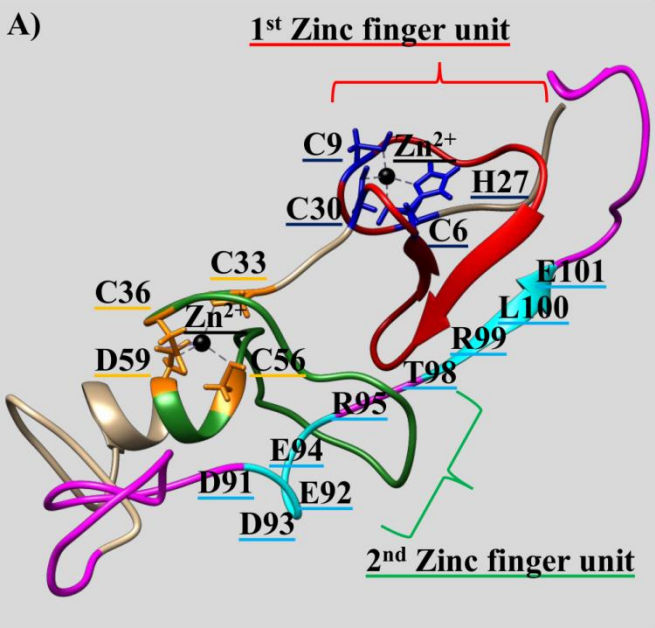


Fig. (7). A) NMR structure of the N-terminal LIM domain of LMO2 (LIM domain Only 2) fused with Ldb1 (LIM domain binding protein 1)-LID (LIM Interacting Domain) (PDB code: 1J2O [81], conformer n.1, Table S5). The first zinc finger unit (residues from C6 to C30) is reported in red whereas, the second zinc finger unit (residues from C33 to D59) is reported in green. The two zinc ions are represented as black spheres. Residues which coordinate the zinc ions of the first and second zinc fingers are colored blue and orange, respectively. The LID region of Ldb1 (D75-E115) is highlighted in magenta with the residues contacted by LMO2-LIM colored cyan (i.e., D91-R95, T98-E101).

From a structural point of view, each zinc finger is characterized by two antiparallel β-hairpins which are perpendicularly wrapped to each other [79]. The first zinc finger motif contains β-hairpin1 and 2, the second one comprises β-hairpin3 and 4 [79]. Furthermore, the second zinc finger unit possesses also a short α-helix at the C-terminal end [79] (Fig. 7A). In addition, there are rubredoxin-type zinc knuckles between the shorter strands of β-hairpin1 and β-hairpin3 and tight turns between the longer strands of β-hairpin2 and β-hairpin4 [79]. Intriguingly, both secondary structure and tertiary fold result to be strongly dependent on the conserved tetrahedral coordination of zinc ions in the zinc finger units whereas, the packing of the two zinc fingers depends on a considerable number of hydrophobic core residues [79].

Although different structural studies described the binding interfaces between LIM domains and other protein partners, differently from other PIDs like the SH2 and the SH3, it is difficult to identify a precise interaction motif common to all LIM domains and further studies are required to better define this aspect [79, 81].

3.2. LIM family proteins and different classifications

The LIM module is specific of eukaryotes, it is well represented in the animal kingdom and also has a certain relevance in plants [77]. Canonical sequences of LIM domains can be different among diverse species [77]. Two subfamilies of plant LIMs are known: DA1/DAR and 2LIM, which contain one and two LIM modules respectively [77]. The structure of the simplest form of LIM domain in plants looks like that of CRPs (Cysteine-Rich Proteins) from animals [77]. Instead, LIM domains motifs in plants possess some variations with respect to those in animals [77]. LIM domains found in plants have two different residues (histidine in 2LIM or cysteine in DA1/DAR) in the last position of the domain, whereas, some CRPs in animals are characterized by the presence of an aspartate as last amino acid of the sequence [77]. Furthermore, LIM domains in plants have a longer C-terminus and less frequently the glycine-rich region (i.e., GRR) after each zinc finger unit [77].

Proteins with LIM domains have been grouped into four families according to their functional domains and localization [82, 83]. For instance, the LHX (LIM Homeobox) and nuclear LMO (LIM only protein) proteins are characterized by two tandem N-terminal LIM-domains and ~~are localized~~ reside in the nucleus [82-84]. Members of this group work as transcription factors or cofactors [82, 84]. The second group is composed by LIM domains only proteins that can be found in the nucleus and cytoplasm; CRP (C-reactive Protein), FHL (Four and Half LIM) and PINCH (Particularly Interesting New Cysteine-Histidine rich protein) protein families belong to this group [82-84]. Members of this subfamily possess two or more tandem LIM domains assembled at the N- or C-terminal ends [82, 84]. Instead, there are examples of LIM proteins (e.g., Zyxin, EPLIN (Epithelial Protein Lost In Neoplasm), and ABLIM (Actin-Binding LIM) protein families)) in which the LIM domain is accompanied by different protein-protein interaction modules such as the PDZ, leucine-aspartate repeats, or actin-target domains [82, 84]. These proteins can be found in both cytoplasm and nucleus [82-84]. The last two groups (in some cases defined as one single group) of LIM domains is composed by LIMK (LIM domain Kinases) and MICAL (Microtubule Associated Monooxygenase, Calponin and LIM domain containing) proteins and possess supplementary functional modules such as mono-oxygenase or kinase catalytic motifs [82-84]. Intriguingly, LIM domain proteins located in the nucleus are involved in gene regulation and cell fate determination whereas, those

positioned in the cytoplasm have a role mainly linked to cytoskeleton organization [84].

Lim proteins can be also classified according to the number of LIM domains they include [85]. Indeed, there is the single LIM domain family which is characterized by only one LIM domain and includes different proteins such as NRAP (Nebulin-Related Anchoring Protein), LASPs 1,2 (LIM and SH3 Proteins), LMO7 (LIM domain-Only protein 7) and PDLIMs 1,3 (PDZ and LIM domains) [85]. Instead the family of proteins with two LIM domains comprise ISL1, CRPs 1,2 (Cysteine and glycine-Rich Proteins), hhLIM (human heart LIM), LIMKs 1,2 (LIM-kinases), LMCD1 (LIM and cysteine-rich domain protein 1), MLP (Muscle LIM Protein), LMO2 (LIM-Only protein 2), and LHX9 (LIM HomeoboX protein 9) [85]. The three LIM domain family is composed by Migfilin, PDLIMs 5,7 (PDZ and LIM domain proteins), RILP (REST (RE1-Silencing Transcription factor)-interacting LIM domain Protein), ZASP (Z-band Alternatively Spliced PDZ-motif), Zyxin and TRIP6 (Thyroid Receptor Interacting Protein 6) [85]. The members of the four LIM family (ABLIM3 (Actin-Binding LIM 3), PXN (PaXilliN) and HIC5 (Hydrogen peroxide-Inducible Clone 5)) are characterized by four LIM domains [85]. Instead, the FHL (Four-and-a-Half LIM) family includes different proteins characterized by a single half LIM domain located at the N-terminus followed by four complete LIM domains [85]. The last group (i.e., five LIM family) is composed by proteins with five LIM domains and comprises the LIMSs (LIM and Senescent cell antigen-like-containing domain proteins), also known as PINCHs [85].

3.3. Lim domains as potential targets in drug discovery

The attention on LIM domains is due to their roles in cardiovascular diseases (e.g., cardiac hypertrophy and cardiomyopathy), fibrosis disease, different types of cancers (including but not limited to T-ALL (T cell Acute Lymphoblastic Leukaemia), Neuroblastoma and breast cancer) [86-89].

Among the proteins with LIM domains, the members of LMO and LIMK families have gathered much interest for their role in cancer initiation and/or progression but also in several other diseases such as Alzheimer's, schizophrenia, neurofibromatosis type 2, psoriatic epidermal lesions, primary pulmonary hypertension, allergic diseases, ocular hypertension and glaucoma, erectile dysfunction, HIV and other viral infections [86, 90]. For instance, LMO1 and LMO2 are over-expressed in patients affected by T-ALL and possible play oncogenic functions whereas, LMO3 and LMO4 are identified as a neuroblastoma-associated oncogene and a human breast cancer auto-antigen associated with poor prognosis, respectively [86]. Therefore, a few efforts were made to develop anticancer therapeutic routes based on targeting LIM proteins [86]. As concerning LMO2, two strategies were already proposed and described as valuable ways for generating treatments against cancer [17]. One of these strategies is based on a single chain antibody which specifically binds to LMO2 and is described as promising instrument for dissecting the role of LMO2 in haematopoiesis and leukaemia and to develop a therapeutic against T-ALL induced by LMO2 [17, 91]. The other strategy consists in the development of a short peptide aptamer characterized by a C-X-X-C motif and able to specifically interact with LMO2 and prevent its action as a T-cell oncogene in a mouse transplantation experiment [17, 92].

A considerable amount of LIMK inhibitors have already been reported [90]. The two LIMK1 and LIMK2 isoforms share a common domains organization, including two LIM domains at the N-terminus, contiguous PDZ and proline/serine rich domains, and a kinase module [90]. The N-terminal LIM domains bind to the kinase domain working as negative modulators of kinase activity [90]. Compound LX7101 inhibits LIMK1 with an IC50 comprised between 32 and 134 nM and LIMK2 with an IC50 value between 1 and 7.3 nM [90] (Fig. 7B). Upon topical application LX7101 induces a decrease in the intraocular hypertension and can be implemented for treating glaucoma [90, 93]. Another valuable example is provided by Pyridocarbazolone, a molecule able to inhibit LIMK1 and LIMK2 with IC50 values equal to 50 nM and 75 nM, respectively (Fig. 7B) [90]. Although this molecule is able to block breast tumor growth and can lower the metastatic load in a mouse model, it is also characterized by a certain toxicity against different cell lines [90].

In this contest, more studies need to be conducted to find out ligands of LIM domains of LIMKs that could be further evaluated to understand their potentials as therapeutics or eventually could work as instruments to better comprehend the role of this domain within kinase signaling.

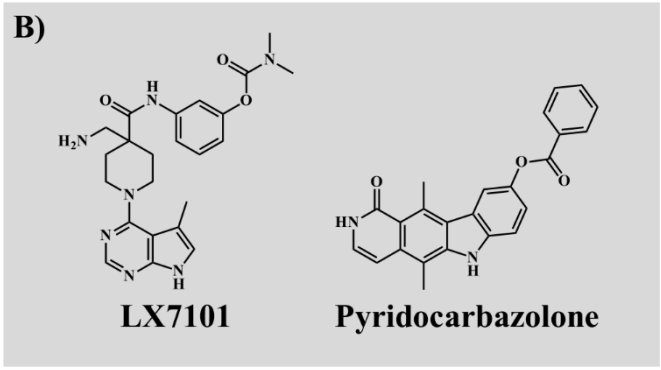


Fig. (**7**). B) Examples of inhibitors which target LIMKs.

4. CH (CALPONIN HOMOLOGY) DOMAIN

CH is a domain composed by ~110 residues, originally discovered in the actin-binding protein calponin but, that can be actually found in cytoskeletal and signal-transduction proteins [18]. This module occurs as both single and multiple copies. For instance, the presence of a single CH unit characterizes proteins involved in the regulation of muscle contraction (e.g., calponin and SM22-α (Smooth Muscle protein 22-alpha)) but, can be also found in different signal-transduction proteins such as Vav and IQGAPs (GTPase-activating proteins that contain calmodulin-binding IQ motifs) [18]. Instead, two CH domains in tandem represent the actin-binding module of different proteins responsible for

cross-linking actin filaments with consequent formation of meshworks or bundles. Examples are given by α-actinin, dystrophin, ABP-120 (Actin-Binding Protein 120), fimbrin and cortexillin [18]. These proteins have in common the presence of CH domains but differ in modular domain organization and in addition, many of them are characterized by the presence of a calmodulin-like domain with EF-hand motifs [18]. The nature of the formed actin types of structures (i.e., meshworks or bundles) depends on the number of repetitive modules which can separate the CH domains from the EF-hand motifs [18].

4.1. Classifications of CH containing proteins

Proteins with CH domains can be grouped in different subfamilies, "1CH", "2CH", "3CH" and "4CH" [11]. The "1CH" group is characterized by one CH domain and includes calponin and different signaling proteins (e.g., Vav (name deriving from the sixth letter of the Hebrew alphabet), IQGAP and Cdc24 (Cell division control protein 24)) [11, 94]. Although a single CH domain is sufficient to bind actin, an entirely functional F-actin binding domain (also named F-ABD) requires at least two CH domains in tandem [11, 94]. Proteins of the "2CH" group have two CH domains in tandem [94]; instead, members of the "3CH" and "4CH" groups are characterized by three or four CH domains, respectively [11, 94]. Members of the last "4CH" subfamily (e.g., the fimbrin/plastin proteins) are monomeric actin cross-linking molecules containing a tandem of two ABDs [11, 94]. In addition, CH domains have also been classified based on the degree of sequence similarity [11]. Indeed, there are three main classes of CH domains: the CH1 and CH2 ABD forming domains and the CH3 module that is generally found in "1CH" proteins [11].
Nevertheless, there are monomeric F-actin interacting proteins (e.g., dystrophin and utrophin) and dimeric cross-linking proteins (like α-actinin, β-spectrin and filamin) each with a single ABD [94].

4.2. Structural features and interaction properties

From a structural point of view, a single CH domain is characterized by a core made up of four α-helices with three of them composing a loose triple helix bundle (Fig. 8A) [11]. Furthermore, there are short α-helices (in a number ranging from one to three) which are located in the loops between the core α-helices [11]. Intriguingly, the structures of tandem CH domains differ from each other for different structural elements such as the core helices lengths, number and location of the secondary helices [95]. Despite these differences, the core structures of different CH domains possess significantly similar features [95].

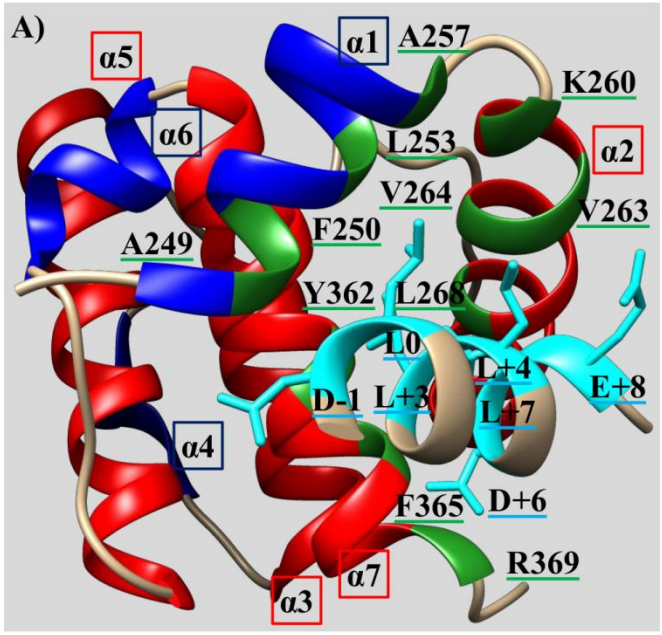


Fig. (**8**). A) X-ray structure of the C-terminal CH domain of α-parvin in complex with the paxillin LD1 (Leucine Rich Domain 1) motif (MDDLDALLADLESTTSHISK, PDB code: 2VZD [96], chains A and C, Table S6). The four core α-helices (α2 (K260-N280), α3 (G293-G305), α5 (S319-G338) and α7 (D353-R369)) are reported in red, additional helical regions (α1 (D248-A257), 3_{10} helix 4 (P309-F313) and α6 (R345-N351)) are shown in blue. The residues in green (A249, F250, L253, A257, K260, V263, V264, L268, Y362, F365 and R369) are the ones whose side chains contribute the peptide binding cleft. The amino acids D-1, L0, L+3, L+4, D+6, L+7, E+8 (numbered with respect to the first residue of the 0-LDXLLXXL-+7 consensus sequence) of the peptide ligand are colored cyan.

An interesting example of proteins provided with a CH domain is represented by the focal adhesion protein α-parvin [96]. α-Parvin CH domain interacts with the consensus sequence 0-LDXLLXXL-7 (the sequence number of the first N-terminal leucine residue of the consensus sequence is set to zero) that can be found in the LD (leucine-rich) regions (specifically LD1, LD2 and LD4) of its binding partner paxillin [96]. More in detail, the side chains of four α-parvin residues (A249, F250, L253, A257) from the α1 helix, three amino acids (V263, V264, and L268) from the α2 helix and two residues (Y362 and F365) from the C-terminal α7 helix form a hydrophobic patch that makes contacts with conserved leucines in the positions 0, +3, +4, and +7 of the consensus binding sequence of LD regions (Fig. 8A) [96]. In addition, there are two positively charged parvin residues (K260 and R369) that can provide additional electrostatic contacts (Fig. 8A) [96]. In detail, K260 and R369 interact with aspartic acids in the position +1 of LD2 and LD4 consensus sequences whereas, make contacts with the aspartic and glutamic acid residues in the positions +8 and +6, respectively of the LD1 consensus motif (Fig. 8A) [96].

Intriguingly, mutagenesis and deletion studies on different CH containing proteins -like α-actinin, dystrophin, filamin, and fimbrin- revealed that there are three principal F-actin binding sites (ABS1,2,3 in Fig. 8B) [97].

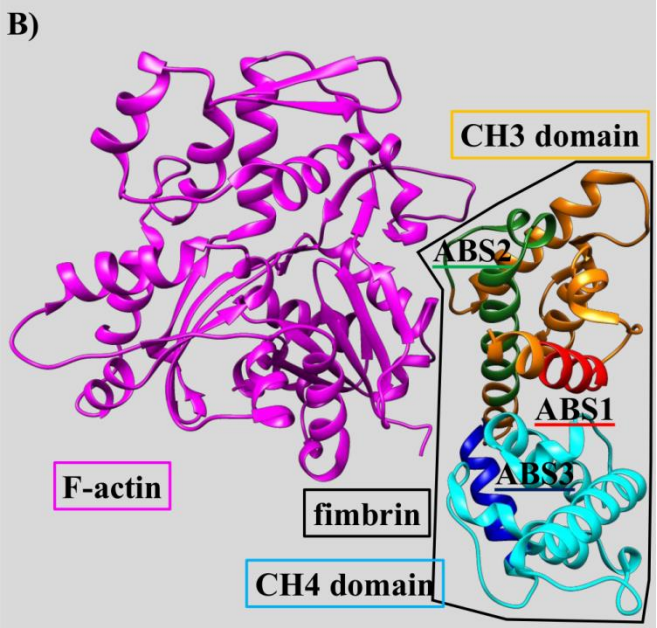


Fig. (**8**). B) High-resolution cryo-EM structure of the F-actin in complex with the ABD2 of fimbrin protein (PDB code: 3BYH [98]). The F-actin is reported in magenta. The fimbrin fragment encompassing the tandem CH3,4 domains is included in the black block. The residues of the CH3 domain (E390-K503) are shown in orange whereas, those of the CH4 domain (E513-L621) in cyan. Residues forming the ABS1 (E395-L405), ABS2 (L470-M498) and ABS3 (E513-V527) [99] are colored red, green and blue, respectively.

4.3. CH domains as potential therapeutic targets

CH domains hold a certain interest as they can be correlated to different diseases [100] like muscular dystrophy [101], otopalatodigital syndromes, which include X-linked pathology involving structures like the craniofacial, the skeletal, the brain, the visceral, and the urogenital ones [102], renal disorders like glomerulosclerosis [103] and a type of hemolytic anemias known as spherocytosis [104]. Generally mutations in CH domains of different proteins have been related to the above mentioned diseases [100].
Although a considerable amount of structural information on CH domains is already available, molecules able to inhibit the functions of CH domains and thus to be used as novel potential therapeutic agents -at the best of our knowledge- are still not available.

5. EH (EPS15 HOMOLOGY) DOMAIN

The EH domain was firstly identified as N-terminal region with three repetitive units in Eps15 (Epidermal growth factor receptor substrate 15), a protein target for the phosphorylation operated by the EGF receptor [105]. To date it is known that EH domains can be found in a large number of proteins belonging to various kingdoms ranging from yeast to human [105]. EH domains are composed by ~100 residues and belong to proteins which are linked in most cases to endocytosis or vesicular transport [105].

There exists four EH domains (EHD) proteins in mammals, that work as modulators of specific steps of the endocytic transport [7, 106]. Among this mammalian family members of EHDs, EHD1 is the best characterized one and is involved in recycling of TfR (Transferrin Receptor) from the endocytic recycling compartment to the plasma membrane [106]. EHD2 has a less clear function but it seems to play a role in internalization of different receptors such as transferrin and GLUT4 (GLUcose Transporter type 4) [106]. Instead, EHD3 is linked to the EE (Early Endosome)-to-Golgi retrograde transport, late biosynthetic conveyance of lysosomal enzymes, maintenance of Golgi morphology [106]. Finally EHD4 seems to play functions related to receptor transport from EE to the ERC (Endocytic Recycling Compartment) and from EE to the lysosomal degradation route [106].
Furthermore, proteins with EH domains can be also linked to the regulation of other processes such as organization of actin cytoskeleton, mitogenic signaling, control of cell proliferation, modulation of nuclear shuttling and DNA repair [7].

5.1. Structural features and interaction properties

The interactions of EH domains with protein partners containing the -NPF- motif hold a key role in different neuronal processes such as synaptic vesicle cycle, internalization of NGF (Nerve Growth Factor), determination of neuronal cell fate, synapses development, trafficking of postsynaptic receptors [8, 107].
In major detail, EH domains recognize mainly three different consensus sequences known as “Class I peptides”, “Class II peptides” and “Class III peptides” [7]. Peptides of the first group (“Class I peptides”) are characterized by the NPF sequence that is the most common recognition sequence of EH domains [7, 105]. “Class II peptides” can be characterized by three different sequences (i.e., WW, FW or SGW) whereas, those of "Class III" possess a H(S/T)F motif and are recognized only by End3p, a yeast protein provided with one EH domain [7].

NMR studies of different EH domains (like EH1 of mouse Eps15, EH2 of human Eps15, EH3 of human Eps15, EH of POB1 (Partner Of RalBP1), and EH of Reps1 (RALBP1 Associated Eps Domain Containing 1)) have been conducted [19]. Interestingly, all these structures are characterized by two closely associated helix-loop-helix motifs (also named EF-hands) which are combined with a short antiparallel β-sheet (Fig. 9A) [19, 105, 108]. More in detail, the EH domain includes four helices (αA, αB, αC and αD), a short β-strand (βA) between αA and αB, a second β-strand (βB) between αC and αD whereas, a tight turn connects αB to αC (Fig. 9A) [108].
Furthermore, EH domains are characterized by the presence of a proline-rich portion located at the C-terminus [105]. Although EF-hands are known for their Ca^{2+}-binding properties, not all EH domains possess the entire set of residues necessary for the interaction with Ca^{2+} ions [105, 108].
As concerning protein-protein interactions, the NPF sequence of “Class I peptides” is completely buried in the binding cleft formed by the αB and αC helices of the helix-loop-helix motifs (Fig. 9A) [19].

For instance, NMR structural studies conducted on the EH2 domain of human Eps15 revealed that the bottom of the binding cleft is composed by three hydrophobic residues (L40, L50 and W54, residue numbers follow those of the PDB structure 1F8H) whereas, the edge of the binding pocket is formed by other four amino acids (G33, K37, V47 and G51) (Fig. 9A) [109, 110]. Furthermore, the residues of the edge contribute to the specificity in the interaction between the EH domain and the consensus motif in its binding partner [109].

In addition, the NP dipeptide motif and flanking residues of "Class I" interactors assume a type I β-turn conformation upon interaction with EH domains (Fig. 9A) [19]. Interestingly, four residues in the NPF peptide ligand make interactions with the EH domain, and in fact, the amino acid located at the +3 position with respect to asparagine (N), is generally involved into contacts with many residues of the binding pocket also contributing to the stabilization of the type I β-turn conformation [19]. On the contrary, an amino acid on the EH domains in position +3 with respect to a conserved Trp (i.e., Trp54 in EH2 from human Eps15) is responsible for the specificity of the different consensus recognition sequences [19]. Indeed, complexes between EH modules and NPF sequences are favored by the presence of an alanine or a serine in this position (+3) of the EH domain whereas, domains recognizing "Class II peptides" (i.e., FW, WW or SWG sequences) present slightly larger amino acids, like cysteine or valine, at the +3 site [19].

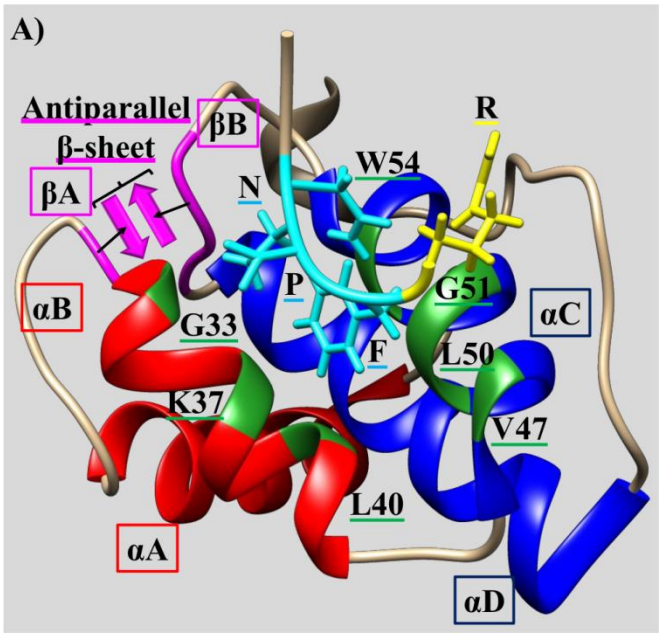


Fig. (**9**). A) NMR structure of the second EH domain of human Eps15 in complex with a "Class I peptide" (PTGSSSTNPFR, PDB code: 1F8H [109] conformer n.1, Table S7). The two α-helices from the first helix-loop-helix motif (αA (K10-D22) and αB (S32-N42)) are reported in red, those of the second helix-loop-helix motif (αC (P46-D58) and αD (D68-K83)) are shown in blue; residues belonging to the short β-strands (βA (F30-L31) and βB (G63-D66)) are colored magenta. G33, K37, L40, V47, L50, G51, W54, which contribute to the binding cleft, are colored green. The NPF consensus sequence in the peptide ligand is shown in cyan; the fourth peptide residue (R), which makes multiple contacts with the binding pocket, is reported in yellow.

5.2. Drug Discovery approaches

EH domains have acquired a considerable importance because of their role in receptors recycling. Two different mechanisms have been described for this process [111]. The first one is the "fast recycling" and consists in the renewing of material from the endosome directly to the membrane [111]. The second one is the "slow recycling" and represents a more regulated route linked to the sorting of material through a larger endocytic recycling compartment (ERC) [111]. Interestingly, EH domains are fundamental for the "slow recycling" of integrins and other receptors, a route that can be also exploited by a solid tumor to enhance invasion and metastatic mechanisms [111]. Indeed, impaired recycling of membrane receptors and decreased cancer invasiveness may be caused by inhibition of EHD1 function, including inactivation of its EH domain [111].

Therefore, it has been suggested that development of molecules able to block EHD1 function by inhibiting protein-protein interactions mediated by its EH domain, may represent a valuable therapeutic route to hinder invasion and metastasis of cancer cells [111].

To achieve this goal, one study reported on the design of peptide inhibitors of EH domains starting from the canonical -NPF- interaction sequence and the notion that multiple negatively charged amino acids in the region located C-terminally with respect to the -NPF- motif usually favor interactions with EH domains [112]. Therefore, linear peptides provided with the NPF core but also one or two additional C-terminal glutamates were synthesized [112]. Furthermore, a tyrosine residue was inserted at the N-terminal region in order to reproduce the features of endogenous peptide interactors of EHD1-EH (Fig. 9B) [112]. The presence of a tyrosine gives the additional advantage to allow a more precise evaluation of ligand concentrations by spectrophotometric measures [112].

These linear peptides were next cyclized through head-to-tail cyclization to favor formation of a β-turn like conformation, that usually characterizes the bound state of peptide ligands to EH domains [112]. The best cyclic peptide, which was derived from this study (i.e., cNPF1), binds EHD1-EH with a dissociation constant K_d= 16.8 μM, as estimated after linking it to a fluorescent probe and conducting direct fluorescence-based interaction assays (Fig. 9B).

This cyclic peptide, as clearly shown by structural studies, is as well characterized by a tight ensemble of well-structured conformations in aqueous solution with the NPF motif forming a β-turn with a structural topology very similar to that assumed by the same consensus sequence in linear NPF peptides in complex with EHD1-EH [112].

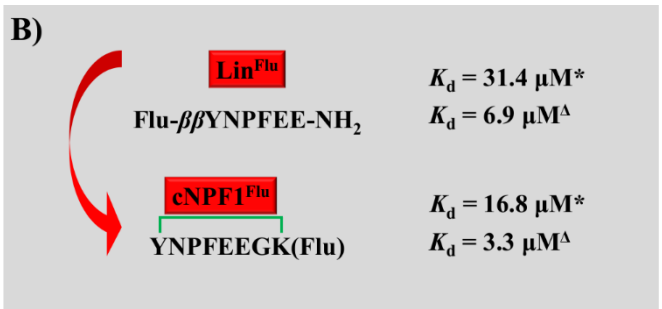


Fig. (**9**). B) Design of cyclic peptides targeting EHD1-EH [112]. Lin = linear; Flu = Fluorescein; NH_2= C-terminal amidation; "*"= values measured under low salt conditions (i.e., 15 mM NaC); "**Δ**"= values measured under physiological salt conditions (i.e., in presence of 150 mM NaCl).

Another recently proposed strategy is based on thioether-stapled macrocyclic inhibitors [111]. Several cyclic compounds targeting EH domains were designed starting from a linear peptide containing the core of the consensus sequence (- NPFE -) recognized by EHD1-EH. Different inhibitors were produced by using diverse N- and C-terminal moieties (See Fig. 9C). In details, firstly all peptide sequences with thiols at the two N- and C-extremities were generated to allow efficient cyclization by means of chemical reactions implementing α,α′-dibromoxylenes or other dibromomethyl-aryl linkers. A tyrosine and an additional glutamic acids were eventually added in a few compounds at the N- and C-termini to potentially enhance binding affinity (Fig. 9C) [111]. Finally, the peptides were converted into thioether-stapled macrocyclic inhibitors by dithiol bis-alkylation (Fig. 9C) [111].

Newly generated molecules were tested in FP (Fluorescence Polarization) competition-type assays by implementing the EH domain of EHD1 in complex with a known fluorescein conjugated cyclic peptide interactor (cNPF1) [111, 112] and also in direct binding experiments using the novel peptides linked to a fluorescent probe [111].

From these experiments the compound Flu-6-OX, provided with a penicillamine, resulted the best EHD1-EH ligand (Fig. 9C) [111]. However, all these newly synthesized molecules do not have relevant biological activities and are unable to affect the slow recycling pathways even if conjugated to the cell-penetrating peptide Tat [111]. These negative biological outcomes have been possibly correlated to a poor cell penetration, intracellular mislocalization or off-target effects [111].

These studies stress out the relevance of targeting EH domains in anticancer drug discovery field. However, much more work needs to be conducted to better comprehend the chances to attack protein-protein interactions mediated by EH domains with small molecules.

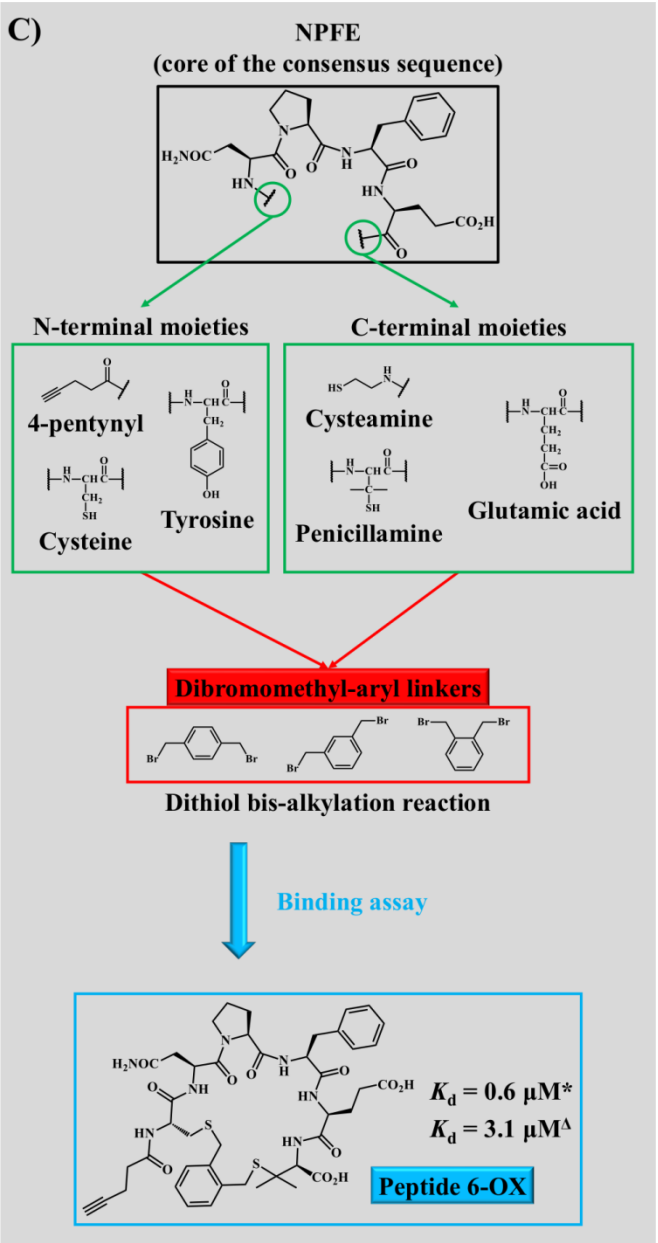


Fig. (**9**). C) Design of thioether-stapled macrocyclic inhibitors of EH domains starting from linear peptides containing the -NPFE- motif [111]. Peptide 6-OX in the cyan box represents the best EH domain ligand. K_d values were obtained through direct interaction assays by implementing a fluorescein-labeled "Peptide 6-OX". "*"= measures conducted under low salts conditions (i.e., in presence of 15 mM NaCl); "**Δ**"= measures conducted under physiological salts conditions (i.e., in presence of 150 mM NaCl).

6. PDZ (POST SYNAPTIC DENSITY PROTEIN 95 (PSD95), DROSOPHILA DISC LARGE TUMOR SUPPRESSOR (DLG1), AND ZONULA OCCLUDENS-1 PROTEIN (ZO-1)) DOMAIN

Proteins with PDZ domains represent a copious family whose members possess various biological functions [113]. This domain owes its name to the first letters of the proteins in which it was firstly found, i.e., the product of the *Drosophila dlg* (discs large) tumor suppressor gene, the ZO-

1 (Zonular Occludens-1) protein and the synaptic protein PSD-95 (also known as SAP90 (Synapse-Associated Protein 90)) [114-116].

6.1. Classifications of PDZ domain containing proteins

There are at least 250 PDZ domains which can be found in more than 150 proteins in humans.
It has been proposed to collect PDZ proteins into three main subfamilies according to different modular domain organization [113]. The first subfamily is composed by proteins without enzyme-like domains and with a number of PDZ domains ranging from two to more than ten [113]. For instance, GRIP (Glutamate Receptor-Interacting Protein) is an AMPA (α-Amino-3-hydroxy-5-Methyl-4-isoxazole-Propionic Acid) receptor which is characterized by the presence of seven PDZ units [113]. Instead, the second subfamily includes the MAGUKs (Membrane-Associated Guanylate Kinases) proteins which are made up of one or three PDZ domains accompanied by one SH3 domain and a GuK (Guanylate Kinase-like) domain. Examples of members of this group are: PSD-95, DLG1, and ZO-1 (i.e., the proteins giving the name to the PDZ domain) [113]. The third family is composed by elements characterized by a PDZ domain and one or more additional domains (like ankyrin, PTB (Phospho Tyrosine Binding), LRR (Leucine Rich Repeats), LIM, WW, C2 (protein kinase C conserved region 2), L27 (Lin-2 and Lin-7), PH (Pleckstrin Homology) and DEP (Dishevelled, Egl-10 and Pleckstrin)) [113].

It has also been proposed to collect PDZ proteins into diverse subgroups based on the organization properties [117]. Thus, a first group (termed "Class I") is composed by PDZ proteins deprived of a catalytic activity and that may or may not possess an enzyme-like domain [117, 118]. Examples of this group are given by InaD (Inactivation-no-after-potential D protein) and MUPP1 (Multi-PDZ domain Protein 1) [117]. Members of "Class II" (e.g., Tyrosine phosphatase 1E and PDZ-RGS3 (Regulator of G-protein Signaling 3) protein) are PDZ proteins with one or at most two catalytic domains [117, 118]. Instead, the last group (termed "Class III") is characterized by proteins with "PDZ-like" domains from plants, metazoan and bacteria, which possess a significant low sequence homology with the canonical PDZ domains [117-119]. A "PDZ-like" fold is characterized by five β-strands (β1-β5) capped by two α-helices (α2 and α3), similar to a canonical PDZ domain fold (as described below), but also by two additional short β-strands at the N- and C-termini (βN and βC, respectively), and a well-defined α-helix (α1) positioned in the space between the β1 and β2 loop [119]. The peculiarity of the members (e.g., HtrA (High temperature requirement protein A)) of this group consists in the frequent co-occurrence of the "PDZ-like" modules with several protease domains [117].

6.2. Structural features and interaction properties

The majority of PDZ proteins encompasses multiple copies of PDZ modules which usually bind to the C-terminus of their biological interactors [113]. Furthermore, they are described as model scaffolds involved in reversible interactions with different biological partners [113]. Therefore, PDZ proteins are responsible for a sort of dynamic coordination of signaling complexes formation and protein networking [113]. Indeed, different types of proteins -like transmembrane receptors, adhesion molecules, cytoskeleton and cytosolic enzymatic proteins- can be involved into protein-protein interactions through PDZ domain mediated complexes [113].

In addition, *in vivo* and *in vitro* binding assays revealed that a PDZ module can be involved in at least two distinct interaction mechanisms, one based on the recognition of specific sequences at the C-terminal ends of proteins and the other consisting in the dimerization with other PDZ modules [114]. For instance, three related proteins (i.e., PSD-95, chapsyn 110 and the human homolog of the *Drosophila Dlg* protein) interact with the C-terminal ends of NMDA (N-Methyl D-Aspartate) and the Shaker-type potassium channels through their PDZ1 and PDZ2 domains [114, 120]. As concerning PDZ hetero-dimerization, a valuable example is given by the PSD95-PDZ2, a module which binds to the PDZ domains from nNOS (neuronal Nitric Oxide Synthase) and SNTA1 (α-1-syntrophin, a member of the muscle-cell cortical proteins linked to the dystrophin complex) [114].

PDZ domains play a fundamental role as scaffolding elements in the nervous system and the binding of PDZ modules to their binding partners is responsible for regulation of crucial neurophysiological mechanisms [113]. For instance, interactions between PSD-95 and the consensus motifs in the C-terminal ends of NMDA receptor and Shaker type K^+ channels are crucial to ensure synaptic plasticity [113]. Other valuable examples are given by MAST2 (Microtubule-Associated Serine/Threonine-protein kinase 2) and NHERF (Na^+/H^+ Exchanger Regulatory Factor), whose PDZ domains recognize a consensus motif located in PTEN (Phosphatase and Tensin homologue deleted on chromosome 10), a protein preferentially expressed in neurons and at the synapses [113]. Interestingly, PTEN functions concern with regulation of two elements of the neuronal survival process (i.e., neurite outgrowth and axonal regeneration) and seem to depend on its interactions with the proteins MAST2 and NHERF [113]. Furthermore, organizational functions at both the pre- and post-synaptic plasma membranes depend on the modular nature of PDZ domains and on their ability to form multivalent interactions [121, 122].

The majority of PDB (Protein Data Bank) structures of PDZ domains is characterized by 80-100 residues and by a fold which consists in five or six β-strands along with two α-helices (Fig. 10) [113, 123-125]. X-ray studies of the third PDZ domain of PSD-95 revealed a compact globular structure with six β-strands (named as β1, β2, β3, β4, β5 and β6) and two α-helices (named as α1 and α2) (Fig. 10A) [118, 123]. More in detail, the β1 strand runs along the up-and-down β-barrel formed by β2-β6 strands and makes hydrogen bonds with the β6 strand (Fig. 10A) [123].

The incorporation of the PDZ domain in proteins with other modules is favored by the N-terminal and the C-terminal ends which are positioned close to each other [118, 125].

Furthermore, one extremity of the β-barrel in PDZ domains is covered by the shorter α1 helix and its flanking loops whereas, the other extremity is capped by the longer α2 helix [123]. The structure of PDZ domains can be also characterized by a third short α-helix which is located at the C-terminus and wrapped against the space external to the β-barrel (Fig. 10A) [123].

Intriguingly, PDZ domains are able to recognize four different consensus sequences located at the C-termini of binding partners [113]: "Class I" consensus interaction motifs (i.e., S/T-X-Φ), "Class II" (Φ-X-Φ), "Class III" (Ψ-X-Φ) and "Class IV" (D-X-V) [113, 114, 125]. In these sequences X represents any residue, Φ indicates a hydrophobic amino acid (V, I, L, A, G, W, C, M and F) and Ψ represents a basic hydrophilic residue (H, R, K) [113]. In addition, a fourth amino acid (in position P-3 considering P0 the last C-terminal residue) contributes to the binding specificity [118, 125, 126].

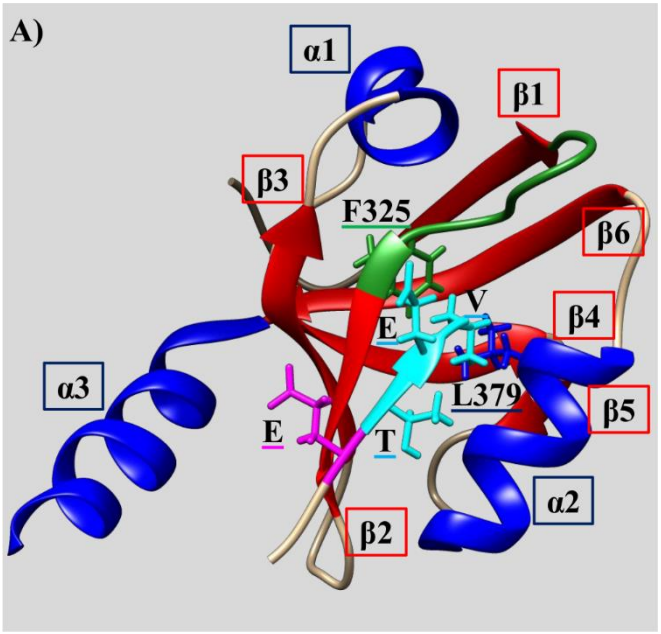


Fig. (**10**). A) X-ray structure of the third PDZ domain of PSD-95 in complex with a peptide possessing a "Class I-like" consensus motif and a modified FMOC moiety at the N-terminus (*FMOC*-KKETEV, PDB code: 5D13, Table S8). The β-strands (β1 (R312-H317), β2 (F325-G329), β3 (I336-I341), β4 (D357-V362), β5 (V365-D366) and β6 (T385-Y392)) are reported in red whereas, the α-helices (α1 (G345-G351), α2 (S371-N381) and α3 (K393-V406)) in blue. The amino acids of the peptide consensus interaction motif: (S/T)-X-Φ are shown in cyan whereas, the E residue in P-3 is shown in magenta. The residues in green are those belonging to the "carboxylate-binding loop" ((R/K)-X-X-X-G-Φ-G-Φ from R318 to F325). The side-chains of a few residues forming the hydrophobic pocket in which the consensus motif inserts (i.e., F325 and L379) are shown.

Although efficient binding to PDZ domain usually requires a motif composed by four residues, these interactions may be further supported by additional longer sequences with more amino acids [113]. For instance, NMR and X-ray studies showed that binding of the C-terminal end of the RABV (Rabies Virus) glycoprotein to the PDZ domain of MAST2 involves a total of 12 C-terminal residues, including the four residues of the consensus sequence [113]. In a similar way, interaction between MAST2-PDZ and the C-terminus of PTEN seems to require a total of 13 residues at the C-terminus of PTEN [113, 126].

Structural studies showed that consensus interaction sequences insert in a hydrophobic groove of the PDZ domain (between β2 and α2) where they make a network of hydrogen bonds with the residues of a loop (between β1 and β2) which is termed the "carboxylate-binding loop" and consists in a well-conserved sequence (i.e., (R/K)-X-X-X-G-Φ-G-Φ, where Φ indicates a hydrophobic residue) (Fig. 10A) [118, 124, 125, 127].

In detail, the C-terminal carboxylate in the peptide ligand is involved in several hydrogen bonds with the main chain amide groups in the PDZ carboxylate binding groove along with an ordered H_2O molecule which is coordinated by a conserved positively charged residue (lysine or arginine) in the same loop. Furthermore, the interaction between the groove of PDZ domain and the binding partners is accompanied by formation of a β-strand in the peptide interactors which, running in an antiparallel fashion with respect to the β2 strand in the PDZ domain, leads to an extension of the β-barrel arrangement [118, 125, 127].

The position of the peptide into the binding groove is steadily locked by two elements, the coordination of the terminal carboxylate group and the extensive β-strand/β-strand interactions involving the PDZ domain (β2-strand) and the binding partner [125, 127]. As a consequence of this structural topology of interaction, the side-chains of the residues in P0 (corresponding to the last C-terminal residue) and P-2 (the second residue N-terminally to the C-terminus) in the interaction consensus binding motif locate themselves at the base of the peptide-binding groove, thus playing a significant role for the specificity of binding [125, 127]. Instead, the side-chains of the residues in P-1 and P-3 positions point towards the surface of the protein and are solvent accessible [127]. The residue in P0 normally is positioned in a wide hydrophobic pocket [125]. For instance, the hydrophobic groove of the PDZ domain in the protein PSD-95 (i.e., F325, L379 and other hydrophobic residues) favors the presence of a valine at P0 whereas, a longer cavity in the PDZ domain of the protein NHERF (Na^+-H^+ Exchanger Regulatory Factor)/EBP50 favors a leucine in the same position of the consensus motif (Fig. 10A) [125]. Instead, residues in P-2 fit a distinct pocket, that in the case of the PSD-95-PDZ3 and other PDZ domains of class1, contains a histidine (H372), or in different PDZ domains may be leucine/methionine or a tyrosine [125]. Indeed, the variety of amino acidic composition among PDZ interaction pockets determines binding to consensus sequences with dissimilar amino acids in position -2; for instance, the presence of a leucine in the second cleft of PDZ domains induces a preference for interactors with serine/threonine in P-2 whereas, a His or a Tyr in PDZ pockets favor ligands with a hydrophobic residue or an aspartic acid in P-2 [125].

A few PDZ domains may even recognize internal motifs if they are part of a specific tertiary structure able to simulate conformationally the chain terminus (i.e., the structural element usually recognized by the PDZ domains) [125, 127]. For instance, the PDZ domains of PSD-95 and SNTA1 (Alpha-1-syntrophin) usually interact with the C-terminal consensus motif of nNOS but, they can also bind the PDZ domain of nNOS by an unusual linear head-to-tail arrangement which depends on a 30-residue extension located in the nNOS PDZ domain [124, 125, 127]. This 30-residue extension is characterized by a β-sheet composed by two short β-strands (3 residues in each strand) [127]. Upon binding to the PDZ domains of either PSD-95 or SNTA1 this small β-sheet in the 30-residue extension assumes a more rigid β-hairpin conformation (named as "β-finger") [124, 125, 127]. The "β-finger" possesses an "ETTF" motif which inserts in the peptide-binding groove and mimics a canonical C-terminal peptide ligand by providing interactions which involve the threonine in P-2 and the phenylalanine in P0 [125, 127].

PDZ proteins can also form diverse homo- and hetero-dimers [124, 128]. For instance, the PDZ module of shank1 protein and the sixth PDZ (or PDZ6) domain of GRIP1 (Glutamate Receptor-Interacting Protein 1) possess unusually long β2-β3 loops and N-terminal β1-strands which favor homo-dimerization [124]. Interestingly, the peptide binding pockets of both proteins are not involved in these interactions and are thus available for the binding with the consensus motifs of the relative binding partners [124]. Another example is provided by the second PDZ (or PDZ2) domain of ZO-1 which is involved in the creation of a characteristic swapped dimeric structure in which the two PDZ units exchange their β1 and β2 strands with each other without altering their overall folds (Fig. 10B) [124]. In this case, the peptide binding grooves position in regions which are at the opposite side of the dimeric arrangement [124]. Interestingly, the peptide binding grooves are usually not involved in the dimerization however, dimerizations between the SNTA1-PDZ and nNOS-PDZ, as well as between PSD-95-PDZ2 and nNOS-PDZ represent exceptions [124]. Indeed, the peptide binding grooves in the SNTA1-PDZ and PSD-95-PDZ2 are clogged by interaction with the C-terminal extension of nNOS [124].

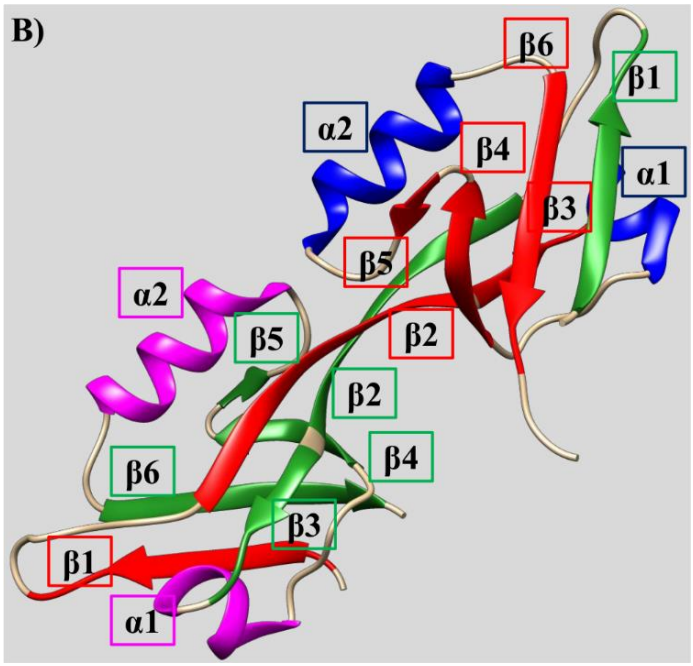


Fig. (**10**). B) X-ray structure of the second PDZ domain of ZO-1 (PDB code: 2RCZ [129]). The six β-strands (β1 (S185-V190), β2 (L200-F207), β3 (K209-Q213), β4 (V228-I232), β5 (T235- V236) and β6 (K255-Q261)) and the two α-helices (α1 (S215-G221) and α2 (S241-R251)) of one unit are reported in green and magenta, respectively. The β-strands (β1 (S185-V190), β2 (L200-F207), β3 (K209-Q213), β4 (V228-I232), β5 (T235-V236) and β6 (K255-Q261)) and α-helices (α1 (S215-G221) and α2 (S241-R251)) of the second unit are reported in red and blue, respectively.

6.3. Drug discovery approaches: Small molecules

PDZ proteins attract attention largely for their ability to behave as scaffolding proteins that regulate signal transduction pathways and as mediators of trafficking membrane proteins [9]. Nevertheless, PDZ domains have been suggested as valuable drug targets for the development of new therapeutics for neurological diseases (e.g., PSD-95, PICK1 (protein interacting with PRKCA 1) and Shank3) and cancers (for example AF6 (ALL1-Fused gene from chromosome 6), MAGI3 (Membrane-Associated Guanylate kinase, WW and PDZ domain-containing protein 3), NHERF1, MINT1 (Munc18-1-INTeracting protein 1), SAP97 (Synapse-Associated Protein 97), Dvl1 (Dishevelled segment polarity protein 1), and GIPC (GAIP C-terminus-interacting protein)) [9].

During the last years many efforts have been centered on finding compound inhibitors of PDZ mediated protein-protein interactions. However, this is a very challenging task due to the relatively elongated and shallow nature of the binding pockets [9, 130]. Indeed, the small molecules targeting PDZ domains, that have been discovered thus far, possess interaction affinities above the low micromolar range, even after optimization by SAR (Structure-Activity Relationship) studies (Fig. 11A upper panel) [9, 130].

Cancer cells are usually provided with stem cells that are resistant to chemo-therapy and whose self-regeneration and pro-cancer outcomes are linked to Wnt/β-catenin signalling,

which in turn can be correlated to the PDZ protein Dvl [131]. An interesting study was conducted by coupling virtual screening and NMR chemical shift perturbation studies with a ^{15}N-labelled PDZ domain, and led to the identification of novel small molecule ligands of hDvl1 (human Dvl1)-PDZ (Fig. 11A, lower panel) [131]. A few of these molecules (named NPL-4001, NPL-4004 and NPL-4012 in Fig. 11A) are better ligands of Dvl-PDZ than the commercially available inhibitor known as "CalBiochem-322338" (Fig. 11A, lower panel). Furthermore, these small molecules, by working as negative regulators of Wnt pathway, exhibit a certain anti-proliferative activity against a TNBC (Triple Negative Breast Cancer) cell line (i.e., BT-20) thus representing promising candidates for anti-TNBC therapies [131].

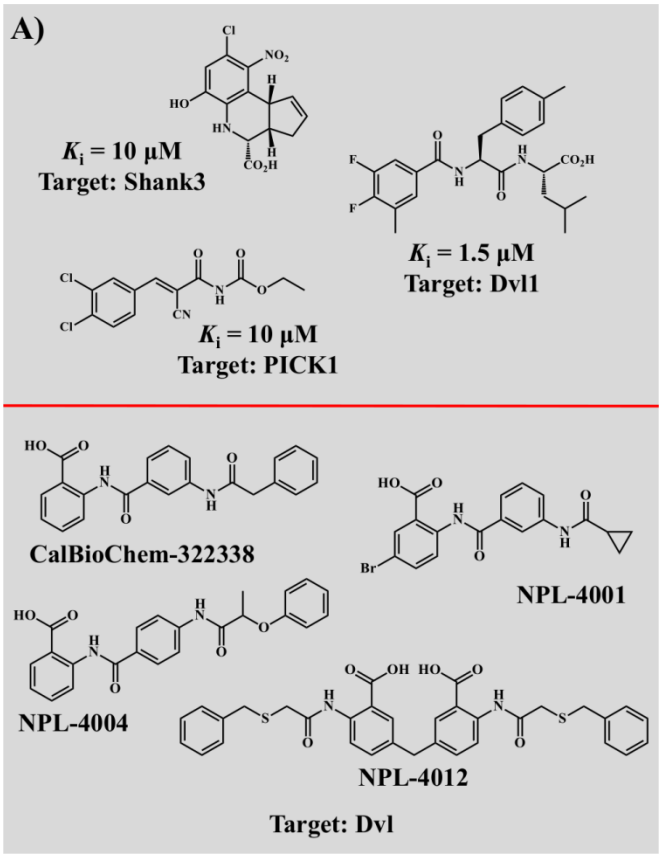


Fig. (**11**). A) (Upper panel) Chemical structures of a few small molecule inhibitors of PDZ domains from different proteins. Values of inhibition constants (K_i) and the therapeutically relevant protein targets are reported. K_i estimates were derived from fluorescence polarization competition-type assays [132-134]. (Lower panel) Chemical structures of small molecule inhibitors of Dvl1-PDZ [131].

6.4. Drug Discovery approaches: Peptides/Peptidomimetics

Several studies were centered on peptide inhibitors of PDZ domains rather than small molecules and many peptides were designed starting from PBM (PDZ Binding Motifs) located at the C-termini of different proteins. An interesting example is provided by the Tat-NR2B9c (or NA-1) peptide (Fig. 11B). Tat-NR2B9C is made up of 20 residues and includes a Tat sequence, that enhances blood-brain barrier permeability, and an additional segment encompassing the last nine C-terminal residues of the GluN2B (also known as NR2B (N-methyl D-aspartate Receptor subtype 2B)) portion of the NMDA receptor [9, 130]. Tat-NR2B9c is able to interact with both the PDZ1 and PDZ2 domains of PSD-95 and has been investigated as potential therapeutic compound in the treatment of cerebral ischemia [9, 130]. It has been implemented in a phase II clinical trial; unfortunately, it presents the disadvantage to bind too weakly to PSD-95 with an inhibition constant K_i equal to 4 µM against PSD-95-PDZ2 [9, 130].

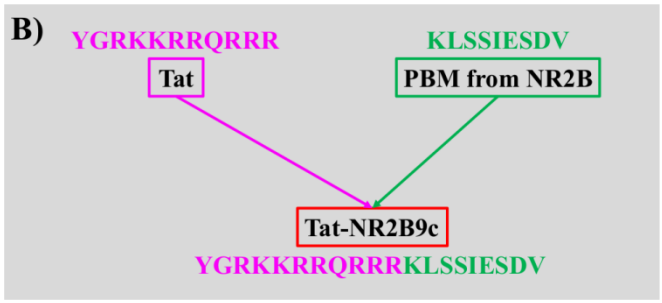


Fig. (**11**). B) Example of a peptide inhibitor derived from a C-teminal sequence of PDZ binding partners.

PDZ1 and PDZ2 modules exist as tandem domains in the protein PSD-95, thus, to overcome the low affinity challenge, dimeric inhibitors were generated and evaluated [9, 130]. In details, to create a more potent PDZ peptide interactor, two peptide ligands possessing low micromolar affinity for the two PDZ1-2 domains in PSD-95 were linked together through diverse PEG (PolyEthylene Glycol) linkers (Fig.11C) [9, 130]. Interestingly, the strategy did lead to molecules with highly improved affinity toward the PDZ1,2 domains of PSD-95 if compared to the affinity of the single peptides for their target PDZ domains (Fig. 11C) [9, 130]. Furthermore, through this kind of dimeric inhibitors, an increased *in vitro* stability in blood plasma was achieved as well [9, 130]. One of the best dimeric compounds was also optimized by inserting a modified linker and a Tat sequence [9, 130]. The final molecule (named as ‘‘Tat-*N*-dimer” Fig. 11C) is characterized by a nanomolar affinity toward the PDZ tandem and by a neuroprotective action which is better if compared to that observed with Tat-NR2B9c (Fig. 11B) in a mouse model of ischemic brain damage [9, 130].

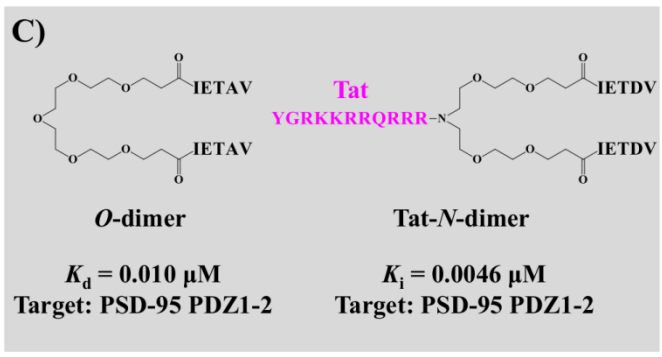


Fig. (**11**). C) Bidentate inhibitors of the tandem PDZ1-2 domain of PSD-95 [135]. K_d and K_i values for O-dimer and Tat-N-dimer were measured as reported in references [136] and [135], respectively. The peptide sequence -IETVD- respect to -IETAV- ensures higher selectivity of binding towards PDZ1-2 domains *vs* PDZ3 domain of PSD-95 [135].

Cellular PDZ proteins with their pivotal functions in cell signaling pathways are easily attacked by viruses eager to exploit cellular function at their own advantage [137]. For instance, in the case of the rabies virus (RBV), following viral infection the RBV G protein interacts with the PDZ domains from key neuronal enzymes -including MAST2 and PTPN4 (Protein Tyrosine Phosphatase Non-receptor type 4)- thus interfering with the interaction network made up by cellular enzymes and their binding partners and producing terrible outcomes in the host [137]. Indeed, viral proteins compete -with affinity and selectivity- with cellular proteins for binding to PDZ domains by simulating PDZ binding motifs [137].

These observations have been exploited for the development of novel and promising therapies as it was found that peptides of viral origin could play a neuroprotective role and induce cell death when delivered inside the cell [137]. Starting from structural and biophysical data related to complexes in between PDZ domains and short viral polypeptides, upon a two-step chemical optimization, it was possible to obtain potent interactors of PTPN4-PDZ an MAST2-PDZ (affinities for best ligands equal to 1 µM and 60 nM, respectively) (Fig. 11D) [137]. These peptides have interesting biological properties and for example peptides against MAST2-PDZ domain are in clinical trial due to their neuro-regeneration ability in animal models [137].

Fig. (11). D) Peptide inhibitor of the PDZ domain of PTPN4 obtained through optimization of a RABV amino acid sequence [138].

Another strategy proposed for the development of PDZ inhibitors is the “peptidomimetics route” (Fig. 11E) that should enhance stability of peptide inhibitors and lead to more drug-like compounds [9, 130]. Peptidomimetics can be generated through different modifications of a simple peptide segment, including peptide cyclization, insertion of unnatural amino acids and the substitution of specific amide bonds with thioamides (Fig. 11E) [9, 130]. A few examples of peptidomimetic inhibitors of PDZ domains are reported in the upper panel of Fig. 11E [139, 140]. In this peptidomimetics contest it's also worth mentioning the lipopeptide CR1166 (i.e., *N*-myristoyl-PSQSK(ε*N*-4-bromobenzoyl)SK(ε*N*-4-bromobenzoyl)A) characterized by cell-permeability and able to inhibit the GIPC (GAIP-interacting protein, C terminus) PDZ domain both *in vitro* and *in vivo* (Fig. 11E, lower panel) [141]. CR1166 exerts also a significant inhibitory action against pancreatic and breast cancer cells and tumors [141].

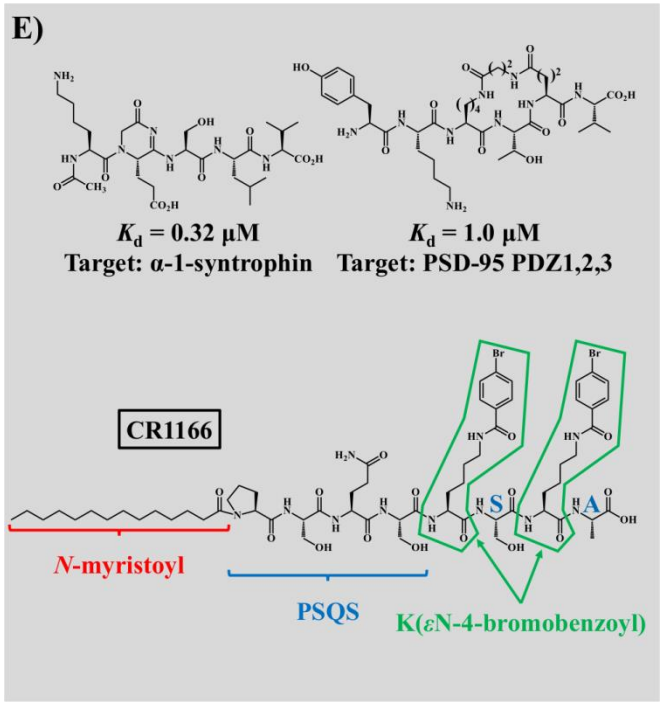


Fig. (**11**). E) Structures of a few peptidomimetic inhibitors of PDZ domains. "*" K_d value obtained for binding to PSD-95-PDZ1 [9, 139, 140].

7. SAM DOMAINS

SAM (Sterile Alpha Motif) domains are small protein modules of roughly 70 amino acids that can be found in all organisms [142]. SAMs can be considered protein-protein interaction modules, as they can form homo- and hetero-dimeric, oligomeric and even polymeric complexes with other SAM domain containing proteins but, they can also bind proteins lacking SAM domains [142-145]. In addition, SAM domains have been reported to bind RNA and lipids [142]. The multiplicity of interaction properties characterizing SAM domains confers to the related proteins a variety of functions [142, 143].

7.1. Structural details of the SAM fold and SAM-SAM interactions

This small modules are usually structured in a five-helix bundle [146-148] (Fig. 12), with the exception of DLC (Deleted in Liver Cancer) 1 and DLC2 SAM domains that are characterized by four helices [142]. SAM domains can exploit different structural binding architectures when interacting with other SAM domains, and the most common binding topology is the head-to-tail orientation, also called “Mid-Loop/End-Helix” (ML/EH) model (Fig. 12A) [146]. This binding mode involves the central region of one SAM domain (Mid-Loop) and the C-terminal α5 helix together with adjacent loops of the other partner SAM domain (End-Helix) (Fig. 12A) [146]. Many ML/EH SAM-SAM complexes are driven by electrostatic interactions [149-151]. In addition, a glycine, positioned on the N-terminal end of the α5 helix on the EH region, seems to be an important

anchoring point for ML surface engagement, as a H-bond in between the glycine H_N and the backbone $_CO$ of a residue belonging to the α2 helix of the ML region has been found in different complexes (Fig. 12A) [152-156]. SAM-SAM associations may also present a tail-to-tail organization, in which α5 helices interact with each other with a reverse orientation (as in the oligomeric structure of the SAM domain from the Ephrin B2 receptor [157]) or with the same orientation (Sly1 SAM domains, in Fig. 12B) [147].

7.2. Drug discovery approaches: peptides targeting SAM domains

Proteins containing SAM domains have been related to different pathological conditions, such as cataracts, cancer, and neurological disorders thus, they potentially represent targets for therapeutics development [142]. For example, the SAM domain from EphA2 receptor (EphA2-SAM) is of interest because this tyrosine kinase receptor is related to different forms of cancer [146]. In particular the interaction between EphA2 and the lipid phosphatase Ship2, that occurs through a heterotypic SAM-SAM interaction, produces pro-oncogenic outcomes, by negatively regulating receptor endocytosis and degradation processes with a consequent enhancement of cancer cells migration [146, 158]. Thus, inhibition of the interaction between SAM domains of EphA2 and Ship2 (Ship2-SAM) may be considered a possible strategy in anti-cancer drug discovery [142, 146].

The first SAM domain of the protein Odin (Odin-SAM1) is another binding partner of EphA2-SAM, and, although Odin seems to modulate receptor stability [159], a clear connection between the SAM-SAM interaction and pro-oncogenic routes has not been established [142, 146]. NMR and X-Ray structures of EphA2-SAM/Ship2-SAM and EphA2-SAM/Odin-SAM1 complexes indicate a common ML/EH topology of binding in which the SAM domain of EphA2 supplies the EH region (Fig. 12A) [149, 150, 152, 160].

SAM-SAM complexes are difficult to target with small molecules due to the large and shallow nature of these protein-protein interaction surfaces, thus, different strategies aimed to identify peptide inhibitors were adopted, and a few weak peptide ligands were identified [161-166].

Recently, the ML portion of Ship2-SAM was used as starting sequence for the design of peptide libraries, that were virtually screened against EphA2-SAM and experimentally tested by a multidisciplinary approach, leading to the discovery of the ShipH1 peptide (sequence: NGWDDLEFLEDIwEEDL, were w= D-Trp) as EphA2-SAM ligand (K_d= 72.4 ± 0.5 μM in MST assays) [166].

The EH interacting region of EphA2-SAM for Ship2-SAM and Odin-SAM1 was also considered in a few peptide design approaches [164, 165]. The $(KRI)_3$ peptide is a Ship2-SAM ligand (K_d= 83 ± 8 μM in SPR (Surface Plasmon Resonance) experiments); the peptide sequence is characterized by the motif "KRIAY", that belongs to the EphA2-SAM α5 helix included in the EH region, repeated thrice in tandem [164]. The $(KRI)_3$ peptide was reported to be more cytotoxic towards PC (Prostate Cancer) 3 cells than NHDF (Normal Human Dermal Fibroblasts) [164]. The stapled peptide S13ST, was also designed starting from the α5 helix of EphA2-SAM (sequence: KRIGVRLPGHQKRXAYSXLGLKDQV, where X=(S)-2-(4'-pentenyl) alanine) [165]. This peptide is characterized by a hydrocarbon stapling introduced to enhance peptide helical content, and is able to interact with the Ship2-SAM ML region (K_d= 52.2 ± 0.7 μM in MST experiments) [165].

Although some success has been obtained in the challenging goal of finding SAM domains peptide ligands, additional efforts must be devoted to enhance their binding affinity and pharmacological properties.

SAM domains therapeutic potential, structural features and interaction properties have been much largely discussed in our recent previous reviews [142, 146].

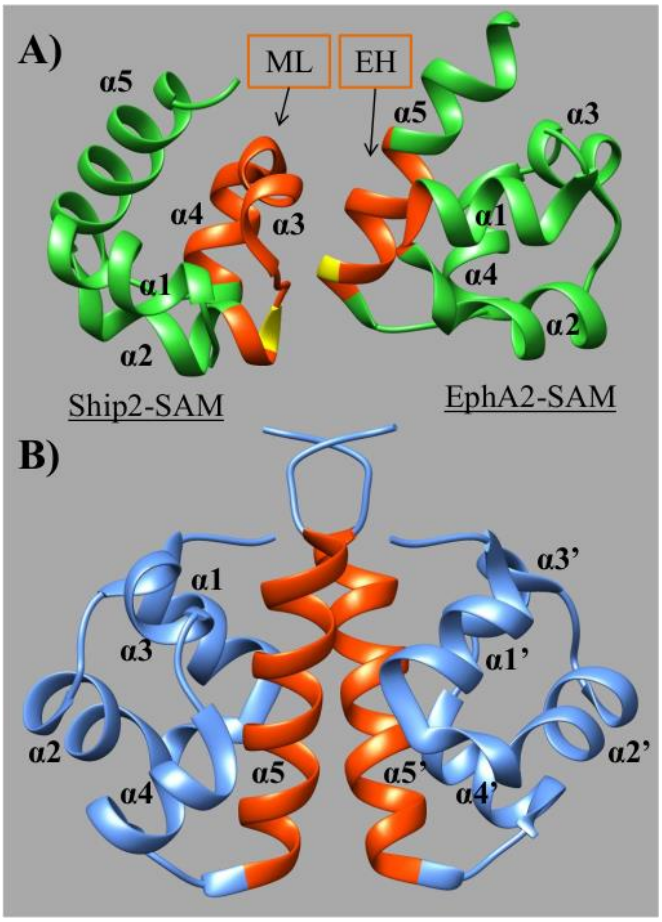


Fig. (12). A) X-Ray structure of the Ship2-SAM/EphA2-SAM complex (PDB code: 5ZRX [152]). ML (residues V1217-E1237 of Ship2-SAM) and EH (residues I917, K918, P953-Y961 of EphA2-SAM) regions are colored orange; G954 on EphA2 and N1219 on Ship2, that are involved in a hydrogen bond at the protein-protein interface, are colored yellow. B) NMR solution structure of the Sly1-SAM homo-dimer (first conformer, PDB code: 6G8O [147]) with α5 interacting helices colored orange. The two SAM domains are cross-linked through a disulphide bridge (obtained mutating S320 at each C-terminus with cysteines) to stabilize the dimeric state [147].

8. CONCLUSION

Most physiological processes as well as pathological conditions are linked to protein-protein interactions mediated by PIDs.

This review reports on PIDs able to recognize simple amino acid sequences. The different protein modules, that we are

describing herein, have diverse folds, that are either completely helical -as in the case of CH domains [94]-, α/β mixed with a variety of structural arrangements (See for example UEV [16], EH [105], PDZ [9] domains) or mainly β- like the SH3 domain [38].

Moreover, PIDs have a multiplicity of interaction preferences; a few of them bind to more than a single consensus motif, for instance, three diverse classes of peptide interactors have been described for EH domains [7]; on the contrary, binding to phospho-inositides has also been reported for CH domains [167]; LIM [77] domains interact with metal ions (Zn^{2+}) too.

These PIDs through their specific protein-protein interactions may play roles in an array of diseases. Different types of cancer and metastatic spread can be linked to SH3 [45], EVH1 [60], LIM [86, 87, 89] and EH [111] domains of specific proteins; SH3 domains may be related also to bones diseases like osteoporosis [44]. LIM domains have a connection with cardiovascular pathologies [88]; UEV domains on the contrary have been related to viral infections (including HIV and HEV infections) [73, 74] whereas, GYF to Parkinson's disease [68].

The association with pathological conditions makes these protein modules particularly interesting in drug discovery and during the last years several studies, as reported above, have been focused on the design and evaluation of compounds able to modulate their interaction properties.

PIDs may play a prominent role in the field of personalized medicine although this aspect to date remains still largely unexplored. In theory, protein interaction patterns may vary from patient to patient. To detect and quantify interactomes in human cancers or other diseases, through studying samples directly taken from the affected patients, should hold a tremendous amount of information [21]. To identify correlations between interactomes and distinctive features of specific diseases, like certain clinical outcomes, as explained in the Introduction Section, is really appealing as could suggest crucial PIDs and specific interactions to be targeted by medicinal chemistry efforts and help building personalized medicine therapies. Of course the set up of fast and reliable methods to study interactomes in cell derived from patients affected by specific diseases is still in the future but, as also highlighted by other authors before us, this should be definitively a goal to pursuit [21, 24].

Structure-based methods to design inhibitors or general modulators of protein/protein interactions mediated by PIDs are in our opinion the best strategies to follow as in principle may enable selectivity.

It would also be interesting to study directly in patients affected by specific diseases, mutations in certain PIDs, clarify how these mutations may influence PID structure and interaction network and finally, create large databases of mutated domains to be exploited to set up personalized medicine therapies. In this context it would be interesting to design compound stabilizing the perturbed 3D fold of the mutated PID, and try to restore interactomes of healthy cells.

In conclusion, this review intends to collect structural features on PIDs in isolation and when bound to peptide ligands and inspire other authors working in the field of the structure-based drug design. The drug discovery strategies presented here, represent only a few examples of the variety of methodologies that can be used to target PIDs, as many others have been reported through the years in the literature. To establish a general and successful structure-based drug discovery route (i.e., a first choice) to adopt when targeting PIDs and their interactomes is -in our opinion- impossible as the outcomes will largely depend on the particular protein/domain under study. We believe that it is always better to combine single strategies together to allow several cycles of compounds optimization and improve iteratively drug like properties.

Targeting protein-protein interactions is never easy and has to face challenging wide and smooth surfaces lacking real binding pockets. In this context, many efforts have been centered on SH3 and PDZ domains for which the larger amount of structural data are available but, the druggability of domains like CHs and LIMs still needs to be mostly investigated.

Nevertheless, much work has been conducted in the field of peptides/peptidomimetics generated by medicinal chemistry efforts starting from known interaction consensus sequences. However, many of the identified peptide-like compounds have still either poor drug-like characters, and/or poor affinities and biological activities. More needs definitely to be done to discover possibly drug like molecules able to efficiently modulate PIDs interaction networks.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ACKNOWLEDGEMENTS

Fondazione Umberto Veronesi is gratefully acknowledged for a fellowship to MV (GRANT 2019).

SUPPORTIVE/SUPPLEMENTARY MATERIAL

Tables listing PDB entries related to complexes in which different protein binding modules are involved.

REFERENCES

[1] Mayer, B.J., Protein-protein interactions in signaling cascades. *Mol. Biotechnol.*, **1999**, *13*(3), 201-213.

[2] Pawson, T.; Raina, M.; Nash, P., Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett.*, **2002**, *513*(1), 2-10.

[3] Pawson, T., Protein Modules and Signaling Networks. *Nature*, **1995**, *373*(6515), 573-580.

[4] Pawson, T.; Nash, P., Assembly of cell regulatory systems through protein interaction domains. *Science*, **2003**, *300*(5618), 445-452.

[5] Liu, B.A.; Engelmann, B.W.; Nash, P.D., High-throughput analysis of peptide-binding modules. *Proteomics*, **2012**, *12*(10), 1527-1546.

[6] Zarrinpar, A.; Bhattacharyya, R.P.; Lim, W.A., The structure and function of proline recognition domains. *Sci. STKE*, **2003**, *2003*(179), re8.

[7] Polo, S.; Confalonieri, S.; Salcini, A.E.; Di Fiore, P.P., EH and UIM: endocytosis and more. *Sci. STKE*, **2003**, *2003*(213), re17.

[8] Montesinos, M.L.; Castellano-Munoz, M.; Garcia-Junco-Clemente, P.; Fernandez-Chacon, R., Recycling and

EH domain proteins at the synapse. *Brain Res. Rev.*, **2005**, *49*(2), 416-428.
[9] Chi, C.N.; Bach, A.; Stromgaard, K.; Gianni, S.; Jemth, P., Ligand binding by PDZ domains. *BioFactors*, **2012**, *38*(5), 338-348.
[10] Brown, S.; Coghill, I.D.; McGrath, M.J.; Robinson, P.A., Role of LIM domains in mediating signaling protein interactions. *IUBM Life*, **2001**, *51*(6), 359-364.
[11] Korenbaum, E.; Rivero, F., Calponin homology domains at a glance. *J. Cell Sci.*, **2002**, *115*(18), 3543-3545.
[12] Dalgarno, D.C.; Botfield, M.C.; Rickles, R.J., SH3 domains and drug design: ligands, structure, and biological function. *Biopolymers*, **1997**, *43*(5), 383-400.
[13] Renfranz, P.J.; Beckerle, M.C., Doing (F/L)PPPPs: EVH1 domains and their proline-rich partners in cell polarity and migration. *Curr. Opin. Cell Biol.*, **2002**, *14*(1), 88-103.
[14] Peterson, F.C.; Volkman, B.F., Diversity of polyproline recognition by EVH1 domains. *Front. Biosci. (Landmark Ed.)*, **2009**, *14*, 833-846.
[15] Nishizawa, K.; Freund, C.; Li, J.; Wagner, G.; Reinherz, E.L., Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proc. Natl. Acad. Sci. U.S.A.*, **1998**, *95*(25), 14897-14902.
[16] Hurley, J.H.; Lee, S.; Prag, G., Ubiquitin-binding domains. *Biochem. J.*, **2006**, *399*, 361-372.
[17] Sang, M.X.; Ma, L.; Sang, M.J.; Zhou, X.L.; Gao, W.; Geng, C.Z., LIM-domain-only proteins: multifunctional nuclear transcription coregulators that interacts with diverse proteins. *Mol. Biol. Rep.*, **2014**, *41*(2), 1067-1073.
[18] Banuelos, S.; Saraste, M.; Djinovic Carugo, K., Structural comparisons of calponin homology domains: implications for actin binding. *Structure*, **1998**, *6*(11), 1419-1431.
[19] Confalonieri, S.; Di Fiore, P.P., The Eps15 homology (EH) domain. *FEBS Lett.*, **2002**, *513*(1), 24-29.
[20] Qiao, F.; Bowie, J.U., The many faces of SAM. *Sci. STKE*, **2005**, *2005*(286), re7.
[21] Haura, E.B., From modules to medicine: How modular domains and their associated networks can enable personalized medicine. *FEBS Lett.*, **2012**, *586*(17), 2580-2585.
[22] Taylor, I.W.; Linding, R.; Warde-Farley, D.; Liu, Y.; Pesquita, C.; Faria, D.; Bull, S.; Pawson, T.; Morris, Q.; Wrana, J.L., Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.*, **2009**, *27*(2), 199-204.
[23] Vincenzi, M.; Mercurio, F.A.; Leone, M., Protein Interaction Domains and post translational modifications: structural features and drug discovery applications. *Curr. Med. Chem.*, **2019**.
[24] Machida, K.; Eschrich, S.; Li, J.; Bai, Y.; Koomen, J.; Mayer, B.J.; Haura, E.B., Characterizing Tyrosine Phosphorylation Signaling in Lung Cancer Using SH2 Profiling. *PLoS One*, **2010**, *5*(10).
[25] Opitz, R.; Muller, M.; Reuter, C.; Barone, M.; Soicke, A.; Roske, Y.; Piotukh, K.; Huy, P.; Beerbaum, M.; Wiesner, B.; Beyermann, M.; Schmieder, P.; Freund, C.; Volkmer, R.; Oschkinat, H.; Schmalz, H.G.; Kuhne, R., A modular toolkit to inhibit proline-rich motif-mediated protein-protein interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **2015**, *112*(16), 5011-5016.
[26] Buday, L.; Downward, J., Epidermal growth factor regulates p21ras through the formation of a complex of receptor, Grb2 adapter protein, and Sos nucleotide exchange factor. *Cell*, **1993**, *73*(3), 611-620.
[27] Ball, L.J.; Jarchau, T.; Oschkinat, H.; Walter, U., EVH1 domains: structure, function and interactions. *FEBS Lett.*, **2002**, *513*(1), 45-52.
[28] Sudol, M.; Sliwa, K.; Russo, T., Functions of WW domains in the nucleus. *FEBS Lett.*, **2001**, *490*(3), 190-195.
[29] Kay, B.K.; Williamson, M.P.; Sudol, M., The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.*, **2000**, *14*(2), 231-241.
[30] Kurochkina, N.; Guha, U., SH3 domains: modules of protein-protein interactions. *Biophys. Rev.*, **2013**, *5*(1), 29-39.
[31] Mayer, B.J., SH3 domains: complexity in moderation. *J. Cell Sci.*, **2001**, *114*(Pt 7), 1253-1263.
[32] Gmeiner, W.H.; Horita, D.A., Implications of SH3 domain structure and dynamics for protein regulation and drug design. *Cell Biochem. Biophys.*, **2001**, *35*(2), 127-140.
[33] Carducci, M.; Perfetto, L.; Briganti, L.; Paoluzi, S.; Costa, S.; Zerweck, J.; Schutkowski, M.; Castagnoli, L.; Cesareni, G., The protein interaction network mediated by human SH3 domains. *Biotechnol. Adv.*, **2012**, *30*(1), 4-15.
[34] Saksela, K.; Permi, P., SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Lett.*, **2012**, *586*(17), 2609-2614.
[35] Aitio, O.; Hellman, M.; Kesti, T.; Kleino, I.; Samuilova, O.; Paakkonen, K.; Tossavainen, H.; Saksela, K.; Permi, P., Structural basis of PxxDY motif recognition in SH3 binding. *J. Mol. Biol.*, **2008**, *382*(1), 167-178.
[36] Lim, W.A., Reading between the lines: SH3 recognition of an intact protein. *Structure*, **1996**, *4*(6), 657-659.
[37] Teyra, J.; Sidhu, S.S.; Kim, P.M., Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains. *FEBS Lett.*, **2012**, *586*(17), 2631-2637.
[38] Kaneko, T.; Li, L.; Li, S.S., The SH3 domain--a family of versatile peptide- and protein-recognition module. *Front. Biosci.*, **2008**, *13*, 4938-4952.
[39] Dikic, I., CIN85/CMS family of adaptor molecules. *FEBS Lett.*, **2002**, *529*(1), 110-115.
[40] Schnoor, M.; Stradal, T.E.; Rottner, K., Cortactin: Cell Functions of A Multifaceted Actin-Binding Protein. *Trends cell biol.*, **2018**, *28*(2), 79-98.
[41] Liu, S.K.; Smith, C.A.; Arnold, R.; Kiefer, F.; McGlade, C.J., The adaptor protein Gads (Grb2-related adaptor downstream of Shc) is implicated in coupling hemopoietic progenitor kinase-1 to the activated TCR. *J. Immunol.*, **2000**, *165*(3), 1417-1426.
[42] Camara-Artigas, A.; Ortiz-Salmeron, E.; Andujar-Sanchez, M.; Bacarizo, J.; Martin-Garcia, J.M., The role of water molecules in the binding of class I and II peptides to the SH3 domain of the Fyn tyrosine kinase. *Acta Crystallogr. F Struct. Biol. Commun.*, **2016**, *72*(Pt 9), 707-712.
[43] Nguyen, J.T.; Porter, M.; Amoui, M.; Miller, W.T.; Zuckermann, R.N.; Lim, W.A., Improving SH3 domain ligand selectivity using a non-natural scaffold. *Chem. biol.*, **2000**, *7*(7), 463-473.

[44] Han, S.J.; Liu, Q.; Wang, F.; Yuan, Z.N., Targeting the SH3 domain of human osteoclast-stimulating factor with rationally designed peptoid inhibitors. *J. Pept. Sci.,* **2016**, *22*(8), 533-539.
[45] Smithgall, T.E., Sh2 and Sh3 Domains - Potential Targets for Anticancer Drug Design. *J. Pharmacol. Toxicol. Methods,* **1995**, *34*(3), 125-132.
[46] Vohidov, F.; Knudsen, S.E.; Leonard, P.G.; Ohata, J.; Wheadon, M.J.; Popp, B.V.; Ladbury, J.E.; Ball, Z.T., Potent and selective inhibition of SH3 domains with dirhodium metalloinhibitors. *Chem. Sci.,* **2015**, *6*(8), 4778-4783.
[47] Oneyama, C.; Nakano, H.; Sharma, S.V., UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug. *Oncogene,* **2002**, *21*(13), 2037-2050.
[48] Oneyama, C.; Agatsuma, T.; Kanda, Y.; Nakano, H.; Sharma, S.V.; Nakano, S.; Narazaki, F.; Tatsuta, K., Synthetic inhibitors of proline-rich ligand-mediated protein-protein interaction: potent analogs of UCS15A. *Chem. biol.,* **2003**, *10*(5), 443-451.
[49] Grover, P.; Shi, H.B.; Baumgartner, M.; Camacho, C.J.; Smithgall, T.E., Fluorescence Polarization Screening Assays for Small Molecule Allosteric Modulators of ABL Kinase Function. *PLoS One,* **2015**, *10*(7).
[50] Chen, S.; Brier, S.; Smithgall, T.E.; Engen, J.R., The Abl SH2-kinase linker naturally adopts a conformation competent for SH3 domain binding. *Protein Sci.,* **2007**, *16*(4), 572-581.
[51] Inglis, S.R.; Stojkoski, C.; Branson, M.M.; Cawthray, J.F.; Fritz, D.; Wiadrowski, E.; Pyke, S.M.; Booker, G.W., Identification and specificity studies of small-molecule ligands for SH3 protein domains. *J. Med. Chem.,* **2004**, *47*(22), 5405-5417.
[52] Naisbitt, S.; Kim, E.; Tu, J.C.; Xiao, B.; Sala, C.; Valtschanoff, J.; Weinberg, R.J.; Worley, P.F.; Sheng, M., Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin. *Neuron,* **1999**, *23*(3), 569-582.
[53] Brakeman, P.R.; Lanahan, A.A.; OBrien, R.; Roche, K.; Barnes, C.A.; Huganir, R.L.; Worley, P.F., Homer: A protein that selectively binds metabotropic glutamate receptors. *Nature,* **1997**, *386*(6622), 284-288.
[54] Tu, J.C.; Xiao, B.; Yuan, J.P.; Lanahan, A.A.; Leoffert, K.; Li, M.; Linden, D.J.; Worley, P.F., Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron,* **1998**, *21*(4), 717-726.
[55] Fedorov, A.A.; Fedorov, E.; Gertler, F.; Almo, S.C., Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. *Nat. Struct. Mol. Biol.,* **1999**, *6*(7), 661-665.
[56] Beneken, J.; Tu, J.C.; Xiao, B.; Nuriya, M.; Yuan, J.P.; Worley, P.F.; Leahy, D.J., Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. *Neuron,* **2000**, *26*(1), 143-154.
[57] Peterson, F.C.; Deng, Q.; Zettl, M.; Prehoda, K.E.; Lim, W.A.; Way, M.; Volkman, B.F., Multiple WASP-interacting protein recognition motifs are required for a functional interaction with N-WASP. *J. Biol. Chem.,* **2007**, *282*(11), 8446-8453.
[58] Zimmermann, J.; Jarchau, T.; Waltr, U.; Oschkinat, H.; Ball, L.J., Letter to the Editor: H-1, C-13 and N-15 resonance assignment of the human Spred2 EVH1 domain. *J. Biomol. NMR,* **2004**, *29*(3), 435-436.
[59] Le Clainche, C.; Carlier, M.F., Regulation of actin assembly associated with protrusion and adhesion in cell migration. *Physiol. Rev.,* **2008**, *88*(2), 489-513.
[60] Gertler, F.; Condeelis, J., Metastasis: tumor cells becoming MENAcing. *Trend. cell biol.,* **2011**, *21*(2), 81-90.
[61] Hunke, C.; Hirsch, T.; Eichler, J., Structure-based synthetic mimicry of discontinuous protein binding sites: Inhibitors of the interaction of Mena EVH1 domain with proline-rich ligands. *ChemBioChem,* **2006**, *7*(8), 1258-1264.
[62] Hopkins, A.L.; Keseru, G.M.; Leeson, P.D.; Rees, D.C.; Reynolds, C.H., The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.,* **2014**, *13*(2), 105-121.
[63] Kofler, M.M.; Freund, C., The GYF domain. *FEBS J.,* **2006**, *273*(2), 245-256.
[64] Ash, M.R.; Faelber, K.; Kosslick, D.; Albert, G.I.; Roske, Y.; Kofler, M.; Schuemann, M.; Krause, E.; Freund, C., Conserved beta-hairpin recognition by the GYF domains of Smy2 and GIGYF2 in mRNA surveillance and vesicular transport complexes. *Structure,* **2010**, *18*(8), 944-954.
[65] Georgiev, A.; Sjostrom, M.; Wieslander, A., Binding specificities of the GYF domains from two Saccharomyces cerevisiae Paralogs. *Protein Eng. Des. Sel.,* **2007**, *20*(9), 443-452.
[66] Freund, C.; Dotsch, V.; Nishizawa, K.; Reinherz, E.L.; Wagner, G., The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nat. Struct. Biol.,* **1999**, *6*(7), 656-660.
[67] Freund, C.; Schmalz, H.-G.; Sticht, J.; Kuhne, R. In *Protein-Protein Interactions as New Drug Targets*. Klussmann E., S.J., Ed.; Springer, **2008**; Vol. *186*, pp 408-422.
[68] Ruiz-Martinez, J.; Krebs, C.E.; Makarov, V.; Gorostidi, A.; Marti-Masso, J.F.; Paisan-Ruiz, C., GIGYF2 mutation in late-onset Parkinson's disease with cognitive impairment. *J. Hum. Genet.,* **2015**, *60*(10), 637-640.
[69] Kofler, M.; Motzny, K.; Beyermann, M.; Freund, C., Novel interaction partners of the CD2BP2-GYF domain. *J. Biol. Chem.,* **2005**, *280*(39), 33397-33402.
[70] Uryga-Polowy, V.; Kosslick, D.; Freund, C.; Rademann, J., Resin-Bound Aminofluorescein for C-Terminal Labeling of Peptides: High-Affinity Polarization Probes Binding to Polyproline-Specific GYF Domains. *ChemBioChem,* **2008**, *9*(15), 2452-2462.
[71] Freund, C.; Kuhne, R.; Yang, H.; Park, S.; Reinherz, E.L.; Wagner, G., Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. *EMBO J.,* **2002**, *21*(22), 5985-5995.
[72] Pornillos, O.; Alam, S.L.; Davis, D.R.; Sundquist, W.I., Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat. Struct. Mol. Biol.,* **2002**, *9*(11), 812-817.
[73] Pornillos, O.; Alam, S.L.; Rich, R.L.; Myszka, D.G.; Davis, D.R.; Sundquist, W.I., Structure and functional interactions of the Tsg101 UEV domain. *EMBO J.,* **2002**, *21*(10), 2397-2406.

[74] Yang, X.; Lennard, K.R.; He, C.; Walker, M.C.; Ball, A.T.; Doigneaux, C.; Tavassoli, A.; van der Donk, W.A., A lanthipeptide library used to identify a protein-protein interaction inhibitor. *Nat. Chem. Biol. ,* **2018**, *14*(4), 375-380.

[75] Im, Y.J.; Kuo, L.; Ren, X.F.; Burgos, P.V.; Zhao, X.Z.; Liu, F.; Burke, T.R.; Bonifacino, J.S.; Freed, E.O.; Hurley, J.H., Crystallographic and Functional Analysis of the ESCRT-I/HIV-1 Gag PTAP Interaction. *Structure,* **2010**, *18*(11), 1536-1547.

[76] Anang, S.; Kaushik, N.; Hingane, S.; Kumari, A.; Gupta, J.; Asthana, S.; Shalimar; Nayak, B.; Ranjith-Kumar, C.T.; Surjit, M., Potent Inhibition of Hepatitis E Virus Release by a Cyclic Peptide Inhibitor of the Interaction between Viral Open Reading Frame 3 Protein and Host Tumor Susceptibility Gene 101. *J. Virol.,* **2018**, *92*(20), e00684-00618.

[77] Srivastava, V.; Verma, P.K., The plant LIM proteins: unlocking the hidden attractions. *Planta,* **2017**, *246*(3), 365-375.

[78] Smith, M.A.; Hoffman, L.M.; Beckerle, M.C., LIM proteins in actin cytoskeleton mechanoresponse. *Trend. cell biol.,* **2014**, *24*(10), 575-583.

[79] Kadrmas, J.L.; Beckerle, M.C., The LIM domain: From the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.,* **2004**, *5*(11), 920-931.

[80] Dawid, I.B.; Breen, J.J.; Toyama, R., LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet.,* **1998**, *14*(4), 156-162.

[81] Deane, J.E.; Mackay, J.P.; Kwan, A.H.; Sum, E.Y.; Visvader, J.E.; Matthews, J.M., Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *EMBO J.,* **2003**, *22*(9), 2224-2233.

[82] Jarvinen, P.M.; Laiho, M., LIM-domain proteins in transforming growth factor beta-induced epithelial-to-mesenchymal transition and myofibroblast differentiation. *Cell. Signal.,* **2012**, *24*(4), 819-825.

[83] Sala, S.; Ampe, C., An emerging link between LIM domain proteins and nuclear receptors. *Cell. Mol. Life Sci.,* **2018**, *75*(11), 1959-1971.

[84] Zheng, Q.; Zhao, Y., The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein-protein interaction. *Biol. Cell,* **2007**, *99*(9), 489-502.

[85] Li, A.; Ponten, F.; dos Remedios, C.G., The interactome of LIM domain proteins: the contributions of LIM domain proteins to heart failure and heart development. *Proteomics,* **2012**, *12*(2), 203-225.

[86] Matthews, J.M.; Lester, K.; Joseph, S.; Curtis, D.J., LIM-domain-only proteins in cancer. *Nat. Rev. Cancer,* **2013**, *13*(2), 111-122.

[87] Tran, M.K.; Kurakula, K.; Koenis, D.S.; de Vries, C.J., Protein-protein interactions of the LIM-only protein FHL2 and functional implication of the interactions relevant in cardiovascular disease. *Biochim. Biophys. Acta,* **2016**, *1863*(2), 219-228.

[88] Liang, Y.; Bradford, W.H.; Zhang, J.; Sheikh, F., Four and a half LIM domain protein signaling and cardiomyopathy. *Biophys. Rev.,* **2018**, *10*(4), 1073-1085.

[89] Grunewald, T.G.; Butt, E., The LIM and SH3 domain protein family: structural proteins or signal transducers or both? *Mol. Cancer,* **2008**, *7*, 31.

[90] Prunier, C.; Prudent, R.; Kapur, R.; Sadoul, K.; Lafanechere, L., LIM kinases: cofilin and beyond. *Oncotarget,* **2017**, *8*(25), 41749-41763.

[91] Nam, C.H.; Lobato, M.N.; Appert, A.; Drynan, L.F.; Tanaka, T.; Rabbitts, T.H., An antibody inhibitor of the LMO2-protein complex blocks its normal and tumorigenic functions. *Oncogene,* **2008**, *27*(36), 4962-4968.

[92] Appert, A.; Nam, C.H.; Lobato, N.; Priego, E.; Miguel, R.N.; Blundell, T.; Drynan, L.; Sewell, H.; Tanaka, T.; Rabbitts, T., Targeting LMO2 with a Peptide Aptamer Establishes a Necessary Function in Overt T-Cell Neoplasia. *Cancer Res.,* **2009**, *69*(11), 4784-4790.

[93] Harrison, B.A.; Almstead, Z.Y.; Burgoon, H.; Gardyan, M.; Goodwin, N.C.; Healy, J.; Liu, Y.; Mabon, R.; Marinelli, B.; Samala, L.; Zhang, Y.L.; Stouch, T.R.; Whitlock, N.A.; Gopinathan, S.; McKnight, B.; Wang, S.L.; Patel, N.; Wilson, A.G.E.; Hamman, B.D.; Rice, D.S.; Rawlins, D.B., Discovery and Development of LX7101, a Dual LIM-Kinase and ROCK Inhibitor for the Treatment of Glaucoma. *ACS Med. Chem. Lett.,* **2015**, *6*(1), 84-88.

[94] Stradal, T.; Kranewitter, W.; Winder, S.J.; Gimona, M., CH domains revisited. *FEBS Lett.,* **1998**, *431*(2), 134-137.

[95] Bramham, J.; Hodgkinson, J.L.; Smith, B.O.; Uhrin, D.; Barlow, P.N.; Winder, S.J., Solution structure of the calponin CH domain and fitting to the 3D-helical reconstruction of F-actin:calponin. *Structure,* **2002**, *10*(2), 249-258.

[96] Lorenz, S.; Vakonakis, I.; Lowe, E.D.; Campbell, L.D.; Noble, M.E.M.; Hoellerer, M.K., Structural Analysis of the Interactions Between Paxillin LD Motifs and alpha-Parvin. *Structure,* **2008**, *16*(10), 1521-1531.

[97] Sjoblom, B.; Ylanne, J.; Djinovic-Carugo, K., Novel structural insights into F-actin-binding and novel functions of calponin homology domains. *Curr. Opin. Struct. Biol.,* **2008**, *18*(6), 702-708.

[98] Galkin, V.E.; Orlova, A.; Cherepanova, O.; Lebart, M.C.; Egelman, E.H., High-resolution cryo-EM structure of the F-actin-fimbrin/plastin ABD2 complex. *Proc. Natl. Acad. Sci. U.S.A.,* **2008**, *105*(5), 1494-1498.

[99] Klein, M.G.; Shi, W.X.; Ramagopal, U.; Tseng, Y.; Wirtz, D.; Kovar, D.R.; Staiger, C.J.; Almo, S.C., Structure of the actin crosslinking core of fimbrin. *Structure,* **2004**, *12*(6), 999-1013.

[100] Gimona, M.; Winder, S.J., The Calponin Homology (CH) Domain. *Protein Science Encyclopedia,* **2008**.

[101] Beggs, A.H.; Hoffman, E.P.; Snyder, J.R.; Arahata, K.; Specht, L.; Shapiro, F.; Angelini, C.; Sugita, H.; Kunkel, L.M., Exploring the Molecular-Basis for Variability among Patients with Becker Muscular-Dystrophy - Dystrophin Gene and Protein Studies. *Am. J. Hum. Genet.,* **1991**, *49*(1), 54-67.

[102] Roberts, R.G.; Gardner, R.J.; Bobrow, M., Searching for the 1 in 2,400,000 - a Review of Dystrophin Gene Point Mutations. *Human Mutation,* **1994**, *4*(1), 1-11.

[103] Robertson, S.P.; Twigg, S.R.; Sutherland-Smith, A.J.; Biancalana, V.; Gorlin, R.J.; Horn, D.; Kenwrick, S.J.; Kim, C.A.; Morava, E.; Newbury-Ecob, R.; Orstavik, K.H.; Quarrell, O.W.; Schwartz, C.E.; Shears, D.J.; Suri, M.;

Kendrick-Jones, J.; Wilkie, A.O.; Group, O.P.-s.D.C.C., Localized mutations in the gene encoding the cytoskeletal protein filamin A cause diverse malformations in humans. *Nat. Genet.*, **2003**, *33*(4), 487-491.
[104] Hassoun, H.; Vassiliadis, J.N.; Murray, J.; Njolstad, P.R.; Rogus, J.J.; Ballas, S.K.; Schaffer, F.; Jarolim, P.; Brabec, V.; Palek, J., Characterization of the underlying molecular defect in hereditary spherocytosis associated with spectrin deficiency. *Blood*, **1997**, *90*(1), 398-406.
[105] Kim, S.; Cullis, D.N.; Feig, L.A.; Baleja, J.D., Solution structure of the Reps1 EH domain and characterization of its binding to NPF target sequences. *Biochemistry*, **2001**, *40*(23), 6776-6785.
[106] Naslavsky, N.; Caplan, S., EHD proteins: key conductors of endocytic transport. *Trends cell biol.*, **2011**, *21*(2), 122-131.
[107] Ioannou, M.S.; Marat, A.L., The Role of EHD Proteins at the Neuronal Synapse. *Sci. Signal.*, **2012**, *5*(221).
[108] Miliaras, N.B.; Wendland, B., EH proteins: multivalent regulators of endocytosis (and other pathways). *Cell Biochem. Biophys.*, **2004**, *41*(2), 295-318.
[109] de Beer, T.; Hoofnagle, A.N.; Enmon, J.L.; Bowers, R.C.; Yamabhai, M.; Kay, B.K.; Overduin, M., Molecular mechanism of NPF recognition by EH domains. *Nat. Struct. Biol.*, **2000**, *7*(11), 1018-1022.
[110] de Beer, T.; Carter, R.E.; Lobel-Rice, K.E.; Sorkin, A.; Overduin, M., Structure and Asn-Pro-Phe binding pocket of the Eps15 homology domain. *Science*, **1998**, *281*(5381), 1357-1360.
[111] Kamens, A.J.; Mientkiewicz, K.M.; Eisert, R.J.; Walz, J.A.; Mace, C.R.; Kritzer, J.A., Thioether-stapled macrocyclic inhibitors of the EH domain of EHD1. *Bioorg. Med. Chem.*, **2018**, *26*(6), 1206-1211.
[112] Kamens, A.J.; Eisert, R.J.; Corlin, T.; Baleja, J.D.; Kritzer, J.A., Structured Cyclic Peptides That Bind the EH Domain of EHD1. *Biochemistry*, **2014**, *53*(29), 4758-4760.
[113] Khan, Z.; Lafon, M., PDZ domain-mediated protein interactions: therapeutic targets in neurological disorders. *Curr. Med. Chem.*, **2014**, *21*(23), 2632-2641.
[114] Fanning, A.S.; Anderson, J.M., Protein-protein interactions: PDZ domain networks. *Curr. Biol.*, **1996**, *6*(11), 1385-1388.
[115] Ranganathan, R.; Ross, E.M., PDZ domain proteins: scaffolds for signaling complexes. *Curr. Biol.*, **1997**, *7*(12), R770-773.
[116] Hata, Y.; Nakanishi, H.; Takai, Y., Synaptic PDZ domain-containing proteins. *Neurosci Res*, **1998**, *32*(1), 1-7.
[117] Fan, J.S.; Zhang, M., Signaling complex organization by PDZ domain proteins. *Neurosignals*, **2002**, *11*(6), 315-321.
[118] Jelen, F.; Oleksy, A.; Smietana, K.; Otlewski, J., PDZ domains - common players in the cell signaling. *Acta Biochim. Pol.*, **2003**, *50*(4), 985-1017.
[119] Lee, H.J.; Zheng, J.J., PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal*, **2010**, *8*, 8.
[120] Saras, J.; Heldin, C.H., PDZ domains bind carboxy-terminal sequences of target proteins. *Trends Biochem. Sci.*, **1996**, *21*(12), 455-458.
[121] Kim, E.; Sheng, M., PDZ domain proteins of synapses. *Nat Rev Neurosci*, **2004**, *5*(10), 771-781.
[122] Garner, C.C.; Nash, J.; Huganir, R.L., PDZ domains in synapse assembly and signalling. *Trends cell biol.*, **2000**, *10*(7), 274-280.
[123] Ponting, C.P.; Phillips, C.; Davies, K.E.; Blake, D.J., PDZ domains: targeting signalling molecules to sub-membranous sites. *Bioessays*, **1997**, *19*(6), 469-479.
[124] Ivarsson, Y., Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett.*, **2012**, *586*(17), 2638-2647.
[125] Harris, B.Z.; Lim, W.A., Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.*, **2001**, *114*(Pt 18), 3219-3231.
[126] Nourry, C.; Grant, S.G.; Borg, J.P., PDZ domain proteins: plug and play! *Sci. STKE*, **2003**, *2003*(179), re7.
[127] Zhang, M.; Wang, W., Organization of signaling complexes by PDZ-domain scaffold proteins. *Acc. Chem. Res.*, **2003**, *36*(7), 530-538.
[128] Ivanov, A.S.; Gnedenko, O.V.; Molnar, A.A.; Mezentsev, Y.V.; Lisitsa, A.V.; Archakov, A.I., Protein-Protein Interactions as New Targets for Drug Design: Virtual and Experimental Approaches. *J. Bioinf. Comput. Biol.*, **2007**, *5*(2b), 579-592.
[129] Fanning, A.S.; Lye, M.F.; Anderson, J.M.; Lavie, A., Domain swapping within PDZ2 is responsible for dimerization of ZO proteins. *J. Biol. Chem.*, **2007**, *282*(52), 37710-37716.
[130] Grillo-Bosch, D.; Choquet, D.; Sainlos, M., Inhibition of PDZ domain-mediated interactions. *Drug Discov. Today Technol.*, **2013**, *10*(4), e531-540.
[131] Hori, K.; Ajioka, K.; Goda, N.; Shindo, A.; Takagishi, M.; Tenno, T.; Hiroaki, H., Discovery of Potent Disheveled/Dvl Inhibitors Using Virtual Screening Optimized With NMR-Based Docking Performance Index. *Front. Pharmacol.*, **2018**, *9*, 983.
[132] Shan, J.F.; Zhang, X.X.; Bao, J.; Cassell, R.; Zheng, J.J., Synthesis of Potent Dishevelled PDZ Domain Inhibitors Guided by Virtual Screening and NMR Studies. *Chem. Biol. Drug Des.*, **2012**, *79*(4), 376-383.
[133] Thorsen, T.S.; Madsen, K.L.; Rebola, N.; Rathje, M.; Anggono, V.; Bach, A.; Moreira, I.S.; Stuhr-Hansen, N.; Dyhring, T.; Peters, D.; Beuming, T.; Huganir, R.; Weinstein, H.; Mulle, C.; Stromgaard, K.; Ronn, L.C.B.; Gether, U., Identification of a small-molecule inhibitor of the PICK1 PDZ domain that inhibits hippocampal LTP and LTD. *Proc. Natl. Acad. Sci. U.S.A.*, **2010**, *107*(1), 413-418.
[134] Saupe, J.; Roske, Y.; Schillinger, C.; Kamdem, N.; Radetzki, S.; Diehl, A.; Oschkinat, H.; Krause, G.; Heinemann, U.; Rademann, J., Discovery, Structure-Activity Relationship Studies, and Crystal Structure of Nonpeptide Inhibitors Bound to the Shank3 PDZ Domain. *ChemMedChem*, **2011**, *6*(8), 1411-1422.
[135] Bach, A.; Clausen, B.H.; Moller, M.; Vestergaard, B.; Chi, C.N.; Round, A.; Sorensen, P.L.; Nissen, K.B.; Kastrup, J.S.; Gajhede, M.; Jemth, P.; Kristensen, A.S.; Lundstrom, P.; Lambertsen, K.L.; Stromgaard, K., A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *Proc. Natl. Acad. Sci. U.S.A.*, **2012**, *109*(9), 3317-3322.
[136] Bach, A.; Chi, C.N.; Pang, G.F.; Olsen, L.; Kristensen, A.S.; Jemth, P.; Stromgaard, K., Design and Synthesis of Highly Potent and Plasma-Stable Dimeric

Inhibitors of the PSD-95-NMDA Receptor Interaction. *Angew. Chem. Int. Ed. Engl.*, **2009**, *48*(51), 9685-9689.
[137] Caillet-Saguy, C.; Maisonneuve, P.; Delhommel, F.; Terrien, E.; Babault, N.; Lafon, M.; Cordier, F.; Wolff, N., Strategies to interfere with PDZ-mediated interactions in neurons: What we can learn from the rabies virus. *Prog. Biophys. Mol. Biol.*, **2015**, *119*(1), 53-59.
[138] Babault, N.; Cordier, F.; Lafage, M.; Cockburn, J.; Haouz, A.; Prehaud, C.; Rey, F.A.; Delepierre, M.; Buc, H.; Lafon, M.; Wolff, N., Peptides Targeting the PDZ Domain of PTPN4 Are Efficient Inducers of Glioblastoma Cell Death. *Structure*, **2011**, *19*(10), 1518-1524.
[139] Hammond, M.C.; Harris, B.Z.; Lim, W.A.; Bartlett, P.A., beta strand peptidomimetics as potent PDZ domain ligands. *Chem. biol.*, **2006**, *13*(12), 1247-1251.
[140] Piserchio, A.; Salinas, G.D.; Li, T.; Marshall, J.; Spaller, M.R.; Mierke, D.F., Targeting specific PDZ domains of PSD-95: Structural basis for enhanced affinity and enzymatic stability of a cyclic peptide. *Chem. biol.*, **2004**, *11*(4), 469-473.
[141] Patra, C.R.; Rupasinghe, C.N.; Dutta, S.K.; Bhattacharya, S.; Wang, E.F.; Spaller, M.R.; Mukhopadhyay, D., Chemically Modified Peptides Targeting the PDZ Domain of GIPC as a Therapeutic Approach for Cancer. *ACS Chem. Biol.*, **2012**, *7*(4), 770-779.
[142] Vincenzi, M.; Mercurio, F.A.; Leone, M., Sam domains in multiple diseases. *Curr. Med. Chem.*, **2018**.
[143] Kim, C.A.; Bowie, J.U., SAM domains: uniform structure, diversity of function. *Trends Biochem. Sci.*, **2003**, *28*(12), 625-628.
[144] Knight, M.J.; Leettola, C.; Gingery, M.; Li, H.; Bowie, J.U., A human sterile alpha motif domain polymerizome. *Protein Sci.*, **2011**, *20*(10), 1697-1706.
[145] Neira, J.L.; Diaz-Garcia, C.; Prieto, M.; Coutinho, A., The C-terminal SAM domain of p73 binds to the N terminus of MDM2. *Biochim. Biophys. Acta Gen. Subj.*, **2019**, *1863*(4), 760-770.
[146] Mercurio, F.A.; Leone, M., The Sam Domain of EphA2 Receptor and its Relevance to Cancer: A Novel Challenge for Drug Discovery? *Curr. Med. Chem.*, **2016**, *23*(42), 4718-4734.
[147] Kukuk, L.; Dingley, A.J.; Granzin, J.; Nagel-Steger, L.; Thiagarajan-Rosenkranz, P.; Ciupka, D.; Hanel, K.; Batra-Safferling, R.; Pacheco, V.; Stoldt, M.; Pfeffer, K.; Beer-Hammer, S.; Willbold, D.; Koenig, B.W., Structure of the SLy1 SAM homodimer reveals a new interface for SAM domain self-association. *Sci. Rep.*, **2019**, *9*(1), 54.
[148] Kong, J.; Wang, M.M.; He, S.Y.; Peng, X.; Qin, X.H., Structural characterization and directed modification of Sus scrofa SAMHD1 reveal the mechanism underlying deoxynucleotide regulation. *The FEBS journal*, **2019**.
[149] Leone, M.; Cellitti, J.; Pellecchia, M., NMR studies of a heterotypic Sam-Sam domain association: the interaction between the lipid phosphatase Ship2 and the EphA2 receptor. *Biochemistry*, **2008**, *47*(48), 12721-12728.
[150] Mercurio, F.A.; Marasco, D.; Pirone, L.; Pedone, E.M.; Pellecchia, M.; Leone, M., Solution structure of the first Sam domain of Odin and binding studies with the EphA2 receptor. *Biochemistry*, **2012**, *51*(10), 2136-2145.
[151] Mercurio, F.A.; Marasco, D.; Pirone, L.; Scognamiglio, P.L.; Pedone, E.M.; Pellecchia, M.; Leone, M., Heterotypic Sam-Sam association between Odin-Sam1 and Arap3-Sam: binding affinity and structural insights. *ChemBioChem*, **2013**, *14*(1), 100-106.
[152] Wang, Y.; Shang, Y.; Li, J.; Chen, W.; Li, G.; Wan, J.; Liu, W.; Zhang, M., Specific Eph receptor-cytoplasmic effector signaling mediated by SAM-SAM domain interactions. *eLife*, **2018**, *7*.
[153] Kim, C.A.; Sawaya, M.R.; Cascio, D.; Kim, W.; Bowie, J.U., Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *J. Biol. Chem.*, **2005**, *280*(30), 27769-27775.
[154] Rajakulendran, T.; Sahmi, M.; Kurinov, I.; Tyers, M.; Therrien, M.; Sicheri, F., CNK and HYP form a discrete dimer by their SAM domains to mediate RAF kinase signaling. *Proc. Natl. Acad. Sci. U.S.A.*, **2008**, *105*(8), 2836-2841.
[155] Stafford, R.L.; Hinde, E.; Knight, M.J.; Pennella, M.A.; Ear, J.; Digman, M.A.; Gratton, E.; Bowie, J.U., Tandem SAM domain structure of human Caskin1: a presynaptic, self-assembling scaffold for CASK. *Structure*, **2011**, *19*(12), 1826-1836.
[156] Leettola, C.N.; Knight, M.J.; Cascio, D.; Hoffman, S.; Bowie, J.U., Characterization of the SAM domain of the PKD-related protein ANKS6 and its interaction with ANKS3. *BMC Struct. Biol.*, **2014**, *14*, 17.
[157] Thanos, C.D.; Goodwill, K.E.; Bowie, J.U., Oligomeric structure of the human EphB2 receptor SAM domain. *Science*, **1999**, *283*(5403), 833-836.
[158] Zhuang, G.; Hunter, S.; Hwang, Y.; Chen, J., Regulation of EphA2 receptor endocytosis by SHIP2 lipid phosphatase via phosphatidylinositol 3-Kinase-dependent Rac1 activation. *J. Biol. Chem.*, **2007**, *282*(4), 2683-2694.
[159] Kim, J.; Lee, H.; Kim, Y.; Yoo, S.; Park, E.; Park, S., The SAM domains of Anks family proteins are critically involved in modulating the degradation of EphA receptors. *Mol. Cell. Biol.*, **2010**, *30*(7), 1582-1592.
[160] Lee, H.J.; Hota, P.K.; Chugha, P.; Guo, H.; Miao, H.; Zhang, L.; Kim, S.J.; Stetzik, L.; Wang, B.C.; Buck, M., NMR structure of a heterodimeric SAM:SAM complex: characterization and manipulation of EphA2 binding reveal new cellular functions of SHIP2. *Structure*, **2012**, *20*(1), 41-55.
[161] Mercurio, F.A.; Scognamiglio, P.L.; Di Natale, C.; Marasco, D.; Pellecchia, M.; Leone, M., CD and NMR conformational studies of a peptide encompassing the Mid Loop interface of Ship2-Sam. *Biopolymers*, **2014**, *101*(11), 1088-1098.
[162] Mercurio, F.A.; Di Natale, C.; Pirone, L.; Scognamiglio, P.L.; Marasco, D.; Pedone, E.M.; Saviano, M.; Leone, M., Peptide Fragments of Odin-Sam1: Conformational Analysis and Interaction Studies with EphA2-Sam. *ChemBioChem*, **2015**, *16*(11), 1629-1636.
[163] Mercurio, F.A.; Marasco, D.; Di Natale, C.; Pirone, L.; Costantini, S.; Pedone, E.M.; Leone, M., Targeting EphA2-Sam and Its Interactome: Design and Evaluation of Helical Peptides Enriched in Charged Residues. *ChemBioChem*, **2016**, *17*(22), 2179-2188.
[164] Mercurio, F.A.; Di Natale, C.; Pirone, L.; Iannitti, R.; Marasco, D.; Pedone, E.M.; Palumbo, R.; Leone, M., The Sam-Sam interaction between Ship2 and the EphA2

receptor: design and analysis of peptide inhibitors. *Sci. Rep.*, **2017**, *7*(1), 17474.
[165] Mercurio, F.A.; Pirone, L.; Di Natale, C.; Marasco, D.; Pedone, E.M.; Leone, M., Sam domain-based stapled peptides: Structural analysis and interaction studies with the Sam domains from the EphA2 receptor and the lipid phosphatase Ship2. *Bioorg. Chem.*, **2018**, *80*, 602-610.
[166] Mercurio, F.A.; Di Natale, C.; Pirone, L.; Marasco, D.; Calce, E.; Vincenzi, M.; Pedone, E.M.; De Luca, S.; Leone, M., Design and analysis of EphA2-SAM peptide ligands: A multi-disciplinary screening approach. *Bioorg. Chem.*, **2019**, *84*, 434-443.
[167] Fraley, T.S.; Tran, T.C.; Corgan, A.M.; Nash, C.A.; Hao, J.; Critchley, D.R.; Greenwood, J.A., Phosphoinositide binding inhibits alpha-actinin bundling activity. *J. Biol. Chem.*, **2003**, *278*(26), 24039-24045.