Understanding Binding Affinity and Specificity of Modular Protein Domains: a Focus in Ligand Design for the Polyproline-Binding Families

Jose C. Martinez [1,*], Francisco Castillo [2], Javier Ruiz-Sanz [1], Javier Murciano-Calles [1], Ana Camara-Artigas [3], Irene Luque [1]

[1] Departamento de Química Física e Instituto de Biotecnología, Facultad de Ciencias, Universidad de Granada, 18071 – Granada, Spain

[2] Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Parque Tecnológico Ciencias de la Salud, Granada, Spain

[3] Departamento de Química Física, Universidad de Almería, Campus de Excelencia Internacional Agroalimentario ceiA3 y CIAMBITAL, 04120 – Almería, Spain

[*] Corresponding Author: jcmh@ugr.es

Running title: Ligand design for polyproline-binding domains

Abbreviations: SH3, Src homology 3; UEV, Ubiquitin E2 variant; EVH1, Ena/VASP homology 1; SH2, Src-homology region 2; PTB, Phospho-tyrosine binding; PDZ, Post-synaptic density protein 95, Disc-large, Zonnula Occludens; APP, avian pancreatic polypeptide; PPII, polyproline II; ITC, isothermal titration calorimetry; DSC, differential scanning calorimetry; HTS, high-throughput screening; Fab, antigen-binding fragment; ΨPro, pseudoproline;

Contents

Abstract

Within the modular protein domains there are five families that recognize proline-rich sequences: SH3, WW, EVH1, GYF and UEV domains. This Chapter reviews the main strategies developed for the design of ligands for these families, including peptides, peptidomimetics and drugs. We also describe some studies aimed to understand the molecular reasons responsible for the intrinsic affinity and specificity of these domains.

Keywords

Ligand design; SH3 domains; WW domains; EVH1 domains; GYF domains; UEV domains; polyproline peptides; protein-protein interactions; structure/function relationships in protein domains;

1. MODULAR DOMAINS FOR PROTEIN RECOGNITION

Modular protein-protein recognition domains provide a very efficient mechanism to control cellular activity, contributing to the localization of enzymes and substrates and interconnecting different signalling pathways. They are present in multi-domain proteins in variable numbers and forming different combinations. The human proteome contains thousands of these domains, which have been classified into more than 70 different families. Each family presents a compact and characteristic folding, at least one binding site to recognize the ligands, and generally small size (between 30 and 150 amino acids). They recognize specific amino acid sequences on their natural targets (Freund, Schmalz, Sticht, & Kuhne, 2008). According to the nature of such recognition sequences, these domains can be classified into different groups: the GYF (named for the presence of this conserved motif in their sequence), the EVH1 (Ena/VASP homology 1), the UEV (Ubiquitin E2 Variant), the WW (named for the presence of two highly conserved tryptophan residues) and the SH3 (Src-homology region 3) domain families recognize proline-rich sequences, whereas the SH2 (Src-homology region 2) and PTB (Phospho-tyrosine binding) families bind to peptide sequences displaying phosphorylated tyrosine residues; other families like the PDZ (Post-synaptic density protein 95, Disc-large, Zonnula Occludens) interact with sequences corresponding to carboxyl termini (Cesareni, Gimona, Sudol, & Yaffe, 2005).

Protein-protein interactions generally involve highly extensive binding interfaces, up to 6000 $Å^2$. Instead, modular protein domains are characterized by recognizing small continuous epitopes, bringing comparatively reduced functional binding interfaces into play, about 1000 $Å^2$. In this regard, the design of inhibitors of protein-protein interactions for the binding sites of modular domains can be particularly affordable. On the other hand, modular domains are specialized in transient interactions, since they are involved in cellular signal transduction processes. Consequently, the interactions mediated by these modular domains are relatively weak, with dissociation constants that typically range within the micromolar order. In addition, these domains present a certain degree of promiscuity, since the same protein can be recognized by more than one domain with comparable affinities. Such characteristics can explain why the binding pockets of modular domains are relatively flat and adaptable (Freund, et al., 2008). In this sense, the binding sites of the modular domains are antagonistic to the catalytic sites of enzymes, which are usually deep and highly structured. Therefore, although the interaction surfaces of the modular domains constitute an attractive intervention point for being relatively narrow, the design of high-quality ligands by improving their affinity and specificity is more complex than for the catalytic sites of the enzymes.

2. MODULAR DOMAINS FOR THE RECOGNITION OF PROLINE-RICH SEQUENCES

Within the modular domains there are five families specialized in the recognition of proline-rich sequences: SH3, WW, EVH1, GYF and UEV domains (Figure 1). Some authors also consider profilin, an actin-binding protein that recognizes proline-rich sequences, as the sixth family (Mahoney, Rozwarski, Fedorov, Fedorov, & Almo, 1999). The polyproline ligands are typically arranged as a polyproline II (PPII) helix. The binding mechanism is also similar, based on the interaction of aromatic residues from the domain binding surface which arrange in grooves, named as "xP", where the proline residues from the ligands dock (Figure 2). The domains contain an additional binding groove which confers specificity within a family, and intervenes in the recognition of the positions that flank the central and usually well-conserved ligand motifs. Such central motifs are characteristic of each domain family (Figure 1). Together, these elements define the intrinsic binding specificity. The proper functioning of the cellular signalling processes in which these domains participate also requires complex networks of highly selective interactions, regulated by a contextual specificity, in which factors such as sub-cellular location or the cooperative effect of multiple interactions may play an important role (Stein & Aloy, 2008; Zarrinpar, Park, & Lim, 2003).

Families are sub-classified by types based on the motif they recognize, but the molecular determinants governing intrinsic specificity within each type are unknown, which would require an individual and prioritized study (Opitz et al., 2015). In this sense, it has become particularly interesting the characterization of those proline-rich binding domains involved in the pathogenesis of viral diseases and hereditary syndromes of certain types of cancer, such as Liddle, Noone and Hopkins (Corbi-Verge & Kim, 2016).

Although the structural analysis has allowed rationalizing the main characteristics of these domain-peptide interactions, the understanding of the origins of the binding energy emerges controversial. Thus, isothermal titration calorimetry (ITC) experiments have revealed a thermodynamic signature for these interactions (very favourable enthalpic contributions opposed by an unfavourable binding entropy) inconsistent with the highly hydrophobic nature of the peptide ligands and their corresponding domain binding sites. Nevertheless, a deeper looking at the crystal structures may reveal some relevant details. Thus, for example, the complex between Abl-SH3 and the peptide p41 (APSYSPPPPP) (PDB code: 1BBZ; (Pisabarro, Serrano, & Wilmanns, 1998)) shows the establishment of a complex hydrogen-bond network mediated by several water molecules buried at the binding interface (Figure 3). This network would explain the observed thermodynamic behaviour (Palencia, Cobos, Mateo, Martinez, & Luque, 2004). The origin of the binding energetics for proline-rich ligands to the Abl-SH3 domain was further

investigated by a comparative calorimetric analysis of a set of p41-related ligands. The striking effects upon the enthalpic and entropic contributions provoked by conservative substitutions at solvent-exposed positions in the ligand confirmed the complexity of the interaction, as well as the energetic relevance that such network of water molecules posses in the interaction (Palencia, et al., 2004). The design of mutations in the Abl-SH3 domain to alter the water-mediated network confirmed such a hypothesis through thermodynamic, structural and molecular dynamics experiments. In the light of these results, a new dual binding mechanism is proposed, where a network of water-mediated hydrogen bonds complements the canonical hydrophobic interactions. The mechanism provides a better description of proline-rich ligand recognition by Abl-SH3 and has important implications for rational design (Palencia, Camara-Artigas, Pisabarro, Martinez, & Luque, 2010).

The systematic analysis of the SH3 structural database reveals that this dual binding mode is universal to SH3 domains. Tightly bound buried-interfacial water molecules mediating the interaction between the peptide ligand and the domain were found in most of the SH3 complexes studied. Moreover, structural waters were also identified in a high percentage of the free SH3 domains. Thus, the analysis enabled the identification of conserved hydration sites in the polyproline-recognition region and the establishment of relationships between hydration profiles and the sequence of both, ligands and SH3 domains (Zafra-Ruano & Luque, 2012).

Water-mediated interactions have also been systematically observed in WW, UEV and EVH1 crystal structures (Figure 3), where similar thermodynamic signatures have been observed for binding. This outcome strongly suggests that the current description of proline-rich sequence recognition by modular protein domains is incomplete and insufficient for a correct structural and energetic understanding of these systems. In the words of the authors, *a new binding paradigm is required that includes interfacial water molecules as relevant elements in polyproline-recognition domains* (Martin-Garcia, Ruiz-Sanz, & Luque, 2012).

3. HIGH-THROUGHPUT SCREENING STRATEGIES FOR THE DESIGN OF BINDING PARTNERS FOR POLYPROLINE-BINDING DOMAINS

Since the first X-ray structure of an SH3 domain was published in 1992 (Musacchio, Noble, Pauptit, Wierenga, & Saraste, 1992), the first reported of a polyproline-binding domain, the main focus has been finding potential binding partners for these modular domains (Musacchio, Gibson, Lehto, & Saraste, 1992). From a practical point of view, high-throughput screening (HTS) methodologies have contributed the most to this point, since

they allowed obtaining moderately or highly related peptide sequences for these domains. The analysis of common features among these sequences allowed knowing the binding abilities of polyproline-binding domains.

Focussing on SH3 domains, the most studied among the five families, phage display screening with fully randomized libraries revealed that the SH3 recognition sequences identified shared a conserved xPxxP motif. Step by step experimental protocols can be found elsewhere (Sparks, Adey, Quilliam, Thorn, & Kay, 1995; Tonikian, Zhang, Boone, & Sidhu, 2007). Biased libraries have been usually designed by including the consensus motif flanked by random sequences, and expanding recognition over a 12 amino acids window but exploring each library a sequence space of six amino acids (as an example: Gx_6G, x_6PPIP, $RSLRPLx_6$, $PPPYPPx_6$; (Rickles et al., 1994)). Screens of such bias against a representative number of SH3 domains effectively converged towards domain-specific sequences. Thus, each SH3 domain has a preferred sequence: Src-SH3 the sequence xxxRPLPPLPxP, Fyn xxxRPLPP(I/L)Pxx, Lyn RxxRPLPPLPxP, PI3K RxxRPLPPLPPP, while Abl-SH3 selects phages containing the sequence PPPYPPPP(I/V)Pxx (Rickles, et al., 1994). Subsequent studies demonstrated that these high-affinity peptide ligands may serve as new tools to interfere with the cellular functions of SH3-mediated processes and constitute the basis for the design of SH3-specific inhibitors of disease pathways (Cheadle et al., 1994; Rickles et al., 1995; Sparks, Quilliam, Thorn, Der, & Kay, 1994).

Parallel studies with WW domains have also yielded a similar perspective. For example, a comprehensive screening of 42 WW domains against several cellulose-bound peptide libraries, including designed peptides and peptides based on naturally occurring sequences, and phosphorylated residues, was carried out. Thirty-two WW domains were classified into six groups according to detected ligand recognition preferences for binding the motifs PPx(Y/pY), (p/ϕ)P(p,g)PPpR, (p/ϕ)PPRgpPp, PPLPp, (p/ψ)PPPPP, and (pS/pT)P (Kasanov, Pirozzi, Uveges, & Kay, 2001; Linn et al., 1997; Otte et al., 2003).

The use of biased phage display libraries has been especially informative when applied to the c-Src-SH3 domain. Thus, Feng et al. (Feng, Kasahara, Rickles, & Schreiber, 1995) found two dodecapeptides that are oriented oppositely at the binding site (defined as class I and class II). Both peptide sequences are constituted by: *i)* a core proline-rich sequence arranged as a PPII helix; and *ii)* a flanking sequence that occupies a large pocket between the RT and n-Src loops of the SH3 domain (Figure 1). However, detailed structural and mutational analyses illustrate how the two flanking sequences exploit the binding surface differently to increase the binding affinity and specificity (Bacarizo & Camara-Artigas, 2013; Bacarizo, Martinez-Rodriguez, & Camara-Artigas, 2015; Camara-Artigas, Ortiz-Salmeron, Andujar-Sanchez, Bacarizo, & Martin-Garcia, 2016; Martin-Garcia,

Luque, Ruiz-Sanz, & Camara-Artigas, 2012). Interestingly, a similar study carried out with Lck-SH3 has shown that phage-display selected peptides represented mainly class I ligands, whereas the comparatively diverse Lck-SH3 binding sites of all analyzed natural binding partners emerged as class II proteins. An explanation for the observed variations between artificial and native ligands, which are not due to significant differences in affinity, suggest that phage display may not irrevocably mirror physiologically relevant domain-ligand interactions (Tran et al., 2005).

Definitely, these studies have provided relevant information at a molecular level on the interactions responsible for the binding affinity and specificity of SH3 and WW domains. In addition, the high-affinity ligands (sometimes improved over 40-fold compared to natural affinities) can be used as tools for the *in vivo* and *in vitro* blocking of natural protein-protein interactions. Nevertheless, as we have shown above, these analyses have produced ligands that in some cases resemble the natural ones but, in many others, such artificial binding motifs clearly differ from them (Cestra et al., 1999; Fazi et al., 2002; Mongiovi et al., 1999).

To overcome these discrepancies, Tong et al. (Tong et al., 2002) describe how to combine phage display and cDNA yeast-two-hybrid results to produce highly reliable interaction predictions, including both artificial and natural sequences respectively. Such predictions were achieved computationally, by identifying potential natural binding partners within the yeast proteome for twenty selected yeast SH3 domains assayed experimentally and, thus, generating a possible natural network of interactions for these SH3 domains. Therefore, to map natural interaction networks, phage display improves when combined with other approaches (Sidhu, Bader, & Boone, 2003). An interesting alternative would be the cloning and further screening of a cDNA library into the surface of a bacteriophage, which profits the main advantages of both approaches (Kurakin & Bredesen, 2002).

The SPOT (Specificity Prediction Of Target) algorithm emerged as an attempt of integrating these “irrational” high-throughput approaches with “rational” structural and energetic information. SPOT was also developed to overcome the limitation of phage display studies which assume that residues contribute individually to the binding. Thus, the sequence alignment by means of tools as WebLogo (https://weblogo.berkeley.edu/) allows knowing the frequency of residues at individual positions; the overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position (Figure 4). Nevertheless, it is well known that correlation among definite positions can exist, and the selection of one amino acid at a definite position can

determine the selection at another one (Stein & Aloy, 2008). Thus, SPOT would predict peptide recognition specificity of any SH3 domain from its sequence. The known structures of SH3/peptide complexes and the domain sequence alignments are used to rationalize the experimental data obtained by panning peptide repertoires displayed in bacteriophages. The results of panning experiments are organized in contact matrices which describe the frequency of occurrence of any specific pair of residues in the SH3 and in the ligand peptide for each SH3/peptide contact position. The matrices are then used to evaluate the probability of interaction between different SH3 domains and a given peptide. Therefore, SPOT can be applied to any family of proteins for which both structural information about at least one protein/peptide complex and energetic data between a few domain/peptide sequences may exist. In our particular example, SPOT was designed to predict the preferred ligand of SH3 domains whose specificity was not known experimentally (Brannetti, Via, Cestra, Cesareni, & Helmer-Citterich, 2000). Today, the development of microarrays technology allows the synthesis and screening of thousands of peptide sequences arranged on cellulose membrane supports (Kramer & Schneider-Mergener, 1998).

SPOT analysis has also been developed for other polyproline-binding domains. Thus, the phage-display-derived recognition motif for CD2BP2-GYF is PPG(W/F/Y/M/L) and was confirmed by SPOT analysis too, which also confirmed that the GYF domain interacts with peptides from human proteins containing a consensus polyproline motif (Figure 1). Epitope mapping by NMR spectroscopy performed for several peptides showed a conserved binding surface. The new interacting partners discovered were verified by yeast two-hybrid analysis (Kofler, Motzny, Beyermann, & Freund, 2005).

In the case of WW domains, and mainly due to their small size, SPOT platforms were developed to analyze the structure–function relationship of FBP28-WW domain through a complete substitution screening within the domain, with variants synthesized as a cellulose-bound peptide array. The functionality of the FBP28-WW domain variants was examined by probing a peptide array of potential ligands (Przezdziak et al., 2006). A similar approach has been used to select a high-affinity WW domain against the extracellular region of VEGF receptor isoform-2 from a combinatorial library using a cell-free molecular display platform. The isolate has nanomolar affinity to VEGFR-2 and inhibits the binding of human VEGF to its receptor (with nanomolar affinity as well). The structure is amenable to cyclisation to improve its proteolytic stability and has the clear advantage over larger protein scaffolds of being synthesised chemically to high yields, offering the potential for therapeutic and non-therapeutic applications (Patel, Mathonet, Jaulent, & Ullman, 2013).

In any case, the strategy of designing artificial modular protein domains with definite affinities and specificities can also be achieved if the randomized residues in the domain are those responsible for these binding characteristics. Thus, SH3 binding properties can be profoundly altered by sequence modifications in the RT-loop and n-Src-loop regions (Bauer & Sticht, 2007; Hiipakka & Saksela, 2007). Going even further, Freund et al. have developed a novel SH3 domain scaffold engineered to bind the CD2BP2-GYF domain and the HIV capsid protein. Phage display screens resulted in the isolation of variants with changes in all randomized positions compared to wild-type SH3. The soluble scaffolds bind with 340 and 600 nM affinity to HIV capsid protein and CD2BP2-GYF, respectively. Similar approaches have been designed to target and inhibit molecular interactions in the context of disease, with both diagnostic and therapeutic potential. Apart from antibodies, several alternative scaffolds have been exploited over the years making use of the fact that individual domain families are best suited for certain target families (Piotukh & Freund, 2012). Antibodies for SH3 domains have also been obtained by phage display of an antigen-binding fragment (Fab) library. The authors have established a rapid and robust HTS methodology that can be applied to generate highly functional and renewable antibodies targeting protein domains on a proteome-wide scale. Affinity assays demonstrated that representative Fabs bind tightly and specifically to their targets. The Fabs were tested in common cell biology applications and confirmed recognition of the full-length antigen in immunoprecipitation, immunoblotting and immunofluorescence assays (H. Huang et al., 2015).

Particularly interesting would be the work carried out by Ranganathan et al., aimed at designing *de novo* artificial WW domains. Their method, termed statistical coupling analysis (SCA), does not use structural or physicochemical information but instead extracts information about essential patterns of amino acids from the evolutionary record. The statistical information used to compute the artificial WW domains is just the encoded in a multiple sequence alignment, where the computer extracts simple statistical energy functions capturing co-evolution between amino acid residues. The artificial proteins showed thermodynamic stabilities and structural arrangements similar to natural WW domains (Socolich et al., 2005). In addition, the authors demonstrated that these artificial WW sequences function like their natural counterparts, showing class-specific recognition of proline-containing target peptides. Moreover, they identified the network of residues responsible for the functional specificity in WW domains. This work demonstrates that a relatively small quantity of sequence information is sufficient to specify the global energetics of amino acid interactions (Russ, Lowery, Mishra, Yaffe, & Ranganathan, 2005).

Definitely, the computational integration of several screening approaches, together with the technological development which allows them to expand the diversity

of sequences to be explored, emerges as the most promising strategy to obtain integrative information on the specificity networks of modular protein domains (Teyra, Sidhu, & Kim, 2012). Coming back to SH3 domains, the combination of phage display, yeast two-hybrid and peptide array screening to independently identify SH3 domain binding partners, followed by the integration of all results using a Bayesian algorithm has allowed the authors to generate a high-confidence yeast SH3 domain interaction map. In addition, this analysis has permitted the prediction of new functionalities for concrete SH3 domains (Tonikian et al., 2009). An additional dimension that combined approaches can include in the protein/peptide binding studies would be the evolution analysis among species of these interaction networks. Specifically, an interesting study carried out in several yeast species points towards a remarkably conserved specificity profile for SH3 domains over a large evolutionary distance. The goal of these kinds of studies would be the unravelling of complex specificity networks of SH3 and other peptide recognition domains in higher eukaryotes, including mammals (Teyra et al., 2017; Verschueren et al., 2015).

On the other hand, the computational rationalization of HTS results has allowed the assembling of a database for peptide recognition modules (http://www.prm-db.org) that will enable many structural, functional, and biological studies (Teyra et al., 2020). These databases will also be enough information to map the specificity profiles in detail, including derived-position weight matrices and binding specificity logos based on multiple peptide ligands. These studies conclude that optimal peptide ligands resemble peptides observed in the existing structures of domain/ligand complexes, indicating that a large majority of the phage-derived peptides are likely to target natural peptide-binding sites and could thus act as inhibitors of natural protein-protein interactions.

3.1. High-Throughput Screening of Cyclic Peptides

The life cycle of enveloped viruses requires extensive assistance from the host cell proteins and pathways. Tsg101 is one of the cellular proteins involved in the budding process of virions. In this process, Tsg101 is recruited from the internal site of the infected cell to the budding site to aid in the release of the HIV-1 virus particles. Depletion of Tsg101 from virus-producing cells can lead to a budding defect. Therefore, Tsg101 is a potentially attractive target for therapeutic intervention in several viral infections. VP40 and Gag matrix proteins are the key viral proteins that drive the budding process by mediating specific virus/host interactions to facilitate the efficient release of virions from the infected cell, concretely, the interaction between some motifs displayed by these proteins (generically named as viral L domains), which contain the consensus sequences

PTAP and PPxY (Figure 1), with the human domains Tsg101-UEV and Nedd4-WW3 respectively (Harty, 2009).

Although the development of inhibitors of the interaction between viral L domains and their cellular target domains will drive the development of wide-spectrum antivirals, few strategies have been developed to discover functional compounds. Currently, these strategies are in their early stages, far from providing molecules with pharmacologically optimal properties and/or in the clinical phase. The works published during the last decade by Professors Tavassoli and Harty demonstrate that it is possible to block with small molecules these interactions. Specifically, Tavassoli et al. have published a strategy for the *in vivo* screening, in *Escherichia coli*, of libraries of cyclic peptides capable of inhibiting the interaction between the TSG101-UEV domain and the L-domain PTAP motif of the p6 region of the HIV-Gag protein. Cyclic peptides have a greater survival time than linear, which increases their therapeutic applicability. This methodology allowed the identification of cyclic penta-peptides that lack the canonical P(T/S)xP motif recognized by Tsg101-UEV. These inhibitors have an alternative specificity that allows them to block the release of HIV viral particles in a cellular model, without interfering in the cellular traffic mediated by the interaction between Tsg101-UEV and the PTAP motif of the cellular protein HGS (Hepatocyte growth factor-regulated tyrosine kinase substrate). Thus, these findings suggest that small molecule therapeutics addressed to inhibit specific interactions between viral and host proteins may have general applicability in antiviral therapy (Lennard, Gardner, Doigneaux, Castillo, & Tavassoli, 2019; Tavassoli et al., 2008). More recently, cyclic peptide inhibitors of the Hepatitis E virus have been found by following a similar strategy (Anang et al., 2018).

3.2. High-Throughput Screening of Drug Libraries

On the other hand, Harty et al. have carried out the computational screening of virtual libraries of commercially available molecules to identify inhibitors of the interaction between the viral L domains PTAP and PPxY with their respective cellular targets, Tsg101-UEV and Nedd4-WW3. They have also actively worked on developing assays in human cells that allowed them to validate the biological activity of the primary compounds identified in their *in silico* screening campaigns. These efforts have given promising results, showing that some of these compounds are capable of blocking the cellular release of several encapsulated viruses (Han et al., 2014).

More recently, a simple, robust, and reliable HTS strategy based on an enzyme-linked immunosorbent assay (ELISA) has allowed identifying compounds that inhibit HIV-1

budding, also by targeting the Tsg101-UEV/Gag interaction. Several hits were identified through screening of a 9600-compound library. Subsequent assays revealed some lead compounds, HSM-9 and HSM-10 (Siarot et al., 2018), or FC-10696 (Han et al., 2021), which have antiviral activity.

4. RATIONAL DESIGN OF LIGANDS FOR POLYPROLINE-BINDING DOMAINS

Parallel to the high-throughput studies, rational design exercises have been carried out with polyproline-binding domains. Although these domains are quite promiscuous, a pioneering work by Pisabarro and Serrano demonstrated that it is possible to design mutations into a peptide ligand to improve their selectivity by using existing biocomputing tools and simple physicochemical reasoning. Thus, they designed mutations in a previously found peptide ligand, able to bind Abl-SH3 and Fyn-SH3 with similar affinity, so that the affinity for Abl-SH3 increased 20-fold, while that for Fyn-SH3 decreased 10-fold. This work demonstrated that both the RT and n-Src loops are responsible for regulating the specificity for proline-rich ligands. Thus, the first N-terminal positions in the peptide ligand would be important for determining the specificity for these SH3 domains, while the remaining at the C-terminus seem to be more important for the affinity and organize the classical PPII helix characterizing such ligands (Figure 2) (Pisabarro & Serrano, 1996). A later analysis showed that for the above examples phage display approaches are in full agreement with rational design, but combinatorial peptide library screens may help rational design by improving affinity and specificity of definite regions, reporting new amino acids that were not selected by phage display and rational design (Santamaria, Wu, Boulegue, Pal, & Lu, 2003).

The ability of certain SH3 domains to bind specifically both class I and class II polyproline ligands (opposite orientations) can be designed as well. A detailed mutational and structural analysis of Fyn-SH3 domain revealed that the conserved Trp in the binding pocket of this family can adopt two different orientations that, in turn, determine the type of ligand (class I or class II) able to bind to the domain. The only exceptions would be ligands that deviate from standard polyproline angles. The motion of the conserved Trp depends on the presence of certain residues located in a key position, near the binding pocket. SH3 domains placing aromatic residues in such key position are promiscuous; by contrast, those presenting beta-branched or long aliphatic residues block the conserved Trp in one of the two possible orientations, preventing binding in a class I orientation. This was experimentally demonstrated by a single mutation in such a key position in Fyn-SH3 (Y132I) that effectively abolished class I ligand binding, while preserving binding to class II ones. This work demonstrates that conformational changes having noticeable effects on

protein-protein interactions are governed by simple rules, highlighting the importance of structural and energetic details to predict protein-protein interactions (Fernandez-Ballester, Blanes-Mira, & Serrano, 2004).

4.1. Strategies to Replace the Highly-Conserved Proline Residues without Altering the Polyproline II Helical Structure of the Peptide Ligands

Since the conformational arrangement of the ligands has revealed as a well-conserved feature of their successful interactions with polyproline-binding domains, a great effort has been invested into the natural or artificial replacement of Pro residues within these sequences which, actually, can encompass a significant part of the sequence. As was already demonstrated by Pisabarro and Serrano (Pisabarro & Serrano, 1996; Pisabarro, Serrano, & Wilmanns, 1998), it is possible to replace some of the Pro residues, postulated to be essential for the interaction with SH3 domains, and still retain a significant affinity in the peptide ligand. Of course, such replacement has to be rationally designed. This fact indicates that the sequence repertoire that could interact with a specific SH3 domain could be larger than previously thought. Phage display may not be helpful in this way, since the highly-conserved xP pockets in the domains have been evolutionarily designed to accommodate Pro residues.

4.1.1. Design of Miniprotein Scaffolds to Assemble the Polyproline II Binding Epitope

An important limitation to all the above design strategies is that the PPII conformation needed to bind to a polyproline-binding domain is maintained in peptides by the presence of the already mentioned highly-conserved Pro residues. Thus, although other amino acids might confer higher specificity and/or affinity to the ligand, Pro residues cannot be mutated because of entropic destabilization (Palencia, et al., 2004; Pisabarro & Serrano, 1996). An alternative strategy to simple mutagenesis will be the design of a stable polypeptide scaffold within which the PPII conformation is preserved by residues of the scaffold itself. Thus, a suitable universal scaffold aimed at any polyproline-binding domain should have a region adopting a PPII conformation. The family of the avian pancreatic polypeptide (APP) is suitable for these purposes. APP is a small protein of 36 residues with a fold consisting of an α-helix packed against a PPII helix, without Cys residues (Figure 5) (Blundell, Pitts, Tickle, Wood, & Wu, 1981). On the solvent-exposed face of the PPII helix there are no Pro residues, which confirm that the PPII conformation is determined by packing against the α-helix. In fact, the interaction between the PPII helix and the α-helix forms a small hydrophobic core, established by an interdigitation of three Pro side-groups of the PPII helix between the mainly non-polar side-chains of one face of

the amphipathic α-helix. In this situation, the exposed residues in contact with the polyproline-binding domain can be replaced by other amino acids or analogues without affecting the PPII conformation.

This miniprotein has been extensively used by Schepartz et al. for the development of peptide ligands for several proteins, optimizing concrete residues on the structure by phage display (Chin, Grotzfeld, Fabian, & Schepartz, 2001). Thus, by optimizing the solvent-exposed residues of the α-helix, they have obtained APP-derivatives with DNA binding properties, as ligands that overcome the protective effects associated with up-regulation of anti-apoptotic Bcl-2 proteins, or phosphorylated peptide ligands that recognize the surface of CBP KIX domain; always achieving high affinities (nanomolar to low micromolar). In addition, since the natural APP miniprotein assembles as a dimer, Schepartz et al. identified, characterized, and replaced two structural elements responsible for APP dimerization by proline switches that single-handedly repack APP signature fold. The result was a monomeric and well-folded miniature protein that may serve as a starting point for the *in vitro* and *in vivo* applications of these molecules (Hodges & Schepartz, 2007).

By using the monomeric version of APP and by following a similar optimization strategy (named by the authors as *protein grafting* (Chin, et al., 2001)), but exploring the PPII helix, they have developed pGolemi, which binds EVH1 domains. pGolemi is monomeric at 10^{-4} M concentration and well-folded, despite the fact that primary sequence differences exceed 50% from wild-type APP. This designed miniprotein binds with high affinity to the EVH1 domain of Mena (Mena1-112) but not to those of VASP (VASP1-115) or Evl (Evl1-115), and also causes an unusual defect in actin-driven *Listeria monocytogenes* motility. Scanning mutagenesis was used to rationalize affinity and specificity, miniature protein secondary structure, and *L. monocytogenes* motility. The NMR analysis confirmed that pGolemi contains the expected APP-like fold and binds Mena1-112 in a manner highly analogous to the proline-rich repeat region of *L. monocytogenes* ActA protein. Thus, this peptide shares the common pancreatic peptide fold with its scaffold, but shows key differences at the N-terminus. The interplay of spatial fixation and flexibility appears to be the reason for its high affinity towards Mena-EVH1. Together with earlier investigations, these structural data shed light also on the specificity determinants of pGolemi and the importance of additional binding epitopes around the residues Thr74 and Phe32 on EVH1 domains regulating paralog specificity (Holtzman, Woronowicz, Golemi-Kotra, & Schepartz, 2007; Link, Hunke, Mueller, Eichler, & Bayer, 2009).

The application of this strategy may help the design of ligands for other polyproline-recognition domains, and even for the *in vivo* application of these miniproteins. In fact, in our research group we rationally designed a structure in which some residues of the APP PPII helix were replaced by a sequence motif, named RP1 (APSYPPPPP), which interacts with the Abl-SH3 domain. This design, APP-RP1, was well folded and, as shown by circular dichroism, its structural content was similar to that of natural APP. The stability of both miniproteins was compared by unfolding experiments; surprisingly, the designed APP-RP1 is almost 20 ºC more thermostable than the wild-type and has a higher Gibbs energy change at 25 ºC too. This increase in stability has an entropic origin. Isothermal titration calorimetry and fluorescence spectroscopy showed that the binding thermodynamics of the APP-RP1 molecule to Abl-SH3 is comparable to that of the shorter RP1 peptide. Furthermore, the mutation by Tyr of the two Pro residues from the consensus xPxxP motif in APP-RP1, demonstrates the effectiveness of the scaffold in enhancing the variability in the design of high-affinity and high-specificity ligands for any SH3 domain (Cobos et al., 2004).

4.1.2. Design of Peptidomimetics

An alternative way to achieve the replacement of Pro residues will be the development of peptidomimetic compounds. This approach is also convenient for developing therapeutic molecules, since natural peptides may suffer natural proteolysis and their cellular penetration is also limited when used *in vivo*.

The cycle of protein-structure-based combinatorial chemistry followed by structure determination of the highest affinity ligands constitutes a powerful tool to discover ligands containing non-peptide binding elements to polyproline-binding domains. In the case of c-Src-SH3 the encoded library used had the form Cap-M1-M2-M3-PLPPLP, in which the Cap and Mi's were composed of a diverse set of organic monomers. The PLPPLP portion provided a structural bias directing the non-peptide fragment Cap-M1-M2-M3 to the SH3 specificity pocket. Fifteen ligands were selected from more than one million of different compounds. The NMR solution structures of the c-Src-SH3 domain complexed with two of such selected ligands showed that the non-peptide moieties of the ligands interact with the specificity pocket of c-Src-SH3, as predicted, but differently to known peptides complexed with this and other SH3 domains. Structural information about the ligands was used to design various homologues, whose affinities for the SH3 domain were measured (Feng, Kapoor, Shirai, Combs, & Schreiber, 1996). The Leu-Pro binding pocket of Src-SH3 was also explored by structure-based split-pool synthesis to discover non-peptide binding elements. Binding characteristics of the protein xP pocket were then explored by comparing a series of ligands that contain subtle variants of the parent non-peptide

binding structure. Further insights into this receptor/ligand interaction were provided by multidimensional NMR structure determination of one of the non-peptide ligands (Morken, Kapoor, Feng, Shirai, & Schreiber, 1998).

A significant advance in the design of peptoids for SH3 and WW domains was achieved by Lim et al. (Nguyen, Turck, Cohen, Zuckermann, & Lim, 1998) by exploring amide N-substituted Pro residues. These domains broadly accept amide N-substituted residues, being Pro the only endogenous N-substituted amino acid. This discriminatory mechanism may explain how these domains achieve specific but low-affinity recognition, a property that is necessary for transient signalling interactions. The authors have exploited the mechanism by screening a series of ligands in which some key prolines were replaced by non-natural N-substituted residues. The results yielded a ligand that selectively bound the Grb2-SH3 domain with 100-times greater affinity. Moreover, the replacement of Pro residues by the N-substituted ones might be a new source to improve ligand specificity (Nguyen, et al., 1998). The same authors have tested the effects of combining multiple peptoid substitutions with specific flanking sequences on ligand affinity and specificity, showing that ligands can be selectively tuned to target a single SH3 domain. In addition, by making multiple substitutions, high-affinity ligands can be generated so that they lack entirely the canonical xPxxP motif. The resulting ligands can potently disrupt natural SH3-mediated interactions. Thus, these hybrid scaffolds yield SH3 ligands with markedly improved domain selectivity, overcoming one of the main challenges in designing inhibitors against these domains. Definitely, these compounds represent important leads in the search for orthogonal inhibitors of SH3 domains (Nguyen et al., 2000).

Even more, it has been published elsewhere (Vidal et al., 2004) the design of the highest affinity peptidomimetic ligands reported so far for the Grb2-SH3 domain. These compounds were designed by combining *N*-alkyl amino acid incorporation in a proline-rich sequence with subsequent dimerization of the peptoid sequence based on structural data and molecular modelling. Because the affinity for Grb2 of the optimized compounds was too high to be measured using the fluorescent modifications that they induce on the Grb2 emission spectrum, a competition assay was developed. In this test, Grb2 was pulled down from a cellular extract by an initial VPPPVPPRRR peptide bound to Sepharose beads. In the presence of competitors, the test quantified the amount of Grb2 displaced from the beads. This methodology enabled the determination of a K_d value in the 10^{-10} M range for the highest affinity Grb2 peptoid analogue dimer (Vidal, et al., 2004).

The N-substitution of amino acids has also been demonstrated as useful for improving some natural ligand interactions. As an example, the phospholipase Cγ1 (PLCγ1) has a regulatory SH3 domain that can specifically recognize and interact with a defined

xPxxP-containing decapeptide segment (185-QP**P**VP**P**QRPM-194) coming from the adaptor protein SLP76. The isolated peptide binds the PLCγ1-SH3 domain with a moderate affinity due to the lack of protein context support. The peptide affinity was noticeably improved by replacing the two key Pro residues, Pro187 and Pro190, of the xPxxP motif with non-natural N-substituted amino acids. Two N-substituted peptides, N-Leu187/N-Gln190 and N-Thr187/N-Gln190, were designed that improved affinity by, respectively, 8.5-fold and 3.4-fold (K_d = 0.67 ± 0.18 and 1.7 ± 0.3 μM, respectively) relative to native peptide (K_d = 5.7 ± 1.2 μM) (Tang, Zhao, Wang, Ye, & Yang, 2019).

The incorporation of substituents of variable size and polarity at the C2 position of pseudoprolines (ΨPro) offers another powerful tool for the design of non-peptide ligands. ΨPro consist of Ser-, Thr-, or Cys-derived proline-like structures with enhanced inherent properties of natural L-Pro (Tuchscherer et al., 2001). ΨPro building blocks exert a dual functionality in: *i)* inducing and stabilizing the relevant PPII conformation; and *ii)* increasing and optimizing the van der Waals contacts and setting hydrogen bonds with the receptor. Most notably, the generation of a library of different substituents at C2 allows the factors contributing to affinity and specificity in protein-protein interactions to be explored, and to further elucidate ligand recognition by polyproline-binding domains at a molecular level (Tuchscherer, et al., 2001).

WW domains have also been used as targets for peptidomimetics design. Apart from the Yap-WW domain studied by Lim et al. (Nguyen, et al., 1998), the Pin1-WW domain has been extensively used, since is involved in cancer developments and is considered as a pharmaceutical target. Thus, finding a high-affinity inhibitor of Pin1 has become an attractive topic. The WW domain of human Pin1 can recognize phosphoserine/phosphothreonine-proline (pS/pT)P motifs. A series of 4-substituted proline derivatives were incorporated into the phosphopeptides and investigated their affinities for the WW domain of Pin1. Isothermal titration calorimetry and fluorescence anisotropy analyses showed that the replacement of Pro with (2S,4R)-4-fluoroproline increased the binding affinity of the peptides, whereas circular dichroism measurements suggested that a more PPII-like structure of phosphopeptides would be the reason. Chemical shift perturbation experiments also indicated that (2S,4R)-4-fluoroproline interacts with Trp34 of the WW domain in the binding site, revealing a strong C–H···π interaction (K. Y. Huang & Horng, 2015).

To explore the druggability of these modular protein domains, a modular toolkit has been developed to generate small molecules able to compete with defined domain/peptide interactions (Opitz, et al., 2015). The modular strategy collects a comprehensive toolkit of chemical fragments (ProMs) designed to replace pairs of

conserved Pro residues in recognition motifs. The authors developed a small, selective, peptidomimetic inhibitor of Ena/VASP-EVH1 domain interactions as proof of principle. Highly invasive MDA MB 231 breast-cancer cells treated with this ligand showed displacement of VASP from focal adhesions, as well as from the front of lamellipodia, and strongly reduced cell invasion. Based on this previously developed non-peptidic micromolar inhibitor, the authors determined 22 crystal structures of Ena-EVH1 in complex with different inhibitors and rationally extended the library of conformationally defined Pro-derived modules to succeed in developing a nanomolar inhibitor of lower size (K_d = 120 nM; MW= 734 Da). In contrast to the previous inhibitor, the optimized compounds reduced extra-vascularization of invasive breast cancer cells in a zebrafish model. In words of the authors, *this study represents an example of successful, structure-guided development of low molecular weight inhibitors specifically and selectively addressing a proline-rich sequence-recognizing domain that is characterized by a shallow epitope lacking defined binding pockets*. Thus, the evolved high-affinity inhibitor may now serve as a tool in validating the basic therapeutic concept, *i.e.*, the suppression of cancer metastasis by inhibiting a crucial protein-protein interaction involved in actin filament processing and cell migration (Barone et al., 2020). The general applicability of this strategy was also illustrated by the design of an ErbB4-derived ligand containing two ProM fragments, targeting the YAP1-WW domain with a fivefold higher affinity (Opitz, et al., 2015).

4.2. Design of Chimeric Proteins

Chimeric proteins have been traditionally designed to target and inhibit molecular interactions in the context of disease, with both diagnostic and therapeutic potential. Apart from antibodies, several alternative scaffolds have been exploited over the years using the fact that individual domain families are best suited for certain target families.

With the aim of improving the NMR NOE pattern for protein structure determination of low/medium affinity complexes of interacting molecules, covalent linkage can be a way to shift the equilibrium between the interacting partners to the bound state. This approach has been developed with the CD2BP2-GYF domain and the target peptide SHRPPPPGHRV from CD2. The peptide was covalently linked at the N-terminus of the domain through a 24-residues linker. In conjunction with general recognition rules for proline-rich sequence recognition the NOEs obtained allowed the accurate modelling of the protein/peptide complex (Freund et al., 2003).

Nature has also been exploited a similar approach among the SH3 domains of the Tec family of tyrosine kinases (Roberts et al., 2016). Since the binding site is far away from the N- and C-terminal tails of these domains, an alternative to shorten the connecting linker has been the design of a chimeric protein (SPCp41) by connecting a circular permutant of the α-spectrin SH3 (Spc-SH3) domain to the proline-rich decapeptide p41 (APSYSPPPPP) with a simple three-residues link. The aim was to obtain a single-chain protein with a tertiary fold that would imitate the binding between SH3 domains and proline-rich peptides. The high-resolution NMR structure of the chimera reproduces perfectly the interactions typically found in SH3/peptide complexes and is remarkably similar to that of the complex between the Spc-SH3 domain and the ligand. NMR relaxation data confirmed the tight binding between the ligand and the SH3 part of the chimera (Candel, Conejero-Lara, Martinez, van Nuland, & Bruix, 2007). Accordingly, DSC experiments indicated that the interactions at the binding interface develop high cooperativity with the rest of the structure since the protein unfolds by a two-state process. The chimera is more stable than the circular permutant by 6-8 $kJ \cdot mol^{-1}$ of Gibbs energy at 25 ºC; in addition, the difference in unfolding enthalpy is approximately 32 $kJ \cdot mol^{-1}$. These values coincide with those found for the binding of proline-rich peptides to SH3 domains (Martin-Sierra et al., 2003). In full agreement, the folding/unfolding kinetics can be correctly interpreted by a two-state process, where the folding transition state produces essentially the same picture shown by the circular permutant S19-P20s (the “nucleus” of the design), and the ligand will dock at the latter stages of folding (Candel, Cobos, Conejero-Lara, & Martinez, 2009). Thus, all conclusions corroborate the effectiveness of the chimera SPCp41 to study energetic, dynamic and structural aspects of SH3/ligand interactions. In fact, the contribution of each Pro residue of the ligand sequence (APSYSPPPPP) to the SH3/peptide interaction was evaluated by producing six single Pro-Ala mutants of the chimeric protein. Structural analyses of the mutant chimeras by circular dichroism, fluorescence and NMR, together with NMR-relaxation measurements, indicate conformational flexibility at the binding interface, which is strongly affected by the different Pro-Ala mutations. A three-state unfolding model was developed to evaluate the unfolding thermodynamics by DSC. The model assumes equilibrium between the “unbound” and “bound” states at the SH3/peptide binding interface and, therefore, allows distinguishing the thermodynamic magnitudes of the interaction at the binding interface from the ones of the unfolding. The resulting thermodynamic magnitudes classify the different Pro residues according to their relevance in the interaction as P2~P7~P10 > P9~P6 > P8, which agrees well with Lim's model for the interaction between SH3 domains and proline-rich peptides. These results demonstrate that this chimeric design may serve as a suitable tool to analyse the energetics of weak bio-molecular interactions such as those involving SH3 domains and

other polyproline-binding domains, using a combination of unfolding experiments and site-directed mutagenesis (Candel, van Nuland, Martin-Sierra, Martinez, & Conejero-Lara, 2008).

The ability to design proteins with desired properties by using protein structural information will allow creating high-value therapeutic and diagnostic products. For example, using the protein structures of lambda lysozyme and the human Crk-SH3 domain, a synthetic protein switch that controls lysozyme activity by sterically hindering its active cleft through the binding of SH3 to its CB1 peptide-binding partner was designed. The fusion protein, including lysozyme and CB1, was tested *in vitro*. In the absence of SH3, the lysozyme-CB1 fusion protein functioned normally. In the presence of SH3, the lysozyme activity was inhibited, and with the addition of excess CB1 peptides to compete for SH3 binding the lysozyme activity was restored. Lastly, this structure-based strategy can be used to engineer synthetic regulation by peptide/domain binding interfaces into a variety of proteins (Pham & Truong, 2012).

REFERENCES

Anang, S., Kaushik, N., Hingane, S., Kumari, A., Gupta, J., Asthana, S., . . . Surjit, M. (2018). 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, 92*(20).

Bacarizo, J., & Camara-Artigas, A. (2013). Atomic resolution structures of the c-Src SH3 domain in complex with two high-affinity peptides from classes I and II. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 5), 756-766.

Bacarizo, J., Martinez-Rodriguez, S., & Camara-Artigas, A. (2015). Structure of the c-Src-SH3 domain in complex with a proline-rich motif of NS5A protein from the hepatitis C virus. *J Struct Biol, 189*(1), 67-72.

Barone, M., Muller, M., Chiha, S., Ren, J., Albat, D., Soicke, A., . . . Kuhne, R. (2020). Designed nanomolar small-molecule inhibitors of Ena/VASP EVH1 interaction impair invasion and extravasation of breast cancer cells. *Proc Natl Acad Sci U S A, 117*(47), 29684-29690.

Bauer, F., & Sticht, H. (2007). A proline to glycine mutation in the Lck SH3-domain affects conformational sampling and increases ligand binding affinity. *FEBS Lett, 581*(8), 1555-1560.

Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P., & Wu, C. W. (1981). X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone. *Proc Natl Acad Sci U S A, 78*(7), 4175-4179.

Brannetti, B., Via, A., Cestra, G., Cesareni, G., & Helmer-Citterich, M. (2000). SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family. *J Mol Biol, 298*(2), 313-328.

Camara-Artigas, A., Ortiz-Salmeron, E., Andujar-Sanchez, M., Bacarizo, J., & Martin-Garcia, J. M. (2016). 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, 72*(Pt 9), 707-712.

Candel, A. M., Cobos, E. S., Conejero-Lara, F., & Martinez, J. C. (2009). Evaluation of folding co-operativity of a chimeric protein based on the molecular recognition between polyproline ligands and SH3 domains. *Protein Eng Des Sel, 22*(10), 597-606.

Candel, A. M., Conejero-Lara, F., Martinez, J. C., van Nuland, N. A., & Bruix, M. (2007). The high-resolution NMR structure of a single-chain chimeric protein mimicking a SH3-peptide complex. *FEBS Lett, 581*(4), 687-692.

Candel, A. M., van Nuland, N. A., Martin-Sierra, F. M., Martinez, J. C., & Conejero-Lara, F. (2008). Analysis of the thermodynamics of binding of an SH3 domain to proline-rich peptides using a chimeric fusion protein. *J Mol Biol, 377*(1), 117-135.

Cesareni, G., Gimona, M., Sudol, M., & Yaffe, M. (2005). *Modular Protein Domains*. Weinheim: Wiley-VCH.

Cestra, G., Castagnoli, L., Dente, L., Minenkova, O., Petrelli, A., Migone, N., . . . Cesareni, G. (1999). The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J Biol Chem, 274*(45), 32001-32007.

Cobos, E. S., Pisabarro, M. T., Vega, M. C., Lacroix, E., Serrano, L., Ruiz-Sanz, J., & Martinez, J. C. (2004). A miniprotein scaffold used to assemble the polyproline II binding epitope recognized by SH3 domains. *J Mol Biol, 342*(1), 355-365.

Corbi-Verge, C., & Kim, P. M. (2016). Motif mediated protein-protein interactions as drug targets. *Cell Commun Signal, 14*, 8.

Cheadle, C., Ivashchenko, Y., South, V., Searfoss, G. H., French, S., Howk, R., . . . Jaye, M. (1994). Identification of a Src SH3 domain binding motif by screening a random phage display library. *J Biol Chem, 269*(39), 24034-24039.

Chin, J. W., Grotzfeld, R. M., Fabian, M. A., & Schepartz, A. (2001). Methodology for optimizing functional miniature proteins based on avian pancreatic polypeptide using phage display. *Bioorg Med Chem Lett, 11*(12), 1501-1505.

Fazi, B., Cope, M. J., Douangamath, A., Ferracuti, S., Schirwitz, K., Zucconi, A., . . . Castagnoli, L. (2002). Unusual binding properties of the SH3 domain of the yeast actin-binding protein Abp1: structural and functional analysis. *J Biol Chem, 277*(7), 5290-5298.

Feng, S., Kapoor, T. M., Shirai, F., Combs, A. P., & Schreiber, S. L. (1996). Molecular basis for the binding of SH3 ligands with non-peptide elements identified by combinatorial synthesis. *Chem Biol, 3*(8), 661-670.

Feng, S., Kasahara, C., Rickles, R. J., & Schreiber, S. L. (1995). Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc Natl Acad Sci U S A, 92*(26), 12408-12415.

Fernandez-Ballester, G., Blanes-Mira, C., & Serrano, L. (2004). The tryptophan switch: changing ligand-binding specificity from type I to type II in SH3 domains. *J Mol Biol, 335*(2), 619-629.

Freund, C., Kuhne, R., Park, S., Thiemke, K., Reinherz, E. L., & Wagner, G. (2003). Structural investigations of a GYF domain covalently linked to a proline-rich peptide. *J Biomol NMR, 27*(2), 143-149.

Freund, C., Schmalz, H. G., Sticht, J., & Kuhne, R. (2008). Proline-rich sequence recognition domains (PRD): ligands, function and inhibition. *Handb Exp Pharmacol*(186), 407-429.

Han, Z., Lu, J., Liu, Y., Davis, B., Lee, M. S., Olson, M. A., . . . Harty, R. N. (2014). Small-molecule probes targeting the viral PPxY-host Nedd4 interface block egress of a broad range of RNA viruses. *J Virol, 88*(13), 7294-7306.

Han, Z., Ye, H., Liang, J., Shepley-McTaggart, A., Wrobel, J. E., Reitz, A. B., . . . Harty, R. N. (2021). Compound FC-10696 Inhibits Egress of Marburg Virus. *Antimicrob Agents Chemother, 65*(7), e0008621.

Harty, R. N. (2009). No exit: targeting the budding process to inhibit filovirus replication. *Antiviral Res, 81*(3), 189-197.

Hiipakka, M., & Saksela, K. (2007). Versatile retargeting of SH3 domain binding by modification of non-conserved loop residues. *FEBS Lett, 581*(9), 1735-1741.

Hodges, A. M., & Schepartz, A. (2007). Engineering a monomeric miniature protein. *J Am Chem Soc, 129*(36), 11024-11025.

Holtzman, J. H., Woronowicz, K., Golemi-Kotra, D., & Schepartz, A. (2007). Miniature protein ligands for EVH1 domains: interplay between affinity, specificity, and cell motility. *Biochemistry, 46*(47), 13541-13553.

Huang, H., Economopoulos, N. O., Liu, B. A., Uetrecht, A., Gu, J., Jarvik, N., . . . Sidhu, S. S. (2015). Selection of recombinant anti-SH3 domain antibodies by high-throughput phage display. *Protein Sci, 24*(11), 1890-1900.

Huang, K. Y., & Horng, J. C. (2015). Modulating the Affinities of Phosphopeptides for the Human Pin1 WW Domain Using 4-Substituted Proline Derivatives. *Biochemistry, 54*(40), 6186-6194.

Kasanov, J., Pirozzi, G., Uveges, A. J., & Kay, B. K. (2001). Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem Biol, 8*(3), 231-241.

Kofler, M., Motzny, K., Beyermann, M., & Freund, C. (2005). Novel interaction partners of the CD2BP2-GYF domain. *J Biol Chem, 280*(39), 33397-33402.

Kramer, A., & Schneider-Mergener, J. (1998). Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods Mol Biol, 87*, 25-39.

Kurakin, A., & Bredesen, D. (2002). Target-assisted iterative screening reveals novel interactors for PSD95, Nedd4, Src, Abl and Crk proteins. *J Biomol Struct Dyn, 19*(6), 1015-1029.

Lennard, K. R., Gardner, R. M., Doigneaux, C., Castillo, F., & Tavassoli, A. (2019). Development of a Cyclic Peptide Inhibitor of the p6/UEV Protein-Protein Interaction. *ACS Chem Biol, 14*(9), 1874-1878.

Link, N. M., Hunke, C., Mueller, J. W., Eichler, J., & Bayer, P. (2009). The solution structure of pGolemi, a high affinity Mena EVH1 binding miniature protein, suggests explanations for paralog-specific binding to Ena/VASP homology (EVH) 1 domains. *Biol Chem, 390*(5-6), 417-426.

Linn, H., Ermekova, K. S., Rentschler, S., Sparks, A. B., Kay, B. K., & Sudol, M. (1997). Using molecular repertoires to identify high-affinity peptide ligands of the WW domain of human and mouse YAP. *Biol Chem, 378*(6), 531-537.

Mahoney, N. M., Rozwarski, D. A., Fedorov, E., Fedorov, A. A., & Almo, S. C. (1999). Profilin binds proline-rich ligands in two distinct amide backbone orientations. *Nat Struct Biol, 6*(7), 666-671.

Martin-Garcia, J. M., Luque, I., Ruiz-Sanz, J., & Camara-Artigas, A. (2012). The promiscuous binding of the Fyn SH3 domain to a peptide from the NS5A protein. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 8), 1030-1040.

Martin-Garcia, J. M., Ruiz-Sanz, J., & Luque, I. (2012). Interfacial water molecules in SH3 interactions: a revised paradigm for polyproline recognition. *Biochem J, 442*(2), 443-451.

Martin-Sierra, F. M., Candel, A. M., Casares, S., Filimonov, V. V., Martinez, J. C., & Conejero-Lara, F. (2003). A binding event converted into a folding event. *FEBS Lett, 553*(3), 328-332.

Mongiovi, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., . . . Di Fiore, P. P. (1999). A novel peptide-SH3 interaction. *EMBO J, 18*(19), 5300-5309.

Morken, J. P., Kapoor, T. M., Feng, G. S., Shirai, F., & Schreiber, S. L. (1998). Exploring the Leucine-Proline Binding Pocket of the Src SH3 Domain Using Structure-Based, Split-Pool Synthesis and Affinity-Based Selection. *J Am Chem Soc, 120*(1), 30-36.

Musacchio, A., Gibson, T., Lehto, V. P., & Saraste, M. (1992). SH3--an abundant protein domain in search of a function. *FEBS Lett, 307*(1), 55-61.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., & Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature, 359*(6398), 851-855.

Nguyen, J. T., Porter, M., Amoui, M., Miller, W. T., Zuckermann, R. N., & Lim, W. A. (2000). Improving SH3 domain ligand selectivity using a non-natural scaffold. *Chem Biol, 7*(7), 463-473.

Nguyen, J. T., Turck, C. W., Cohen, F. E., Zuckermann, R. N., & Lim, W. A. (1998). Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors. *Science, 282*(5396), 2088-2092.

Opitz, R., Muller, M., Reuter, C., Barone, M., Soicke, A., Roske, Y., . . . Kuhne, R. (2015). A modular toolkit to inhibit proline-rich motif-mediated protein-protein interactions. *Proc Natl Acad Sci U S A, 112*(16), 5011-5016.

Otte, L., Wiedemann, U., Schlegel, B., Pires, J. R., Beyermann, M., Schmieder, P., . . . Oschkinat, H. (2003). WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci, 12*(3), 491-500.

Palencia, A., Camara-Artigas, A., Pisabarro, M. T., Martinez, J. C., & Luque, I. (2010). Role of interfacial water molecules in proline-rich ligand recognition by the Src homology 3 domain of Abl. *J Biol Chem, 285*(4), 2823-2833.

Palencia, A., Cobos, E. S., Mateo, P. L., Martinez, J. C., & Luque, I. (2004). Thermodynamic dissection of the binding energetics of proline-rich peptides to the Abl-SH3 domain: implications for rational ligand design. *J Mol Biol, 336*(2), 527-537.

Patel, S., Mathonet, P., Jaulent, A. M., & Ullman, C. G. (2013). Selection of a high-affinity WW domain against the extracellular region of VEGF receptor isoform-2 from a combinatorial library using CIS display. *Protein Eng Des Sel, 26*(4), 307-315.

Pham, E., & Truong, K. (2012). Engineered regulation of lysozyme by the SH3-CB1 binding interaction. *Protein Eng Des Sel, 25*(6), 307-311.

Piotukh, K., & Freund, C. (2012). A novel hSH3 domain scaffold engineered to bind folded domains in CD2BP2 and HIV capsid protein. *Protein Eng Des Sel, 25*(10), 649-656.

Pisabarro, M. T., & Serrano, L. (1996). Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain. *Biochemistry, 35*(33), 10634-10640.

Pisabarro, M. T., Serrano, L., & Wilmanns, M. (1998). Crystal structure of the abl-SH3 domain complexed with a designed high-affinity peptide ligand: implications for SH3-ligand interactions. *J Mol Biol, 281*(3), 513-521.

Przezdziak, J., Tremmel, S., Kretzschmar, I., Beyermann, M., Bienert, M., & Volkmer-Engert, R. (2006). Probing the ligand-binding specificity and analyzing the folding state of SPOT-synthesized FBP28 WW domain variants. *Chembiochem, 7*(5), 780-788.

Rickles, R. J., Botfield, M. C., Weng, Z., Taylor, J. A., Green, O. M., Brugge, J. S., & Zoller, M. J. (1994). Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO J, 13*(23), 5598-5604.

Rickles, R. J., Botfield, M. C., Zhou, X. M., Henry, P. A., Brugge, J. S., & Zoller, M. J. (1995). Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A, 92*(24), 10909-10913.

Roberts, J. M., Tarafdar, S., Joseph, R. E., Andreotti, A. H., Smithgall, T. E., Engen, J. R., & Wales, T. E. (2016). Dynamics of the Tec-family tyrosine kinase SH3 domains. *Protein Sci, 25*(4), 852-864.

Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B., & Ranganathan, R. (2005). Natural-like function in artificial WW domains. *Nature, 437*(7058), 579-583.

Santamaria, F., Wu, Z., Boulegue, C., Pal, G., & Lu, W. (2003). Reexamination of the recognition preference of the specificity pocket of the Abl SH3 domain. *J Mol Recognit, 16*(3), 131-138.

Siarot, L., Chutiwitoonchai, N., Sato, H., Chang, H., Fujino, M., Murakami, T., . . . Aida, Y. (2018). Identification of human immunodeficiency virus type-1 Gag-TSG101 interaction inhibitors by high-throughput screening. *Biochem Biophys Res Commun, 503*(4), 2970-2976.

Sidhu, S. S., Bader, G. D., & Boone, C. (2003). Functional genomics of intracellular peptide recognition domains with combinatorial biology methods. *Curr Opin Chem Biol, 7*(1), 97-102.

Socolich, M., Lockless, S. W., Russ, W. P., Lee, H., Gardner, K. H., & Ranganathan, R. (2005). Evolutionary information for specifying a protein fold. *Nature, 437*(7058), 512-518.

Sparks, A. B., Adey, N. B., Quilliam, L. A., Thorn, J. M., & Kay, B. K. (1995). Screening phage-displayed random peptide libraries for SH3 ligands. *Methods Enzymol, 255*, 498-509.

Sparks, A. B., Quilliam, L. A., Thorn, J. M., Der, C. J., & Kay, B. K. (1994). Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J Biol Chem, 269*(39), 23853-23856.

Stein, A., & Aloy, P. (2008). Contextual specificity in peptide-mediated protein interactions. *PLoS One, 3*(7), e2524.

Tang, W., Zhao, Z., Wang, C., Ye, T., & Yang, B. (2019). Molecular design and optimization of hepatic cancer SLP76-derived PLCgamma1 SH3-binding peptide with the systematic N-substitution of peptide PXXP motif. *J Mol Recognit, 32*(12), e2806.

Tavassoli, A., Lu, Q., Gam, J., Pan, H., Benkovic, S. J., & Cohen, S. N. (2008). Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. *ACS Chem Biol, 3*(12), 757-764.

Teyra, J., Huang, H., Jain, S., Guan, X., Dong, A., Liu, Y., . . . Sidhu, S. S. (2017). Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure, 25*(10), 1598-1610 e1593.

Teyra, J., Kelil, A., Jain, S., Helmy, M., Jajodia, R., Hooda, Y., . . . Sidhu, S. S. (2020). Large-scale survey and database of high affinity ligands for peptide recognition modules. *Mol Syst Biol, 16*(12), e9310.

Teyra, J., Sidhu, S. S., & Kim, P. M. (2012). Elucidation of the binding preferences of peptide recognition modules: SH3 and PDZ domains. *FEBS Lett, 586*(17), 2631-2637.

Tong, A. H., Drees, B., Nardelli, G., Bader, G. D., Brannetti, B., Castagnoli, L., . . . Cesareni, G. (2002). A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science, 295*(5553), 321-324.

Tonikian, R., Xin, X., Toret, C. P., Gfeller, D., Landgraf, C., Panni, S., . . . Boone, C. (2009). Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. *PLoS Biol, 7*(10), e1000218.

Tonikian, R., Zhang, Y., Boone, C., & Sidhu, S. S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat Protoc, 2*(6), 1368-1386.

Tran, T., Hoffmann, S., Wiesehan, K., Jonas, E., Luge, C., Aladag, A., & Willbold, D. (2005). Insights into human Lck SH3 domain binding specificity: different binding modes of artificial and native ligands. *Biochemistry, 44*(45), 15042-15052.

Tuchscherer, G., Grell, D., Tatsu, Y., Durieux, P., Fernandez-Carneado, J., Hengst, B., . . . Feller, S. (2001). Targeting Molecular Recognition: Exploring the Dual Role of Functional Pseudoprolines in the Design of SH3 Ligands. *Angew Chem Int Ed Engl, 40*(15), 2844-2848.

Verschueren, E., Spiess, M., Gkourtsa, A., Avula, T., Landgraf, C., Mancilla, V. T., . . . Distel, B. (2015). Evolution of the SH3 Domain Specificity Landscape in Yeasts. *PLoS One, 10*(6), e0129229.

Vidal, M., Liu, W. Q., Lenoir, C., Salzmann, J., Gresh, N., & Garbay, C. (2004). Design of peptoid analogue dimers and measure of their affinity for Grb2 SH3 domains. *Biochemistry, 43*(23), 7336-7344.

Zafra-Ruano, A., & Luque, I. (2012). Interfacial water molecules in SH3 interactions: Getting the full picture on polyproline recognition by protein-protein interaction domains. *FEBS Lett, 586*(17), 2619-2630.

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

FIGURE LEGENDS

Figure 1. Representative complexes of the six families of proline-rich recognition modules (SH3, WW, EVH1, GYF, UEV and Profilin). The domains are represented in ribbons, the ligand peptides in PPII conformation are shown in sticks, and relevant protein residues are represented as red sticks. The respective PDB codes are 4HVV (SH3), 2OEI (WW), 1EVH (EVH1), 1L2Z (GYF), 1M4Q (UEV) and 1CJF (Profilin). The consensus-binding motif of each family is shown. All the structures have been drawn using PyMol (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.).

Figure 2. c-Src-SH3 domain in complex with the high-affinity synthetic peptide APP12 (APPLPPRNRPRL) (PDB entry 4HVV). The xP pockets in the surface of the SH3 domain are

represented with the residues in sticks. The APP12 peptide is represented in yellow sticks. The specificity region comprises residues in the n-Src and RT-loops, where Arg7 residue interacts with Asp99 and Trp118. The structure has been drawn using PyMol (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.).

Figure 3. Water-mediated interaction patterns in SH3, WW and UEV domains. The domains are shown in light red ribbons, ligands are depicted as marine blue sticks, and relevant protein residues are represented as light red sticks. Interfacial waters are shown as non-bonded red spheres. Hydrogen bonds are depicted as discontinuous green lines. **(A)** Abl-SH3 in complex with the rationally designed peptide p41 (APSYSPPPPP) (1BBZ). **(B)** Dystrophin–WW domain in complex with a β-dystroglican peptide (1EG4). **(C)** Tsg101–UEV domain in complex with a proline-rich viral late domain sequence (3OBQ). Water molecules are those identified elsewhere (Martin-Garcia, Ruiz-Sanz, et al., 2012). All the structures have been drawn using PyMol (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.).

Figure 4. Phage display results of the panning against a WW domain of a fully randomized X_{12} library displayed in M13-pVIII protein. The resulting sequences are aligned in the left and the corresponding Web-logo representation (https://weblogo.berkeley.edu/logo.cgi) in the right. In the Web-logo, the total height indicates the degree of conservation of that position in the alignment; within each position, the height of the symbol indicates the frequency of the corresponding amino acid. Colour code corresponds to the nature of amino acids (black = apolar, green = polar, red = negative charge, blue = positive charge).

Figure 5. Crystal structure of the Avian Pancreatic Polypeptide (APP) showing the hairpin arranged between an α-helix (blue ribbon and sticks) and a PPII helix (yellow sticks). The PDB code is 1PPT (Blundell, et al., 1981). The structure has been drawn using PyMol (The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, LLC.).

Figure 1

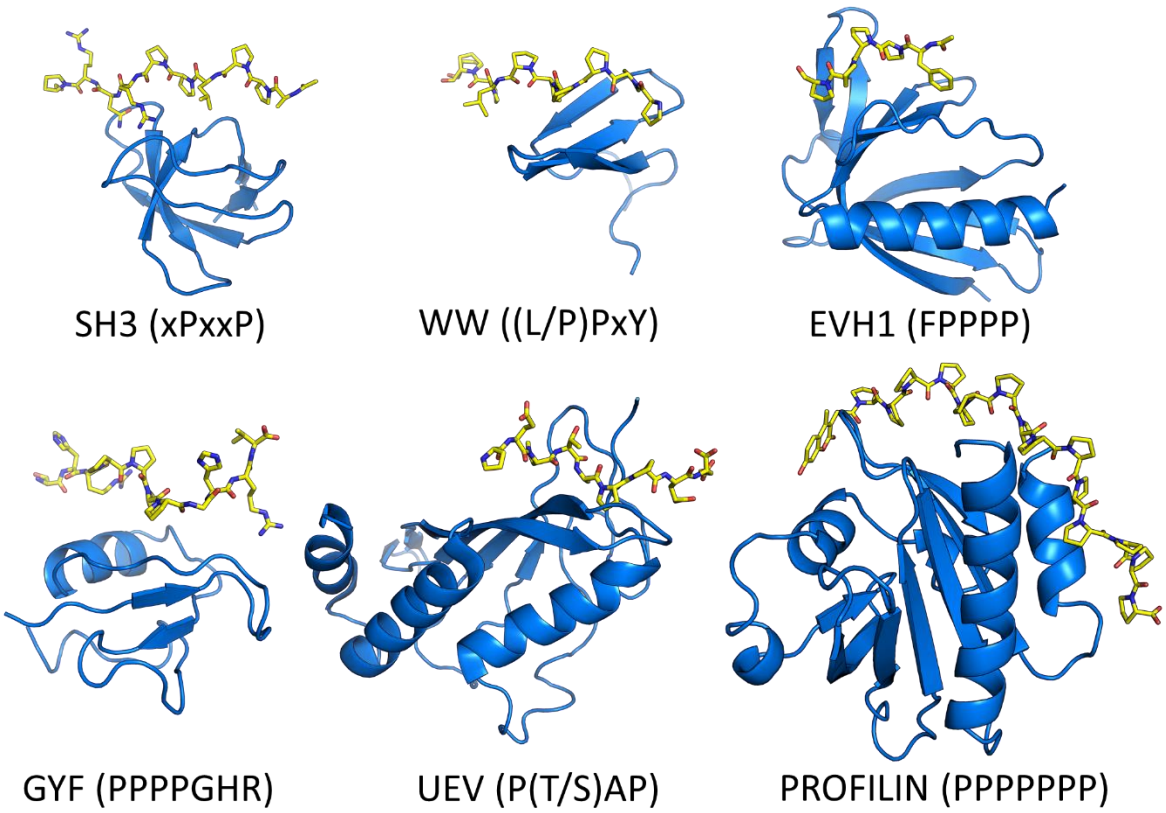
SH3 (xPxxP)
WW ((L/P)PxY)
EVH1 (FPPPP)
GYF (PPPPGHR)
UEV (P(T/S)AP)
PROFILIN (PPPPPPP)

Figure 2

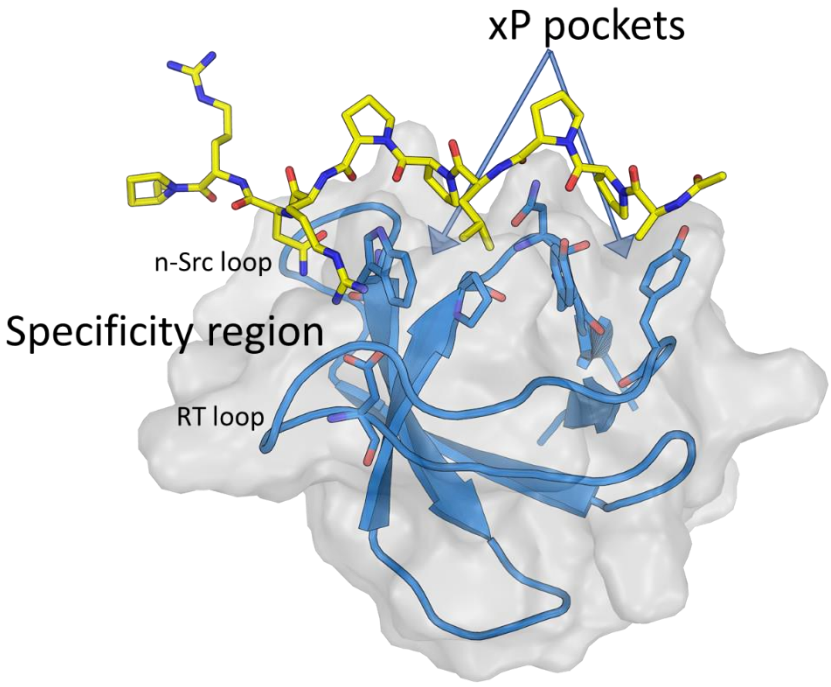

Figure 3

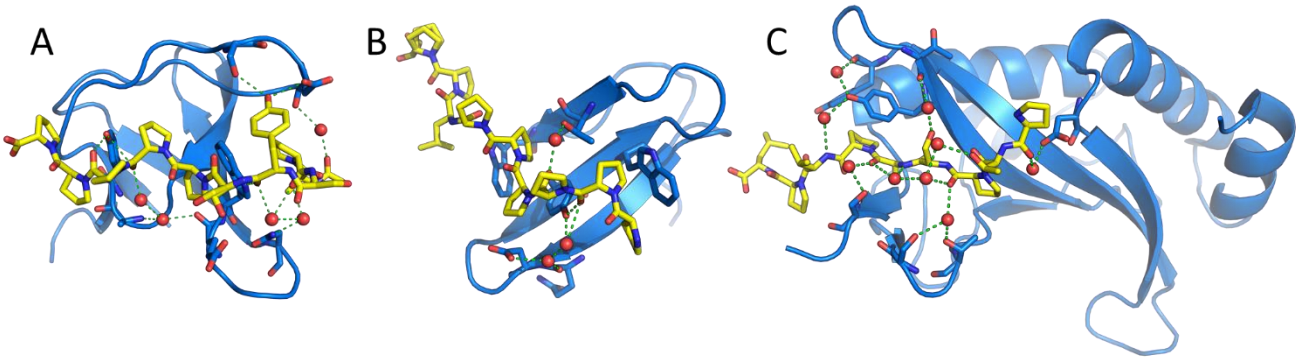

Figure 4

```
---VLITYDELS-SWC
-EDFLIEYAEVSS---
----PPDYDAVVKGLL(x2)
----PPDYEAIYKFV-
MSDTLPSYQTLC----
KFTTLPSYLSIF----
NIVPLPSYASLF----(x2)
WD--LPLYSEIILP--(x3)
---SLPEYHDLVPYW-
----LPLYCELYFDTC(x6)
WFPMLPSYWDLF----
----LPLYEDLYYYTF
```

Figure 5