Journal of Structural Biology 185 (2014) 147–162

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Designing protein function

Modular peptide binding: From a comparison of natural binders to designed armadillo repeat proteins

Christian Reichen, Simon Hansen, Andreas Plückthun *

Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland

ARTICLE INFO

Article history:
Available online 3 August 2013

Keywords:
Protein engineering
Libraries
Proteomics
Protein design
Repeat proteins
Peptide binding
Protein–peptide interactions

ABSTRACT

Several binding scaffolds that are not based on immunoglobulins have been designed as alternatives to traditional monoclonal antibodies. Many of them have been developed to bind to folded proteins, yet cellular networks for signaling and protein trafficking often depend on binding to unfolded regions of proteins. This type of binding can thus be well described as a peptide–protein interaction. In this review, we compare different peptide-binding scaffolds, highlighting that armadillo repeat proteins (ArmRP) offer an attractive modular system, as they bind a stretch of extended peptide in a repeat-wise manner. Instead of generating each new binding molecule by an independent selection, preselected repeats – each complementary to a piece of the target peptide – could be designed and assembled on demand into a new protein, which then binds the prescribed complete peptide. Stacked armadillo repeats (ArmR), each typically consisting of 42 amino acids arranged in three α-helices, build an elongated superhelical structure which enables binding of peptides in extended conformation. A consensus-based design approach, complemented with molecular dynamics simulations and rational engineering, resulted in well-expressed monomeric proteins with high stability. Peptide binders were selected and several structures were determined, forming the basis for the future development of modular peptide-binding scaffolds.

1. Introduction

In recent years the need for specific protein-binding molecules for experimental biology and medicine has been growing, especially for proteomic approaches. The ultimate goal is to map – ideally – every epitope, including posttranslational modifications, with specific protein detection reagents (Stoevesandt and Taussig, 2012). In sharp contrast to the increasing demand, the speed of generating such binders has not increased proportionally. Monoclonal antibodies obtained by traditional immunization are still the most frequently used binders, and their generation has essentially remained unchanged over the last 40 years. Within the last 1–2 decades, recombinant methods have provided greater control over the selection process (Plückthun et al., 2000) and revolutionized the generation of *therapeutic* proteins, notably therapeutic antibodies; in contrast, the impact on *proteomic* reagents has been modest so far. One reason is that the generation of specific binders, first in the form of antibody libraries (Knappik et al., 2000), later with alternative scaffold libraries (Binz et al., 2005), depends on the selection from a library, which has to be carried out for every target individually, and the performance and specificity of each selected binder must be evaluated individually and in detail – just as with traditional antibodies from immunization. Thus, no direct advantage is derived from having carried out previous selections: every target is a new challenge.

A modular binder, where each unit contributes to the interaction with the target molecule in a predefined manner, could overcome this limitation. Modular binding is difficult to achieve for folded protein targets. Therefore, such a modular approach is currently only conceivable for linear targets like RNA, DNA or peptides in extended conformation, and perhaps for oligosaccharides. Nucleic acids are the prime example of the use of this principle in biology: it is the prerequisite for forming a double strand. Our focus here, however, is on protein-based binding of peptides, which is a much greater challenge.

Nature has developed proteins with a modular binding mode for nucleic acids, and some of these have been technologically exploited, such as the transcription activator-like effector (TALE) repeats (Deng et al., 2012; Mak et al., 2012) or zinc-finger proteins (Klug, 2010). In both cases, repeats or domains can be linked in tandem to recognize sequence motifs in nucleic acids of various lengths. Whereas one zinc-finger domain binds three nucleotides of the DNA or RNA target, one TALE repeat recognizes one nucleotide on one strand of a dsDNA. Using this modular binding principle, artificial TALE repeat proteins were engineered to bind any consecutive nucleotide sequence of choice (Boch et al., 2009).

* Corresponding author. Address: Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Fax: +41 44 635 57 12.
E-mail address: plueckthun@bioc.uzh.ch (A. Plückthun).

1047-8477/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2013.07.012

The key question is now, how to achieve a conceptually similar modular system for peptides in extended conformation.

In the first part of this review the peptide-binding scaffolds that have been structurally described are compared and analyzed for their ability to potentially bind peptide targets in a modular fashion. To anticipate our conclusions below: we found that repeat proteins are of particular interest, since their tandem structure allows rigid stacking of repeat units and the formation of an extended structure. Among them, armadillo repeat proteins (ArmRP) form a continuous peptide-binding surface and each module binds, in a first approximation, two consecutive amino acids. An asparagine residue, conserved in almost every repeat, keeps the peptide in extended conformation through binding to the peptide backbone. Other amino acids of the binding surface provide the specificity for the target peptide by interacting with the peptide side chains. Based on these considerations, Armadillo repeat proteins (ArmRPs) were chosen as template to generate a modular peptide-binding scaffold.

In the second part of this review we describe the different protein engineering steps to develop designed ArmRPs with regular tandem repeats and favorable biophysical properties. Although modular binding along the extended binding surface of designed ArmRPs has not yet been shown, designed ArmRP have demonstrated their applicability in pull–down experiments or as detection agents in Western blots (Varadamsetty et al., 2012) and form the basis for the future development towards a modular peptide-binding system.

2. Peptide-binding: strategies and scaffolds

Protein–protein interactions are essential for living cells, and in many of these interactions both proteins are globular domains. Many of these have not only been well characterized, but also numerous protein–protein interaction scaffolds were explored for the generation of designed binding molecules as promising alternatives to traditional monoclonal or recombinant antibodies. Most of these scaffolds, such as, e.g., (in alphabetical order) Adnectins, Affibodies, Anticalins or Designed Ankyrin Repeat Proteins (DARPins) bind usually to the surfaces of folded proteins and thus do not form the focus of our review, as they have been reviewed elsewhere (Binz et al., 2005; Boersma and Plückthun, 2011; Caravella and Lugovskoy, 2010; Hosse et al., 2006; Löfblom et al., 2011; Mintz and Crea, 2013).

In the past decade, it became clear, however, that about 15–40% of all interactions in the cell are peptide–protein interactions (Petsalaki et al., 2009). By this term, we do not mean short oligopeptides, but rather unstructured regions of proteins which can bind as linear peptide targets. They can be described as unfolded terminal regions of protein domains, unstructured loops within a domain, disordered linkers between two domains (London et al., 2010) or parts of intrinsically unstructured proteins. Such transient and in general low-affinity but highly specific interactions between a globular protein and short linear peptide regions have been found in many highly dynamic cellular networks involved in signaling, regulation and protein trafficking (Diella et al., 2008; Pawson and Nash, 2003).

The challenge of binding to peptides – in contrast to stable folded proteins domains – is that peptides are usually flexible (as we are excluding structured peptides here). They thus lose a large amount of configurational entropy upon association (Killian et al., 2009). The analysis of peptide–protein binding strategies by London and coworkers (London et al., 2010) revealed that the loss of configurational entropy upon binding is minimized by the rigidity of the protein interface and compensated by an over-representation of hydrogen bonds between peptide and protein. These additional hydrogen bonds are accomplished mainly by interaction of the peptide backbone with the binding surface.

Many peptide–protein interaction domains have been characterized (Pawson and Scott, 1997; Pawson and Nash, 2003). For better comparison, we describe below all well-studied peptide-binding scaffolds according to (i) fold, (ii) target, (iii) binding mode, (iv) typical affinity and (v) natural function and illustrate them in Fig. 1. This comprehensive survey will then allow us to justify potential choices of scaffolds for the engineering of modular binding.

2.1. Antibodies

The antigen-binding variable domains (V_H and V_L) of antibodies are composed of a conserved two-β-sheet framework and six hypervariable loops, known as complementarity determining regions (CDRs) (Fig. 1A) (Sundberg, 2009). Variable in length and sequence, CDRs determine the shape of the binding site (Collis et al., 2003; MacCallum et al., 1996). Antibodies can bind to folded proteins, peptides, DNA, carbohydrates and other substances. Anti-peptide antibodies have a binding site which is usually an intermediate between the generally very deep binding pocket of anti-hapten antibodies and the relatively large and flat binding surface of protein binders (Almagro, 2004). Peptides are often bound in a groove along the dimer interface formed by the V_H and V_L domains, and sometimes one amino acid of the peptide binds in a central cavity, like a hapten. Nevertheless, the binding mode of anti-peptide antibodies is not conserved, and different peptides assume many different orientations and conformations, such as extended chains, β-turns or α-helices (Stanfield and Wilson, 1995; Sundberg, 2009). Anti-peptide antibodies can bind their target with very high affinities with reported dissociation constants in the nM range or even below (Ferrières et al., 2000; Pope et al., 2009; Luginbühl et al., 2006; Zahnd et al., 2004) and are a central component of the adaptive immune system of higher vertebrates.

2.2. MHC-I and MHC-II

The membrane-anchored and heterodimeric major histocompatibility complexes (MHC-I and MHC-II) are both composed of three domains, one peptide-binding domain and two immunoglobulin-like domains. The peptide-binding domain is composed of an eight-stranded β-sheet platform laterally enclosed by two α-helices (Yaneva et al., 2010). MHC-I proteins bind short extended peptides (8–9 amino acids in length) that originate from the intracellular degradation of (endogenous) proteins. MHC-II proteins target longer peptides (up to 20 residues) (Fig. 1A) that originate from proteolysis of engulfed extracellular (exogenous) proteins. The peptide groove of MHC-I complexes is encoded by a single protein and is closed, explaining the strict length limits of the peptide antigen (Fig. 1A). In class II MHC proteins, the binding groove is formed by two protein chains and is open at both ends, allowing MHC-II to bind longer peptides. Although MHC-II are able to bind long and highly variable peptides, the low stability and yield of MHC-II complexes makes working with them a demanding task and puts some limit on their utility as a biotechnological tool. The affinities for both complexes span a wide range. Dissociation constants between low nM and high μM have been measured (Christinck et al., 1991; Fahnestock et al., 1994; Froloff et al., 1997; Morgan et al., 1997; Sadegh-Nasseri et al., 1994). Both complexes are involved in the mammalian cellular immune response: they are presenting bound peptides at the cell surface to engage the T-cell receptor and thus activate T-cells (Neefjes et al., 2011; Rudolph et al., 2006; Vyas et al., 2008).

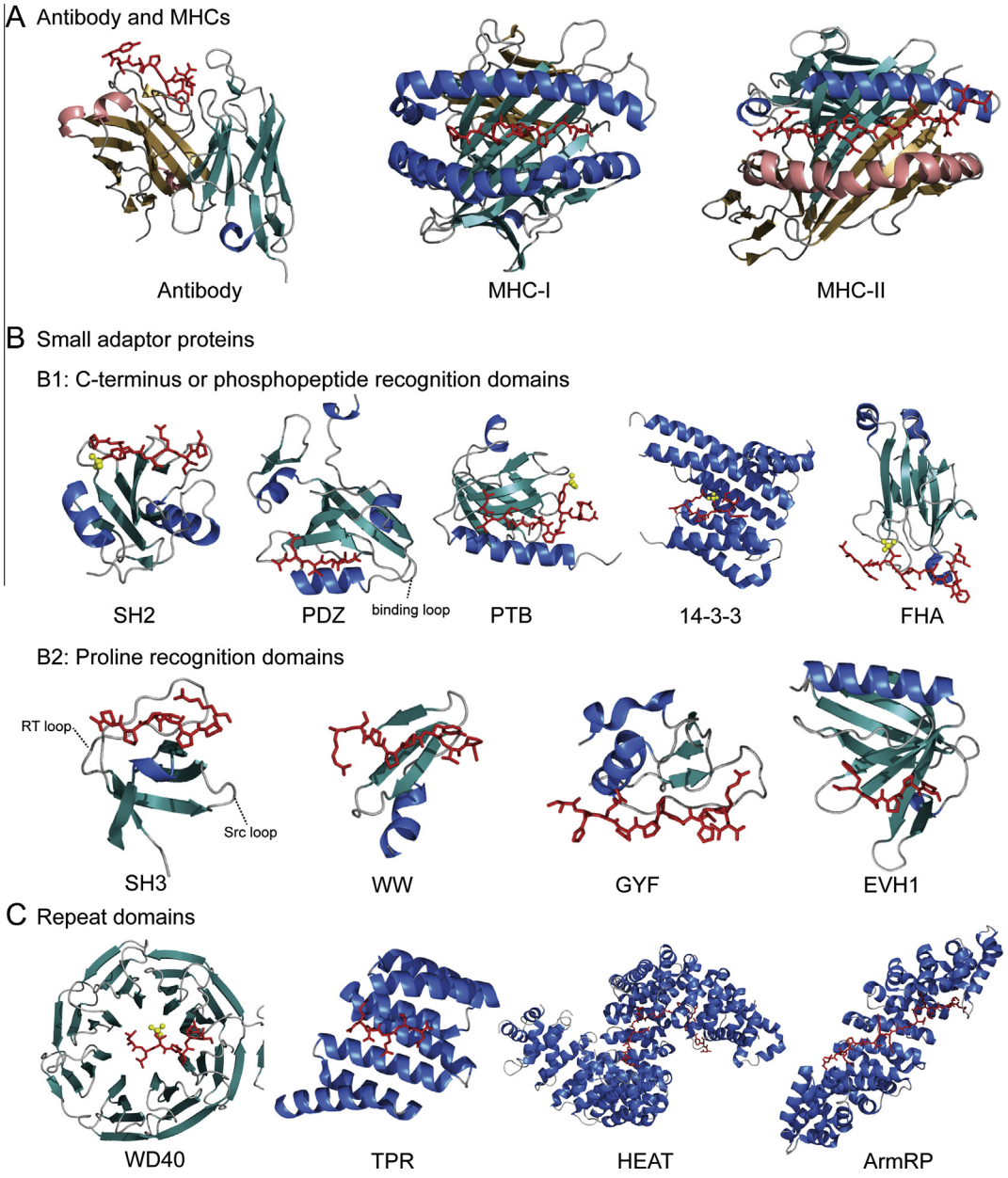

Fig.1. Collection of peptide-binding proteins. Bound peptides are shown in red stick representation, with N- and C-terminus at the right and left side, respectively. For each binder, α-helical secondary structures are shown in blue, β-sheets in cyan and loops in grey. For heterodimeric proteins, chain B is colored in salmon, brown and grey, respectively. Phosphoryl groups are indicated in yellow ball-and-stick representation. (A) Structure of variable heavy and light chain domains (V_H and V_L) of an antibody bound to hemagglutinin peptide (PDB ID 1HIM (Rini et al., 1992)), structure of MHC-I in complex with viral peptide SEV9 (PDB ID 1FZK (Rudolph et al., 2001)) and MHC-II protein binding autoantigen glutamic acid decarboxylase peptide (GAD) (PDB ID 1ES0 (Corper et al., 2000)). (B) Examples of small adaptor proteins, binding to C-terminal- or phospho-peptides (B1) or to proline-rich peptide sequences (B2). (B1) Structure of SH2 domain of Src binding to phosphotyrosine peptide pYEEI (PDB ID 1SPS (Waksman et al., 1993)). PDZ domain from synaptic protein PSD-95 in complex with its C-terminal peptide (PDB ID 1BE9 (Doyle et al., 1996)). NMR structure of the PTB domain of IRS-1 in complex with its phosphotyrosine peptide derived from the IL-4 receptor (PDB ID 1IRS (Zhou et al., 1996)). Structure of the ζ isoform of 14–3-3 domain with a phosphoserine peptide bound (PDB ID 1QJB (Rittinger et al., 1999)). FHA domain of Rad53p in complex with a phospho-threonine peptide (PDB ID 1G6G (Durocher et al., 2000)). (B2) Selection of proline recognition domains. SH3 domain of Sem5 complexed to mSos-derived peptide (PDB ID 1SEM (Lim et al., 1994)). WW domain of dystrophin binding to a proline-rich motif in the tail of β-dystroglycan (PDB ID 1EG4 (Huang et al., 2000)). NMR structure of GYF domain of CD2 antigen binding protein 2 (CD2BP2) in complex with the T-cell surface antigen CD2 peptide (PDB ID 1L2Z (Freund et al., 2002)). Structure of Enabled (Mena) EVH1 domain complexed with a consensus peptide ligand (PDB ID 1EVH (Prehoda et al., 1999)). (C) Peptide-binding repeat domains. WD40 domain of Cdc4 complexed with phosphothreonine-CPD peptide from cyclin E (PDB ID 1NEX (Orlicky et al., 2003)). TPR domain of adaptor protein Hop (TPR2A) in complex with the C-terminal peptide of Hsp90 (PDB ID 1ELR (Scheufler et al., 2000)). HEAT domain of karyopherin β2 in complex with the NLS of hnRNP A1 (PDB ID 2H4M (Lee et al., 2006)). ArmRP of yeast importin-α (karyopherin α) binding to bipartite nuclear localization signal of *Xenopus* nucleoplasmin (PDB ID 1EE5 (Conti and Kuriyan, 2000)).

2.3. Small adaptor proteins

Cellular processes, e.g., the dynamic process of signal transduction, depend on precise temporal and spatial assembly of macromolecular complexes, mediated by small peptide recognition modules (PRMs) (Pawson and Scott, 1997; Pawson and Nash, 2003). These interaction domains are 50–200 amino acids in size, usually fold independently and bind specific core peptide motifs (Cesareni et al., 2005; Teyra et al., 2012). Many of such PRM have been identified (Kuriyan and Cowburn, 1997) and SH2, PDZ, PTB,

14-3-3, FHA, SH3, WW, GYF and EVH1 are described in more detail below (Fig. 1B). As these small adaptor proteins are involved in building up signaling networks, binding events must be tightly regulated and have to be reversible. Therefore, network-mediating domains have, on average, only low to medium affinities to their peptide target (μM to high nM range).

2.3.1. SH2 domain

The SH2 (Src homology 2) domain is 100 amino acids in length and forms a seven-stranded β-sheet (a four-stranded sheet directly connected to a three-stranded β-sheet), flanked by two α-helices (Fig. 1B). SH2 domains are specialized in recognizing short motifs containing a phosphorylated tyrosine (pTyr), which is C-terminally followed by three to five amino acids (Songyang et al., 1997). The pTyr is bound by a conserved positively charged pocket. Selectivity for peptides is achieved by different configurations of surface loops, regulating the accessibility for the C-terminal amino acids of the peptide to binding pockets (Kaneko et al., 2011; Schlessinger and Lemmon, 2003; Yaffe, 2002). SH2 domains bind their peptide targets with affinities in the range of 0.1–1 μM (Kuriyan and Cowburn, 1997) and are involved in downstream signaling of receptor tyrosine kinases.

2.3.2. PDZ domain

The PDZ domain has been found (and named accordingly) in three proteins: postsynaptic density 95 (PSD-95), discs large (Dlg) and zonula occludens-1 (ZO1). The PDZ domain (about 80–90 amino acids in length) folds into five to six β-strands and one or two α-helices (Fig. 1B) (Nourry et al., 2003; Reimand et al., 2012). PDZ domains bind mostly to unstructured regions at the carboxy-termini of proteins; however, some are capable of binding to internal protein sequences or phospholipids (Kaneko et al., 2011). The typical peptide bound by PDZ domains is a four-amino-acid motif with a consensus sequence (x[T/S]xϕ-COOH) (ϕ being any hydrophobic residue (Aasland et al., 2002)), however several other peptide classes have been described with up to seven residues in length (Nourry et al., 2003; Songyang et al., 1997; Stiffler et al., 2007; Tonikian et al., 2008). The C-terminus of the bound peptide is recognized by the carboxylate-binding loop. The remainder of the peptide binds as additional β-strand in a groove between a β-strand and an α-helix of the domain (Kaneko et al., 2011). Reported dissociation constants are in the double digit nM to double digit μM range (Jaulin-Bastard et al., 2002; Songyang et al., 1997; Stiffler et al., 2007). PDZ domains are often involved in protein–protein interactions as multi-domain adaptors and regulate e.g., the localization of components of signaling pathways to the membrane (Reimand et al., 2012).

2.3.3. PTB domain

The phosphotyrosine-binding domain (PTB) was first found in the scaffold protein Shc (Blaikie et al., 1994). PTB domains are approximately 200 amino acids in length and fold into a β-sandwich, formed by seven antiparallel β-strands and a C-terminal α-helix. This so-called pleckstrin homology (PH) superfold is also found in the EVH1-domain (see below) (Schlessinger and Lemmon, 2003; Uhlik et al., 2005; Yaffe, 2002). The target peptide usually contains an NPxY-motif and forms a conserved β-turn in complex with the PTB domain (Uhlik et al., 2005). However, many variations of the NPxY motif suggest a wider binding specificity range of PTB domains. For example, the tyrosine residue is phosphorylated in the binding targets of Shc, IRS, Dok and SNT, but is unphosphorylated or even replaced by a phenylalanine in substrates for the Dab-like PTB domains (Uhlik et al., 2005). PTB domains bind their peptide ligand as an anti-parallel pseudo-β-sheet (antiparallel β-sheet augmentation, already seen in PDZ domains (Harrison, 1996)), forming extensive contacts with one β-strand and the C-terminal α-helix (Fig. 1B). Reported affinities of PTB domains are around 1–200 μM (Wolf et al., 1995), and are found in cell signaling processes (Schlessinger and Lemmon, 2003).

2.3.4. 14-3-3 Domain

The highly conserved acidic 14-3-3 proteins obtained their name based on their particular migration pattern on two-dimensional DEAE-cellulose chromatography and starch gel electrophoresis. 14-3-3 domains contain approximately 230 amino acids (based on all 34 structures in the Protein Data Bank), forming a bundle of nine antiparallel α-helices with a concave surface including an amphipathic ligand-binding-groove (Fig. 1B) (Rittinger et al., 1999). 14-3-3 domains typically form rigid dimers (homo- or heterodimers with different 14-3-3 isoforms), with a cup-like shape. Initially, 14-3-3 domains have been found to bind to phosphorylated serines in unstructured N- and C-terminal regions of the target (mode I: R[S/ϕ] + pSxP or mode II: Rx[ϕ/S] + pSxP) (reviewed in Obsil and Obsilova (2011)). In addition, significantly different binding motifs have been identified, including phosphorylated threonines, unphosphorylated sequences (e.g., amphipathic WLDLE sequence) or the C-terminal $pS/pTx_{1-2}COOH$ motif (mode III). Binding to phosphoserine-containing peptides is mediated by a basic pocket formed by residues placed on two helices of the domain. Residues placed on two other helices of the domain interact further with the peptide regions adjacent to the phosphorylation site (Rittinger et al., 1999). Dissociation constants lie in the two- to three digit nM range (Muslin et al., 1996; Rittinger et al., 1999; Yaffe et al., 1997). 14-3-3 domains are involved in intracellular signal transduction and cell cycle regulation (Rittinger et al., 1999).

2.3.5. FHA domain

FHA (forkhead-associated) domains with a length of about 80–100 amino acids consist of a 10–11 stranded β-sandwich. Sequence insertions in the loops connecting the β-strands form the main part of the binding site and are often composed of helical secondary structure (Fig. 1B) (Yaffe and Smerdon, 2004). In contrast to its very strict specificity towards pThr, the FHA domain recognizes very diverse patterns in the residues surrounding the pThr residue (Mahajan et al., 2008). The bound peptide has the consensus motif: [R/D/H][ϕ/Ω][A/Q/I/V][K/A/P]pT[V/I/E/L/Q/M][ϕ/C][D/ϕ/π] (Aasland et al., 2002; Yaffe and Smerdon, 2004). Unlike WW and 14-3-3 domains, FHA domains bind with high affinity exclusively to pThr but not to pSer peptides. The binding site carries four of the seven most highly conserved residues in the FHA domain, of which three bind directly to the pThr residue of the bound peptide. The lowest dissociation constants of FHA domains are around 100 nM (Byeon et al., 2005; Pennell et al., 2010). FHA domains are involved in transcription, DNA damage repair and cell cycle control (Durocher and Jackson, 2002; Yaffe and Smerdon, 2004).

2.3.6. Proline recognition domains (PRD)

SH3, WW, GYF, EVH1 and UEV domains, as well as profilin belong to the proline recognition domain (PRD) family that recognizes proline-rich sequences (Kofler and Freund, 2006). Repetitive proline-rich sequences are found in many proteins and often function as docking sites for signaling modules (Fig. 1B2) (Zarrinpar et al., 2003). Proline is the only natural amino acid with a cyclic side chain, creating substituted amide nitrogen and restrictions for its dihedral angles. Consequently, proline-rich-ligands have a propensity to form extended left-handed polyproline type II (PPII) helices. The PPII-helix is triangular in cross section (prism) and generally contains proline residues in at least every third amino acid position (Ball et al., 2002). Although PRD targets vary in their proline-rich sequences, every target peptide forms a PPII helix upon binding. SH3, WW and GYF have a relatively flat binding surface and are binding the PPII helix prism on one flat surface. In

contrast, the EVH1 domain with its wedge-shaped hydrophobic groove binds the PPII helix prism by its apex (Zarrinpar et al., 2003).

2.3.6.1. SH3 domain. SH3 domains (Src homology 3) of about 60 amino acids form a five-stranded β-sheet (Fig. 1B). Typically, peptide sequences carrying the PxxP motif flanked by a charged residue are targeted (Kaneko et al., 2011; Teyra et al., 2012). The ligand is bound on the surface flanked by two loops, the RT- and the Src- loop. Recognition of the proline-rich-peptide ligand in its PPII-helix conformation is achieved by insertion of the ridges of the PPII-helix into a complementary pair of grooves on the SH3 surface. These grooves are defined by a series of nearly parallel, well conserved aromatic residues (Zarrinpar et al., 2003). The two loops appear to play a key role for the shape and specificity of the binding surface. The observed dissociation constants are usually between 1 and 10 μM (Kuriyan and Cowburn, 1997; Mayer, 2001) but also a few sub-μM affinities have been reported (Lee et al., 1995; Posern et al., 1998). SH3 domains are abundant in eukaryotes (536 SH3 domains have been identified in the human proteome) and are involved in several processes including cytoskeleton regulation, receptor tyrosine kinase pathways and endocytosis (Teyra et al., 2012).

2.3.6.2. WW domain. WW domains are named after two conserved tryptophan (W) residues involved in peptide binding. They are highly compact binding domains, consisting of only about 40 amino acids, forming an antiparallel three stranded β-sheet (Fig. 1B) (Macias et al., 2002; Zarrinpar et al., 2003). Recognition motifs of WW domains can be divided into four peptide classes: (i) PPxY, (ii) PPLP, (iii) P–R repeats and (iv) (pS/pT)P. Structures of WW domain–ligand complexes are similar to SH3 domains, with peptides adopting a PPII-helix conformation and binding surfaces containing nearly parallel aromatic residues. WW domains differ from SH3 domains in that they typically have only one proline binding groove, as compared to the two adjacent proline binding grooves found in SH3 domains (Zarrinpar et al., 2003). Like in SH3-domains, specificity is regulated by variable loops adjacent to the groove. Dissociation constants for the interaction of WW domains with their peptide targets are high nM to low μM for interactions with proline-rich ligands and low μM for pS/pT containing peptides (Kay et al., 2000). WW domains are found in many proteins that play a role in cell signaling or regulation (Lu et al., 1999).

2.3.6.3. GYF domain. GYF domains are named after a conserved Glycine–Tyrosine–Phenylalanine sequence motif. GYF domains with about 60 amino acids have a conserved sequence ($WxYx_{6-11}GPFx_{4-}Mx_2Wx_3GYF$) and fold into a main α-helix packed against a small anti-parallel β-sheet (Fig. 1B). They interact with proline-rich-peptides with a conserved PPG motif. The main helix is tilted away from the sheet, providing space for an array of stacked aromatic side chains which create the binding site for the PPG motif of the peptide ligand (reviewed in Kofler and Freund (2006)). Monovalent interactions between GYF domains and peptide targets have reported dissociation constants of about 200 μM (Freund et al., 2002). The GYF domain is thought to be involved in splicing and splicing-associated processes, immune cell function and antigen presentation (Kofler and Freund, 2006).

2.3.6.4. EVH1 domain. EVH1 (enabled/vasodilator-stimulated phospho-protein homology 1) domains are about 115 residues in length and adopt the pleckstrin homology (PH) fold, as found for PTB-domains (Peterson and Volkman, 2009). Different families with different consensus core target peptide sequences could be identified within the EVH1 domain targets (PPxϕP for Ena/VASP; PPxxF for Homer; extended ligands up to 25 amino acids for WASP). The mechanism of binding is always based on binding the peptide in a PPII helix. In contrast to PTB domains, which bind their targets between an α-helix and a β-strand by antiparallel-β-sheet augmentation, EVH1 domains bind their peptide targets by a groove formed only by β-strands (Fig. 1B). In PTB domains this groove is hidden by an extra α-helix, which is missing in EVH1 domains. EVH1 domains have concave binding sites, which only allow binding of the PPII helix via its apex site (Zarrinpar et al., 2003). EVH1 domains typically exhibit low affinities towards their target peptides, reported K_d values are between 1 and 500 μM and are involved in modulation of the actin cytoskeleton and signal transduction (Ball et al., 2002).

2.4. Repeat proteins

Repeat protein families can be divided into closed and open structures: while closed structures have a defined number of repeats, as found, e.g., in β-propeller proteins like WD40 repeat proteins, open structures can have highly variable numbers of repeats, as found in the α-solenoid structures of tetratricopeptide repeats (TPRs), HEAT repeat proteins and ArmRPs (Fig. 1C) (Kippert and Gerloff, 2009). All of these repeat proteins have been shown to possess an intrinsic ability to bind peptides, taking advantage of the binding surface generated on their repetitive structure. Repeat proteins consist of tandem arrays of small structural motifs of 20–50 amino acids in length. In the case of open repeat proteins, these motifs stack together to form elongated stable structures. Extended structures have a larger surface-area-to-volume ratio than typical globular proteins and therefore a larger potential binding area (Grove et al., 2008; Main et al., 2005). These features make them attractive for using their stable scaffolds as a basis for engineering binding molecules (Andrade et al., 2001a, 2000; Boersma and Plückthun, 2011). In this review only repeat proteins binding primarily to peptide targets will be discussed. Designed Ankyrin Repeat Proteins (DARPins), even though they can be selected to peptide targets, will be excluded from this analysis, because their peptide-binding mode is not conserved (Batyuk et al., unpublished data). Natural Ankyrin repeat proteins have also been reported to bind peptides, but also in different binding modes (Collins et al., 2008; Xu et al., 2012).

2.4.1. Beta-propeller proteins

Beta-propellers are the most common closed repeat structures; they are formed by 4–10 repeats of four-stranded β-sheets (the so called "blades") (Chen et al., 2011). Beta-propellers are built up from WD40, kelch (Adams et al., 2000) or other types of repeats, which differ in sequence but share a common fold (Chen et al., 2011). A WD40 repeat comprises a 44–60 amino acid sequence that typically contains the GH dipeptide 11–24 residues from its N-terminus and the WD dipeptide at the C-terminus of each repeat (Smith et al., 1999). A kelch repeat is 44–56 amino acids long and contains a GG motif and conserved Y and W residues located C-terminally of the GG dipeptide. Target peptides often carry post-translational modifications. They are usually positioned on the top surface, close to the entry site of the central channel (top view shown in Fig. 1C). However, binding to the bottom region or the circumference has also been observed and illustrates the broad spectrum of possible binding sites and variety of ligands (Grove et al., 2008; Stirnimann et al., 2010). Affinities are usually in the high nM or low μM range (Couture et al., 2006; Jennings et al., 2006; Nash et al., 2001; Oliver et al., 2009) but also K_d values below hundred nM have been reported (Song et al., 2008). Beta-propeller proteins are involved in different processes like signal transduction, RNA processing, vesicular trafficking, cytoskeleton assembly, cell cycle regulation and others (Smith et al., 1999; Stirnimann et al., 2010).

2.4.2. TPR proteins

The tetratricopeptide repeat (TPR) motif consists of 34 residues on average (hence the name) and adopts a helix-turn-helix fold (reviewed in D'Andrea and Regan (2003), Zeytuni and Zarivach (2012)). In nature, 3–16 adjacent TPR units form an overall superhelical structure, topologically identical to the structure of 14-3-3 proteins, with an inner concave binding surface (Fig. 1C). Ligands of TPR proteins are highly variable in sequence and secondary structure. Extended coil or α-helical conformations or both have been found as the secondary structure elements recognized within the peptidic ligands. The binding mode of TPRs is variable: often peptides are accommodated in a groove formed by the helices, but also interactions where the peptide is recognized by loop regions have been found (D'Andrea and Regan, 2003; Scheufler et al., 2000). Low μM dissociation constants were measured between TPR proteins and their targets (Scheufler et al., 2000). TPR proteins are involved in a wide range of processes, e.g., transcription, protein translocation, cell cycle regulation or chaperone activity (D'Andrea and Regan, 2003).

2.4.3. HEAT proteins

The HEAT motif was discovered in a diverse group of proteins; it was named after four members: huntingtin, elongation factor 3, the PR65/A subunit of protein phosphatase 2A (PP2A), and the lipid kinase TOR (target of rapamycin) (Andrade and Bork, 1995). The canonical HEAT repeat, composed of 37–47 amino acids, is built from two helices, A and B, which form a helical hairpin (Fig. 1C). The B helices form a concave ligand-binding surface, which in importin β1 and β2 is highly conserved. HEAT repeat proteins have highly divergent repeat sequences, which restrict consensus sequence approaches to subfamilies. The HEAT protein importin-β binds the N-terminal portion of importin-α, which adopts a helical secondary structure upon binding, while other parts are bound in an extended conformation (Cingolani et al., 1999). Karyopherin-β2, another well characterized HEAT protein, binds to the folded protein Ran-GTP (Chook and Blobel, 1999). Affinities of HEAT proteins have mainly been determined for importin β and are usually in the nM range (Ben-Efraim and Gerace, 2001; Catimel et al., 2001; Lam et al., 1999; Lott et al., 2010). HEAT motifs are found in proteins that are involved in various processes, such as vesicle trafficking, nuclear import, cell cycle control, protein biosynthesis or protein phosphorylation (Andrade and Bork, 1995).

2.4.4. Armadillo repeat proteins (ArmRP)

The name armadillo originates from the appearance of a *Drosophila* mutant with a defect in a segment polarity protein (Nüsslein-Volhard and Wieschaus, 1980; Riggleman et al., 1989), which was later identified as *Drosophila* β-catenin (see below). The structure of natural ArmRPs is a right-handed superhelix or solenoid, formed by 4–12 stacked tandem armadillo repeat motifs, each consisting of approximately 42 amino acids folded into three α-helices, named H1, H2, and H3 (see Fig. 2D). The structure of the armadillo repeat (ArmR) is similar to that of the HEAT motif, even though it has three helices while the HEAT repeat has two. ArmR helices H1 and H2 correspond to the strongly bent helix A of the HEAT motif, and helix H3 corresponds to helix B, suggesting that ArmRPs and HEAT proteins have a common ancestor but diverged into different structural families (Andrade et al., 2001b; Kippert and Gerloff, 2009). Natural ArmRPs bind to peptides of various lengths in extended conformation. Depending on the ArmRP subfamily, the target peptide sequence contains conserved positively or negatively charged residues important for the interaction (see consensus target sequence in Fig. 2A and B). ArmRPs interact with their targets via the concave surface formed by adjacent H3 helices.

The binding mode is highly conserved. A polypeptide in extended conformation is bound in an antiparallel orientation relative to the ArmRP, forming an asymmetric double helix (the peptide forming the thinner strand and the ArmRP forming the bigger "complementary strand"). This antiparallel binding mode is maintained by an array of conserved asparagine residues, making hydrogen bonds to the backbone of the extended peptide, and several other residues of the binding surface, contributing interactions to the peptide side chains, mediating the specificity for a peptide sequence (Andrade et al., 2001b; Conti et al., 1998). ArmRPs can reach high affinities down to two digit nM dissociation constants depending on the peptide target (Catimel et al., 2001; Hodel et al., 2001). ArmRPs are abundant in eukaryotes and involved in signaling, nucleocytoplasmic transport and cell adhesion (reviewed in Tewari et al. (2010)).

2.5. Applications of peptide-binding domains

We wish to reiterate that we are limiting our discussions to peptides as targets. Many peptide-binding domains lend themselves to protein engineering, and libraries have been made from many of them (see below and Table 1). Antibodies are by far the most used scaffolds and many libraries have been made from the natural repertoire or were designed using synthetic genes (Mondon et al., 2008). The differences between protein-binding and peptide-binding antibodies have been analyzed, and while each group has features in common, the binding mode for peptides is not conserved, regarding the secondary structure of the peptide, its orientation or contacting residues of the antibody (Almagro, 2004; MacCallum et al., 1996; Wilson et al., 1994). Nonetheless, anti-peptide antibodies with affinities in the low picomolar range can be obtained by directed evolution (Luginbühl et al., 2006; Zahnd et al., 2004).

The main application for anti-peptide antibodies is probably the detection of the unfolded protein from which the peptide has been derived, typically in western blots, and similarly in array formats with spots of denatured proteins on a solid surface. Of particular interest is the detection of posttranslational modifications, which very frequently occur in unstructured regions of proteins. Binders recognizing such intrinsically unstructured regions have the advantage that it is uncritical whether the protein of interest denatures in the assay. The principal applications of non-antibody binders would be the same as described here for antibodies – ultimately the performance and ease of generation will determine which format is most attractive.

MHCs can bind a plethora of different peptides but they have not yet been used as a general peptide binding reagent in biotechnology or biomedicine. Furthermore, they are labor-intensive to produce. The peptide binding pocket of MHC-I is closed at both ends and can thus only harbor peptides of defined length, precluding many applications with whole proteins as targets. MHC-I complexes were engineered as single chain trimers (SCTs), in which the peptide, the heavy chain of MHC-I and β2 microglobulin are connected. They may serve for probing T-cell interactions in vaccination or be used in diagnostics (reviewed in Hansen et al. (2010)). MHC-II complexes were selected by yeast surface co-display with the peptide, allowing to probe and alter peptide-binding specificity (Jiang and Boder, 2010).

For small adaptor domains the recognition surface is usually limited such that only very short sequence stretches can be bound, with the consequence that the affinity is typically only modest. Furthermore, each of the scaffolds has a preference for a sequence motif, as summarized above. Therefore, in order to develop such scaffolds into very general sequence recognition tools, several will have to be combined into one protein. This can be done in a beads-on-a-string-like manner, but the sequence recognized will be discontinuous, and constructs may become prone to intermolecular binding. In many instances, however, it is of interest to use these

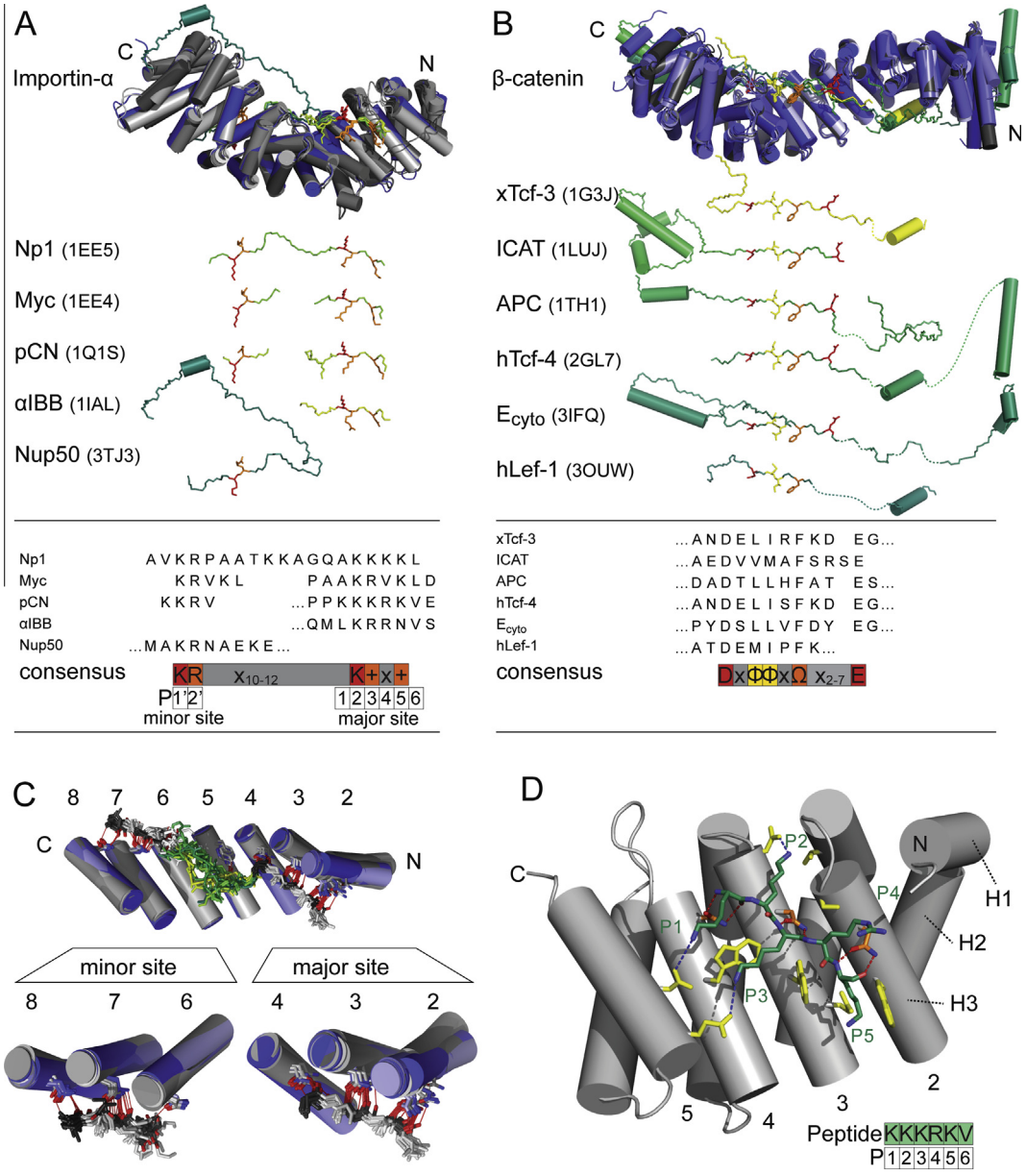

Fig.2. Peptide-binding analysis in natural ArmRPs. The α-helices of the right-handed superhelical structures of different ArmRPs (colored from grey to blue, with N- and C-termini labeled as N and C, respectively) are represented by cylinders. Peptides, bound in antiparallel orientation, are shown in stick representation, colored from yellow to green for different peptides. Dashed lines indicate residues of the peptides not resolved in the crystal structures. (A) Top: Structural superposition of five members of the importin-α family, representing classes of peptides with different structural homology. Middle: structural alignment of superimposed peptides: bipartite *Xenopus laevis* nucleoplasmin (Np1), monopartite human c-myc NLS (Myc), phosphorylated SV40 CN (pCN), importin-β binding domain of importin-α (αIBB) and human nucleoporin Nup50 (Nup50) (PDB ID: 1EE5 (Conti and Kuriyan, 2000), 1EE4 (Conti and Kuriyan, 2000), 1Q1S (Fontes et al., 2003b), 1IAL (Kobe, 1999), 3TJ3 (Pumroy et al., 2012)). Highly conserved peptide residues are highlighted by their side chain (colored from red [most conserved] to yellow). Bottom: sequence alignment of peptides and consensus sequence. (B) Analogous to (A) for members of the β-catenin family: xTcf-3, ICAT, APC, hTcf4, E_{cyto} and hLef1 (PDB ID: 1G3J (Graham et al., 2000), 1LUJ (Graham et al., 2002), 1TH1 (Xing et al., 2004), 2GL7 (Sampietro et al., 2006), 3IFQ (Choi et al., 2009), 3OUW (Sun and Weis, 2011)) (Figure adapted from Choi et al. (2006)). (C) Top: Analysis of conserved peptide binding in importin-α family. All available structures in complex with a peptide were superimposed and peptide-binding surfaces (repeat 2–8 formed by helix 3, with the repeat numbers indicated) are shown, forming two conserved binding sites (major and minor). In both sites, conserved Asn residues (shown as sticks) in each ArmR make hydrogen bonds to the peptide backbone (shown as red dashed lines). Consecutive amino acids of the peptides in the binding sites are colored alternatively dark and light grey, indicating the binding of approximately two amino acids per repeat. Bottom: Detailed view of the above superposition. (D) Detailed view of the major binding site of yeast importin-α (PDB ID 1BK6 (Conti et al., 1998)) in complex with NLS peptide (green), making six backbone hydrogen bonds (red dashed lines) with the conserved Asn residues (orange). Interaction residues of importin-α with the side chains of the peptide are shown in yellow. Ionic interactions are indicated by blue dashed lines.

domains for recognizing variants of their natural sequences, e.g., to identify the specificity of natural interaction partners, and libraries of many scaffolds have been constructed and selections have been performed successfully (see Table 1): SH2 domains (Malabarba et al., 2001), PDZ domains (Ernst et al., 2010; Huang et al., 2008; Junqueira et al., 2003; Reina et al., 2002; Schneider et al., 1999), FHA domains (Pershad et al., 2012), WW domains (Dalby et al., 2000) and SH3 domains (Hiipakka and Saksela, 2007; Hiipakka et al., 1999, 2001; Panni et al., 2002). Of these, SH2 domains and FHA domains are of particular interest as they specifically recognize phosphorylated tyrosine and threonine residues within the peptides, respectively. For practical utility as detection reagents, the challenge will be to increase their affinity without compromising specificity.

Repeat proteins have a longer continuous binding surface, which makes them interesting as recognition agents for longer

Table 1
Summary of studies employing peptide-binding scaffolds to create novel or improved peptide-binding reagents.

Scaffold	Fold (amino acids)	Library diversity	Selection technology	Targets[a]	Affinity of binders	References
Antibody[h] and MHC[h]						
scFv	V_L–V_H (235)	n.d.[b]	Ribosome display	BoPrP	1 pM	Luginbühl et al. (2006)
MHC-II	α1α2/β1β2 (191/190)	1×10^7	Yeast codisplay	FLU	n.d.	Jiang and Boder (2010)
Small adaptor proteins						
SH2	α_2/β_7 (95)	5×10^4	Filter screening	mTag-pY324, PDGFR-pY716, Shc-pY239	1–684 nM	Malabarba et al. (2001)
PDZ	α_2/β_6 (101)	n.d.[b]	Yeast two-hybrid	pa-, pb-, pdrl-, pelk-1-peptide	160–240 nM	Schneider et al. (1999)
PDZ	α_2/β_5 (101)	n.a.[c]	Rational design	Eg5-, pol-, hyd-peptide	1–96 μM	Reina et al. (2002)
PDZ	α_2/β_6 (124)	7×10^5	Yeast two-hybrid	c-Myc	n.d.	Junqueira et al. (2003)
PDZ/FN3[d]	α_1/β_6 (96)/β-sandwich (91)	10^9	Phage display	ARVCF peptide	4–56 nM	Huang et al. (2008)
PDZ	α_1/β_6 (103)	10^9	Phage display	P-WT and 14 peptides	0.03–8 μM	Ernst et al. (2010)
FHA	β_{11}	5×10^9	Phage display	jun-B, ATF2, MAPK1, MAPK3, jun-D	0.9–1.0 μM	Pershad et al. (2012)
WW	β_3 (52)	2.5×10^8	Phage display	WW1 peptide	3.3–102 μM	Dalby et al. (2000)
SH3	β_5 (ca. 60)	10^7	λ-Phage display	Abl1, Src peptide	nM–μM	Panni et al. (2002)
SH3	β_5 (57)	1.37×10^8	Phage display	HIV-1 Nef	down to 7 nM	Hiipakka et al. (1999)
SH3	β_5 (57)	4.2×10^8	Phage display	SIVmac Nef	n.d.	Hiipakka et al. (2001)
SH3	β_5 (57)	4.2×10^8	Phage display	CD3ε, ADAM15, PAK1, Sos1, p85α	n.d.	Hiipakka and Saksela (2007)
Repeat proteins						
TPR	α_2 (18+34·n)[e]	n.a.	Rational design	Hsp90 peptide	200 μM	Cortajarena et al. (2004), (2010a)
TPR	α_2 (18+34·n)	n.a.	Rational design	Hsp90 peptide	1 μM	Cortajarena et al. (2008)
TPR	α_2 (18+34·n)	n.a.	Rational design	Hsp90 mutants	1.2–74 μM	Jackrel et al. (2009)
TPR	α_2 (18+34·n)	n.a.	Rational design	Hsp90 peptide	down to 60 nM	Kajander et al. (2009)
TPR	α_2 (18+34·n)	2.7×10^8	Split GFP reassembly assay	c-Myc tag, Dss1	10–108 μM	Cortajarena et al. (2010b), Jackrel et al. (2010)
TPR	α_2 (18+34·n)	324	Rational design/split GFP assay	PTIEEVD peptide	70 μM	Grove et al. (2010)
HEAT	α_2 (31·n)	$3{\cdot}10^8$	n.a. [f]	n.a.	n.a.	Urvoas et al. (2010)
WD	β_4-propeller (42·n)	n.a.[g]	n.a.	n.a.	n.a.	Nikkhah et al. (2006)
ArmRP	α_3 (71+42·n)	10^{11}	Ribosome display	NT peptide	7 μM	Varadamsetty et al. (2012)

[a] Nomenclature according to original publication.
[b] Not determined.
[c] Not applicable.
[d] FN3: fibronectin type III domain.
[e] *n* Refers to the number of internal repeats.
[f] No selections reported, library created to test effect of randomization on protein properties.
[g] Only consensus design reported.
[h] List of studies employing these peptide-binding scaffolds is reduced to one representative.

peptides (Grove et al., 2008). Additionally, they are usually less constrained towards a preferred sequence motif, compared to small adaptor molecules, which makes them suitable for the design of general peptide-binding scaffolds. Several studies have been conducted where libraries based on repeat scaffolds were designed and binders were selected or designed rationally (see Table 1): beta-propellers (Nikkhah et al., 2006), TPR proteins (Cortajarena et al., 2008, 2010a,b, 2004; Grove et al., 2010; Jackrel et al., 2009, 2010; Kajander et al., 2009), HEAT proteins (Urvoas et al., 2010).

In addition to the advantage of large continuous binding surfaces shared between all rigid tandem repeat proteins, ArmRP show a highly conserved binding mode, with the putative ability to bind peptides in a modular way (Grove et al., 2008; Parmeggiani et al., 2008). It is this combination of properties that make ArmRP highly interesting scaffolds.

3. Armadillo repeat proteins as modular peptide-binding scaffolds

From the comparison of peptide-binding proteins, ArmRPs are of particular interest because they uniquely combine three features in one scaffold. First, ArmRPs are by their nature extended rigid structures based on a modular assembly, second, ArmRPs bind their target peptides in an extended conformation and third, they do so with a conserved modular recognition mechanism. Only this combination would permit modular binding of peptides.

Other repeat proteins, such as beta-propellers, TPR and HEAT proteins have the typical modular scaffold design, but they do not bind the target peptide in a repeat-wise manner. Likewise, small adaptor proteins bind peptides in a highly conserved way, but to combine several different domains rigidly into one protein is difficult (see above).

This feature of modular binding has been described in more detail for proteins from the importin-α and β-catenin subfamily, which belong to the best characterized subfamilies of natural ArmRPs (Coates, 2003; Hatzfeld, 1999). β-Catenin homologs are involved in cell adhesion and signaling, whereas importin-α homologs are involved in the nucleocytoplasmic transport of proteins.

3.1. Importin-α

Importin-α homologs are found across all eukaryotic kingdoms (Coates, 2003; Tewari et al., 2010), where they transport, in the "classic" nuclear import pathway, proteins from the cytoplasm into the nucleus, through nuclear pore complexes (Marfori et al., 2011). Importin-α, consisting of 10 ArmRs, binds to the nuclear localization sequence (NLS) of a target protein. Together with importin-β (a HEAT repeat protein), they form a trimeric cargo protein complex that enters the nucleus.

Classical NLS sequences (cNLS) contain either one or two stretches of basic amino acids, and are therefore grouped into two

classes: monopartite and bipartite NLSs. The negatively charged binding groove of importin-α contains the Asn residue in almost every repeat (Asn37, according to the nomenclature in Parmeggiani et al., 2008) that is characteristic for ArmRPs. It makes conserved bidentate hydrogen bonds to the backbone of the peptide target. In addition, Trp residues at position 33 are also highly conserved in the binding site and their side chains are aligned in an almost parallel arrangement, building hydrophobic pockets to accommodate the aliphatic portion of Lys or Arg side chains of the peptide target (Conti et al., 1998). This Asn and Trp ladder along the binding surface is disrupted in ArmR 5 and 6, thus segregating the binding groove into the "major" (comprised of residues from ArmR 2–4) and the "minor" site (comprised of residues from ArmR 6–8) (Fig. 2C) (Marfori et al., 2011). Bipartite NLSs, defined by the sequence KRx$_{10-12}$K[K/R]x[K/R] (Kosugi et al., 2009), can therefore bind in an extended conformation, contacting both major and minor binding site, with a linker separating the N- and C-terminal basic cluster. Monopartite NLSs, defined by the consensus sequence K[K/R]x[K/R] (Fontes et al., 2003a) can bind to either of the two sites, however, they bind with higher affinity to the major site.

In the past decade, several importin-α structures in complex with NLS peptides have been solved (Catimel et al., 2001; Chen et al., 2005; Conti and Kuriyan, 2000; Conti et al., 1998; Cutress et al., 2008; Dias et al., 2009; Fontes et al., 2000, 2003a,b; Giesecke and Stewart, 2010; Hirano and Matsuura, 2011; Kobe, 1999; Matsuura and Stewart, 2005; Mynott et al., 2011; Pumroy et al., 2012; Takeda et al., 2011; Tarendeau et al., 2007; Yang et al., 2010). Based on structural homologies of the bound peptide, five peptide classes have been defined, represented by (i) the bipartite peptide nucleoplasmin (Np1), and (ii) the monopartite peptides c-myc (Myc), bound to major and minor site; (iii) pCN, a subgroup of c-myc; (iv) αIBB, bound only to the major site and (v) Nup50, bound only to the minor site (Fig. 2A).

In all classes, major and minor binding sites fix peptide backbones in an extended conformation into the binding groove by Asn37 bidentate hydrogen bonds, as illustrated in Fig. 2C. Larger backbone fluctuations of the peptides are visible for linker regions. Target peptide residues, critical for binding the major and minor binding site, are named P1-P6 and P1′-P5′, respectively (Fontes et al., 2003a; Marfori et al., 2012). The side chains of these residues interact with binding pockets on the ArmRP, which are named correspondingly.

The almost strictly conserved Lysine P2 appears to be the most critical binding residue, followed by well-defined P3 and P5 (mainly [K/R]). In all three pockets, a negatively charged residue often makes strong ionic interactions with the target peptide side chains (Fig. 2D). The P4 pocket is more tolerant, and P1 and P6 have little preference for any residue (Fig. 2A: peptide consensus sequence) (Marfori et al., 2011). Sequences flanking the basic clusters, including also phosphorylation modifications, are contributing to the interaction, indicating that importin-α is able to bind a broad range of amino acid sequences (Fontes et al., 2003b; Hübner et al., 1997), although at least one of the basic clusters in the target consensus sequence (KRx$_{10-12}$K+x+, see Fig. 2A) seems to be essential for binding and is found in all NLS sequence families (Kosugi et al., 2009).

3.2. β-Catenin

β-Catenin is involved in adherence junctions, where it binds to the cytoplasmic domain of cadherin cell adhesion molecules (Shapiro and Weis, 2009), but also functions as a transcriptional activator in the Wnt signaling pathway, where it binds to Tcf/Lef family transcription factors. The β-catenin ligand is bound in an extended conformation through the central domain, which consists of 12 ArmRs (Huber et al., 1997). In contrast to importin-α with two separated binding sites, β-catenin has a positively charged continuous binding groove formed by repeats 5–9. Several crystal structures of β-catenin in complex with either protein or peptide targets have been determined including transcription factors xTcf-3 (Graham et al., 2000) or hTcf-4 (Graham et al., 2001; Poy et al., 2001; Sampietro et al., 2006), inhibitor of β-catenin and Tcf (ICAT) (Daniels and Weis, 2002; Graham et al., 2002; Ha et al., 2004), the tumor suppressor protein adenomatous polyposis coli (APC) (Eklof Spink et al., 2001; Xing et al., 2004), E-cadherin (E_{cyto}) (Choi et al., 2009; Huber and Weis, 2001) and lymphoid enhancer-binding factor (Lef1) (Sun and Weis, 2011). A polar ladder, resembling the Asn37-ladder in importin-α, is found in β-catenin, where 7 out of 12 repeats carry Asn, His or Gln residues (Asn in five repeats, His and Gln each in one repeat) (Andrade et al., 2001b).

As shown in Fig. 2B, peptides with a conserved Dxϕϕxϴx$_{2-7}$E motif bind to the protein in extended conformation. Negatively charged Asp and Glu residues in the peptide make strong and specific ionic interactions with two conserved Lys residues (named "charged buttons": Lys435 and Lys312) on the binding surface of the β-catenin (Graham et al., 2000; Sun and Weis, 2011; Xu and Kimelman, 2007). Only for Lef-1, no interaction has been found for the conserved Glu, highlighted in the disordered region of the peptide. Additional interaction sites, apart from the conserved binding groove of β-catenin, can be involved in target binding, increasing affinity and specificity (Choi et al., 2006).

3.3. Modular peptide binding

In both subfamilies, ArmRP exhibit, apart from variable binding sites for side chains, a conserved binding mechanism toward peptides in extended conformation, including strong hydrogen bonding with Asn residues at position 37 (Asn ladder along the concave binding surface) to the peptide backbone. Superposition of several peptide complexes indicates that, with the conserved fixation of the backbone at every second peptide bond of the peptide, approximately two amino acids can be bound per repeat, as illustrated for importin-α in Fig. 2D.

The repetitive structure of ArmRPs, where repeats are packed tightly on each other, allows binding of consecutive dipeptide units. As these dipeptide units are directly connected, the resulting oligopeptide can be bound by a conserved and modular binding mechanism over the ArmRP scaffold, created by several repeats. However, in all natural ArmRP structures, this modular fixation of the peptide backbone by Asn37 bidentate hydrogen bonds is limited in space to three consecutive repeats. As illustrated in Fig. 2D, yeast importin-α in complex with the NLS peptide (Conti et al., 1998) shows the longest uninterrupted conserved backbone binding, forming six hydrogen bonds. Based on the analysis of curvature variations between two neighboring natural ArmRs, we expect conserved binding to more than three repeats could be achievable.

ArmRPs appear to allow binding to a broad spectrum of peptide sequences. One limitation of the conserved binding mechanism of ArmRP is that the binding of polyproline sequences would not be possible in a modular mode, since the backbone nitrogen is involved in the formation of the pentameric ring and is thereby not accessible to Asn37. However, several aromatic residues arranged in parallel, found in the Trp ladder in importin-α, resemble the binding sites of those PRDs which allow binding of a proline-rich peptide in PPII-helix conformation (Zarrinpar et al., 2003). Another limitation, in the initial phases of engineering, i.e., when staying close to the natural ArmRP sequences, is the apparent necessity of charged amino acids, as seen in NLS sequences binding to importin-α (positively charged Arg or Lys residues (Kosugi et al., 2009)), or in β-catenin binding to conserved negatively charged Glu and Asp residues. These critical residues for interaction have been termed "hot spot residues".

It will have to be evaluated, if specific binding with similar affinities as found in natural ArmRP is possible to peptides in extended conformation with a limited number or even no charged amino acids. From the analysis of a large set of peptide-binding structures unrelated to ArmRPs, the most frequently used peptide hot-spot residues are aromatic residues, leucine and isoleucine (London et al., 2010). Therefore, we expect that binding to peptides without charged hotspots should be possible in principle for designed ArmRPs.

Screening and selection for individual repeats with affinity for a specific dipeptide (in context of a longer peptide) would ultimately provide a collection of single repeats with predetermined specificity (Fig. 3). Such single pre-selected repeats (building blocks) could be reassembled into a tandem repeat protein with a defined binding specificity. A modular approach could effectively bypass the current *in vitro* selection procedure for every individual target peptide and binders could be generated by rapid gene assembly instead of new selections for every target peptide.

4. Design of armadillo repeat proteins towards stability

Consensus design approaches, where a multiple alignment of a family of homologous proteins is used to calculate a consensus sequence, have been applied successfully in the generation of synthetic antibody libraries (Knappik et al., 2000) and later for repeat proteins, including TPR (D'Andrea and Regan, 2003) and DARPins (Forrer et al., 2004).

For ArmRPs an approach based on consensus repeat design was insufficient to engineer proteins with the desired biophysical properties, including protein expression, folding, solubility and stability (Parmeggiani et al., 2008). Several additional steps were necessary, using computational (Alfarano et al., 2012; Parmeggiani et al., 2008) and structural protein engineering approaches (Madhurantakam et al., 2012), to obtain ArmRP with improved characteristics compared to natural ArmRP and with a complete and stable α-solenoid fold. This iterative process of protein stability engineering, which involved a combination of consensus design, homology and force-field based modeling, molecular dynamics simulation, NMR and X-ray crystallography is summarized in Fig. 5 and described in more detail below.

The rather low sequence similarity between members of the ArmRP family (e.g., between the structurally aligned human importin-α and β-catenin consensus, the identity is only 26%) and the significant deviations between individual repeat sequences (identity as low as 30% (Hatzfeld, 1999)) complicate the definition of an overall armadillo consensus and the identification of cross-species homologs of ArmRP, although some advances in prediction algorithms have been made (Andrade et al., 2000, 2001b; Kippert and Gerloff, 2009).

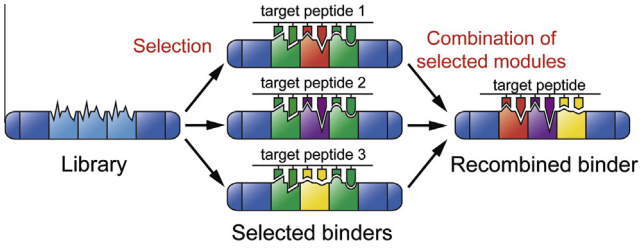

Fig.3. The modular repeat architecture of ArmRPs allows binding to uninterrupted longer peptides. Screening and selection of surface-randomized repeats (light blue) against different peptide targets would provide a collection of single repeats with predetermined specificity (colored red, violet and yellow). Such predetermined units would then be reassembled into a novel repeat protein, exhibiting a defined binding specificity.

In a first step, a consensus repeat sequence was defined, derived from a combination of the subfamilies importin-α and β-catenin (termed C-type for combined in Parmeggiani et al., 2008). A set of artificial short consensus ArmRPs with the overall constitution of $Y_zC_xA_z$ was generated. Y denotes an N-terminal capping repeat derived from yeast importin-α, A is an artificial C-terminal capping repeat and subscript *z* refers to the design cycle ("generation") of the capping repeats (see below), given as a roman numeral (I–III). Subscript *x* is the number of (identical) internal repeats of type C present in the protein. This first consensus design led to well-expressed but molten globule-like proteins ($Y_IC_4A_I$) (see Table 2) (Parmeggiani et al., 2008). Alternatively, proteins with internal repeats based on a consensus design taken only from importin-α (I-type) or β-catenin (T-type) were created, and they produced stable albeit dimeric proteins, and were thus not further pursued.

The stability of ArmRPs is mediated mainly by the hydrophobic core packing, through nonpolar interactions. Molten globule-like features are expected to be a result of non-optimal packing of this hydrophobic core (Munson et al., 1996). The consensus-designed ArmRP with molten globule-like properties ($Y_IC_4A_I$) was converted into a monomeric, stable, folded protein ($Y_IM_4A_I$) with cooperative unfolding behavior by stabilization of the hydrophobic core (Table 2). Three point mutations (G12A, A32L and L38I) were introduced into the hydrophobic core in every internal repeat (named M-type), based on a computational modeling approach in combination with molecular dynamics (MD) simulations (Figs. 4 and 5) (Parmeggiani et al., 2008). All newly introduced residues are found in the consensus sequence of the importin-α subfamily, indicating that the M-type sequence resembles more the importin-α than the β-catenin subfamily, which is consistent with the overrepresentation of importin-α (133) over β-catenin (110) sequences for the combined consensus design. Importantly, the M-type proteins are monomeric.

Heteronuclear NMR experiments with protein $Y_IM_4A_I$ revealed an electrostatic repulsion between two closely spaced lysine residues in every internal repeat (K26 and K29) (Fig. 4). Mutation of these two lysines to glutamines (to give the so-called $\overline{M}$-type repeat) resulted in a pH-independent, more stable packing of the novel ArmRP $Y_I\overline{M}_4A_I$ (see Fig. 5 and Table 2) (Alfarano et al., 2012).

MD simulations based on homology models of $Y_IM_4A_I$ have shown an increased conformational instability in the capping repeats (Y_I and A_I), compared to the internal repeat (M or $\overline{M}$) (Alfarano et al., 2012). To reduce this flexibility, several cap mutations were tested in silico and validated experimentally. In the C-cap, the Q38L and F39Q point mutations increased the protein stability in temperature-induced unfolding experiments by 7 °C compared to the original protein, whereas an increase of 1.5 °C was achieved by introducing mutations V33R and R36S and by deletion of residue R41 in the N-cap (see Fig. 4). These stabilizing effects from the N- and C-terminal caps (named Y_{II} and A_{II} for second generation) were shown to be additive (Alfarano et al., 2012) and the thermal and chemical stability of the final construct $Y_{II}\overline{M}_4A_{II}$ (with improved caps) was increased by 9.5 °C and by 0.7 M GdmCl, respectively (Table 2).

Protein $Y_{II}\overline{M}_4A_{II}$, including MD-designed capping repeats and internal repeats with removed electrostatic repulsions, was the first designed ArmRP for which a crystal structure could be solved with a resolution of 2.5 Å (Madhurantakam et al., 2012). As predicted by MD simulation and verified experimentally, the most stability-relevant residue introduced by mutation was Q38L in the C-cap (A_{II}). It fills the hydrophobic pocket between the last internal repeat and the C-cap and thereby inhibits the formation of a small cavity (Alfarano et al., 2012). Yet, this crystal structure also revealed that the N-cap (Y_{II}) was not folded correctly onto the remainder of the protein, as the loop connecting the N-cap with the first internal repeat was not formed, and instead the chain

Table 2
Biophysical properties of designed ArmRPs.

Constructs[a]	Refs.	Residues (repeats)[b]	pI[c]	MW_{calc} (kDa)[d]	Oligomeric state[e]	MW_{obs} (kDa)[f]	$MW_{obs/calc}$[g]	CD_{222} (MRE)[h]	T_m (°C)[i]	CD GdmCl (M)[j]
$Y_IC_4A_I$	1	253 (6)	5.1	26.9	Monomer	50.2[k]	1.86	−12,393[k]	n.d.	n.d.
$Y_IM_4A_I$	1, 2	253 (6)	5.1	27.1	Monomer	32.3[k]	1.19	−19,255[k]	71	3.5
$Y_I\overline{M}_4A_I$	2	253 (6)	4.5	27.1	Monomer	32.3[k]	1.19	−19,162[k]	76	3.7
$Y_{II}\overline{M}_4A_{II}$	2	252 (6)	4.5	26.9	Monomer	31.2[k]	1.16	−20,401[k]	86	4.4
$Y_{II}\overline{M}_3A_{II}$	2, 3	210 (5)	4.6	22.6	Monomer	27.5[k]	1.22	−19,015[k]	77	3.6
$Y_{III}\overline{M}_3A_{II}$	3	210 (5)	4.8	22.5	Monomer	27.4[k]	1.22	−20,259[k]	81	3.8
$Y_I\overline{M}_5A_I$	4	295 (7)	4.4	31.5	Monomer	38.6[k]	1.23	−20,435[k]	80	4.2
VG_328[l]	4	295 (7)	4.4	31.7	Monomer	39.9[k]	1.26	−20,199[k]	74	3.3
Importin-α[m]	1	435 (10)	5.5	48.2	Monomer	43.0[k]	0.9	−14,646[k]	43	n.d.
β-Catenin[n]	1	528 (12)	8.7	57.6	Monomer	52.8[k]	0.9	−17,207[k]	58	n.d.

[1] Parmeggiani et al. (2008); [2] Alfarano et al. (2012); [3] Madhurantakam et al. (2012); [4] Varadamsetty et al. (2012).
[a] Capping repeats ($Y_I/Y_{II}/Y_{III}$ and A_I/A_{II}) and internal repeats (C, M, $\overline{M}$) are given in Fig. 4.
[b] The number of residues includes the $MRGSH_6GS$ tag; the number of repeats includes capping repeats.
[c] Isoelectric point (pI).
[d] Molecular weight calculated from the sequence; masses were confirmed by mass spectrometry.
[e] Oligomeric state as indicated by multi-angle static light scattering.
[f] Observed molecular weight as determined by size exclusion chromatography.
[g] Ratio between observed (size exclusion chromatography) and calculated molecular weight ($MW_{obs/calc.}$).
[h] Mean residue ellipticity at 222 nm expressed as deg·cm^2/dmol.
[i] Transition midpoint (T_m) observed in thermal denaturation measured by CD.
[j] Midpoint of transition in GdmCl-induced denaturation, measured by CD.
[k] Normalized to value from Alfarano et al. (2012).
[l] Binder VG_328 is derived from a $Y_I\overline{M}L_3\overline{M}A_I$ library, where L is a randomized library module.
[m] Armadillo domain of human importin-α1.
[n] Armadillo domain of mouse β-catenin.

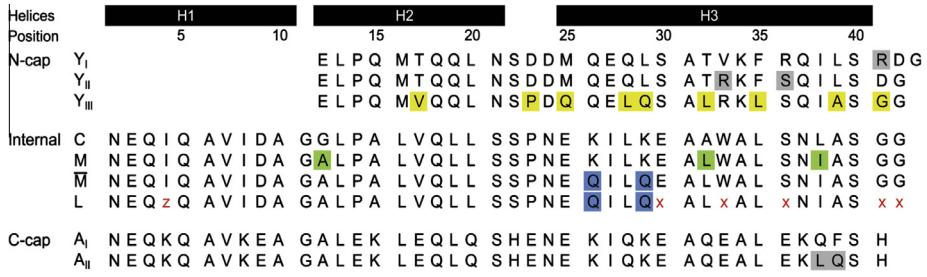

Fig.4. Sequence alignment of the N-terminal capping repeats (Y_I, Y_{II} and Y_{III}), internal repeats (C, M, $\overline{M}$ and L) and C-terminal repeats (A_I and A_{II}). Mutations introduced in the internal repeats to improve packing of the hydrophobic core and decrease charge repulsion on the surface are colored green and blue, respectively. Residues mutated or deleted in the terminal caps upon MDS analysis are colored grey. N-terminal cap mutations introduced to prevent domain swapping are colored yellow. The nomenclature for designed ArmRPs assigns two α-helices (shown in black rectangles) for the N-terminal cap, and three for the internal repeats and C-cap, respectively. Experimentally tested proteins consisted always of an N-cap, several internal repeats and a C-cap (e.g. $Y_IM_4A_I$). Library module L is based on internal repeat $\overline{M}$ and contains in total six randomized residues (z: E,H,K,I,Q,T or R, and x: all 20 amino acids except P,C and G).

adopted a continuous α-helix. Thereby N-caps were involved in a domain-swapping event between symmetry-related molecules. The consensus design strategy had apparently removed a special feature within the first repeat of natural consensus proteins, which encode at the beginning of the first internal repeat a helix breaker.

The N-cap was therefore redesigned, based on the sequence of the well-folded internal repeats. Nine point mutations were introduced, including a helix breaker, D41G in the linker between the N-cap and the first internal repeat (and this redesigned N-cap was named Y_{III}) (Figs. 4 and 5) (Madhurantakam et al., 2012). A crystal structure of an ArmRP ($Y_{III}\overline{M}_3A_{II}$) with the redesigned N-cap (Y_{III}) was determined, showing that the N-cap now folded correctly on the remainder of the protein (Fig. 5). The Y_{III}-cap further improved thermal and chemical stability of the protein by 4.5 °C and 0.2 M GdmCl in the denaturation midpoint, respectively. Additional stability could be gained by increasing the number of internal repeats, as had been observed for other repeat proteins (Kajander et al., 2005; Parmeggiani et al., 2008; Tripp and Barrick, 2007; Varadamsetty et al., 2012; Wetzel et al., 2008).

Natural ArmRPs (β-catenin and importin-α) and the designed ArmRPs with reengineered caps ($Y_{III}\overline{M}_3A_{II}$) share the typical solenoid fold of this protein class, as intended by the design, but thermal and chemical stability is higher for the designed proteins (Table 2). The structure of the internal repeats of $Y_{III}\overline{M}_3A_{II}$ is similar to the minor NLS-binding site of importin-α (PDB ID 1BK6) with a RMSD of 0.71 Å (Madhurantakam et al., 2012). Even though most residues responsible for binding to the NLS peptide (KKKRKV) are identical or similar in both structures (notably the conserved Asn37 and Trp33 residues), the conserved binding pocket P1′ is missing in the designed ArmRP. The absence of this critical binding pocket could explain the observed low affinity of designed ArmRPs of that generation to the NLS peptide, measured by ELISA. Stronger binding signals were obtained if the number of internal repeats was increased from three to five repeats (unpublished data). Based on the new structures of designed ArmRPs, it is possible to engineer stronger binding to the NLS peptide and to find peptides ideal for binding (unpublished data). Nevertheless, to reach the ultimate goal of obtaining binders against a broad spectrum of peptides,

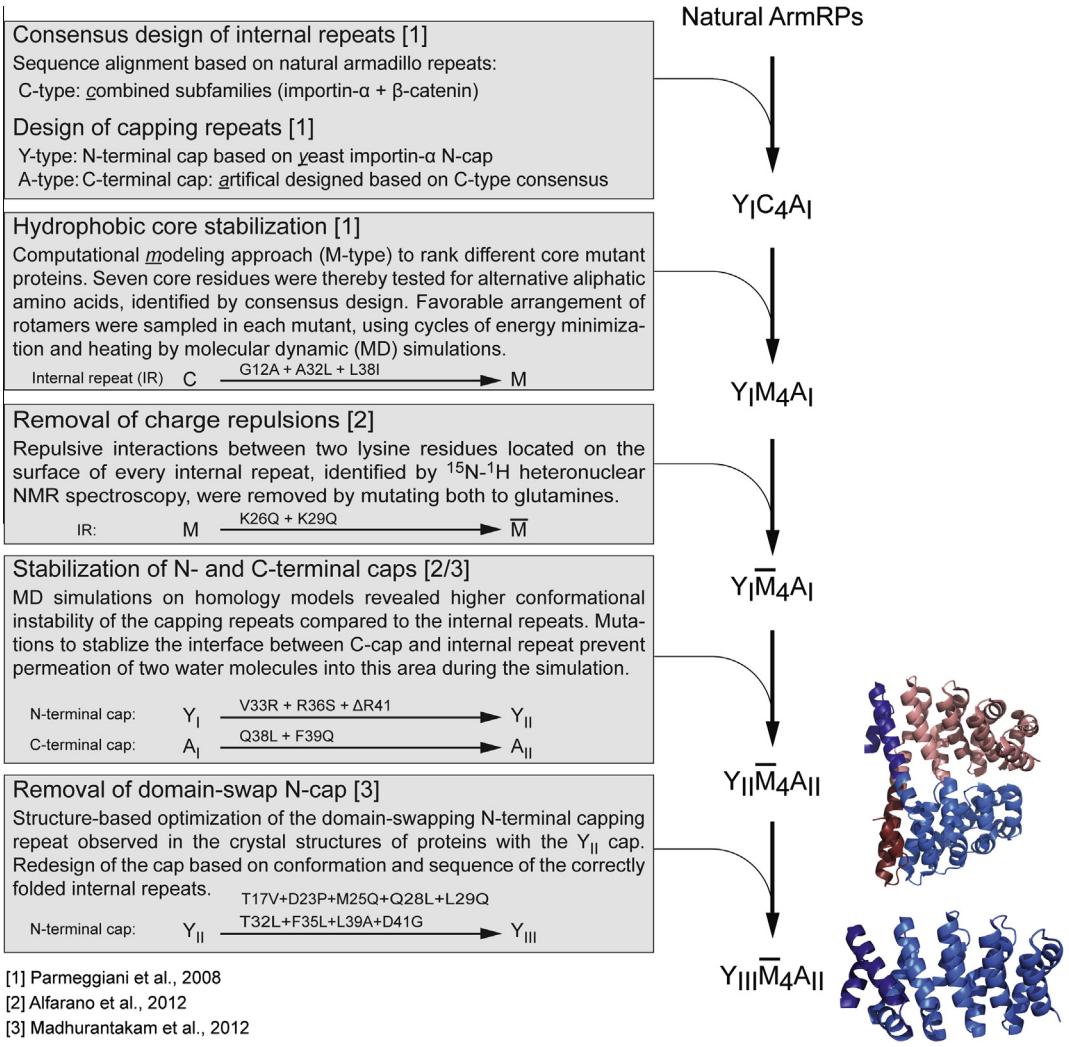

Fig.5. Development of designed ArmRPs towards stability, summarized in five protein engineering steps. Stability, measured by heat- and GdmCl-denaturation, increases from $Y_IC_4A_I$ towards $Y_{II}\overline{M}_4A_{II}$ constantly. Designed proteins (e.g., $Y_zC_xA_z$) are composed of an N-terminal capping repeat (Y), a defined number (subscript *x*) of identical internal repeats (e.g., type C) and a C-terminal capping repeat (A). The subscript *z* is a roman number (I–III) indicating the generation of the capping repeats. Crystal structures are shown for $Y_{II}\overline{M}_3A_{II,}$ with a domain swapped N-terminal cap and $Y_{III}\overline{M}_3A_{II}$ (PDB ID 4DBA and 4DB6).

rational engineering will have to be complemented with directed evolution (see next section).

5. Selection for binding

In many solenoid proteins, the binding surface for peptides is formed by secondary structure elements, whereas in most non-solenoid peptide-binding proteins binding is mediated through variable loops, as seen in antibodies (Sundberg, 2009), SH2- (Yaffe and Smerdon, 2004), PDZ- (Kaneko et al., 2011), FHA- (Mahajan et al., 2008), SH3- (Kaneko et al., 2011), WW- (Zarrinpar et al., 2003) and GYF-domains (Kofler and Freund, 2006). In ArmRP, as described above, the binding surface is formed mainly by the helices H3, comprising solvent exposed residues for target binding along with residues forming the hydrophobic core. The first combinatorial library was based on the $Y_I\overline{M}_xA_I$ design and residue positions for randomization (position 4, 30, 33, 36, 40 and 41) were identified from structures of complexes of natural armadillo proteins and bound target peptides (Varadamsetty et al., 2012). Residues 30–40 are located in helix H3, residue 4 in helix H1 and residue 41 in a loop after helix H3, respectively. These six positions were randomized in each library repeat (and these randomized repeats are called L-type), but only a selection of amino acids was allowed (see Fig. 4). Randomization may lead to decreased helical propensity, surface-exposed hydrophobic residues or introduction of electrostatic repulsions, which can all contribute to poorer biophysical properties. Therefore, the right randomization strategy and the intrinsic stability of the scaffold are very important components, since incorrectly folded library members tend to expose hydrophobic regions of the core, which can result in unspecific binding, diminishing the efficacy of the selection.

To increase the stability of the library members, two non-randomized internal repeats flanking the three randomized repeats were introduced into the library (to result in the $Y_I\overline{M}L_3\overline{M}A_I$ library format) (Varadamsetty et al., 2012). The final experimental diversity of the library was estimated to be greater than 10^{11} (theoretical diversity $\sim 10^{21}$). Characterization of randomly picked clones from the library revealed that almost all were monomeric and stable, showing that the designed scaffold is stable enough to sustain randomization.

Neurotensin (NT, QLYENKPRRPYIL), a 13-amino-acid peptide ligand of the G-protein coupled neurotensin receptor, was used as target peptide in a first selection, since NT has no secondary

structure in solution (Nieto et al., 1986). Two binders against NT differing in only a single residue on the binding surface were isolated by ribosome display. Among them, binder VG_328 showed better binding and was thus further analyzed.

VG_328 has biophysical characteristics comparable to ArmRPs with the non-randomized consensus sequence (see Table 2) and showed good specificity towards NT. Alanine scanning experiments identified four residues of NT (Pro7, Arg8, Arg9 and Tyr11) that are the main contributors to binding. The dissociation constant of binder VG_328 was determined by surface plasmon resonance to be approximately 7 μM (Table 1). This moderate affinity can be explained by only four amino acid side chains of the peptide contributing to binding and by the fact that selection emphasized specificity and not affinity. However, it was shown that VG_328 could still be used as detection agent in Western blots or in pull-down experiments (Varadamsetty et al., 2012). The details of binding will have to await a structure determination of the complex by crystallography or NMR.

While the selection of a binder from the ArmRP library constituted an important proof of principle and resulted in a protein with residue-specific sequence recognition, the affinity was moderate, and the peptide does not necessarily bind in the conserved binding orientation. It should be noted that this first selection has been carried out in a "classical" manner, typical for any selection on any scaffold, i.e., by using a diverse library across three adjacent repeats, and without particular constraints regarding binding orientation.

Therefore, current developments of this technology are concentrating on the uniqueness of the ArmRP scaffold, by using information from recently determined structures of engineered ArmRP-peptide complexes (unpublished data), by enforcing the antiparallel peptide orientation during selections, and by exploiting the modularity of the ArmRP-peptide interaction.

6. Conclusions

Peptide-protein interactions, which are essential for many cellular processes like signaling and protein transport, are mediated by several classes of peptide-binding domains. Small adaptor proteins, although specific for their peptide target, have usually a lower affinity and a narrower recognition spectrum than repeat proteins with their elongated binding surface. From the comparison of peptide-binding proteins, ArmRPs were identified as the ones to allow modular binding to peptides in extended conformation. Modular binding would be the prerequisite to efficiently bypass the current bottleneck of generating peptide-binding proteins by selections against each peptide individually.

Based on a consensus design approach in combination with several methods for protein stability engineering, stable, correctly folded and highly expressed ArmRPs have been designed. The first specific binders were selected *in vitro* from a combinatorial library of the first generation. The designed ArmRP of the latest generation are stable enough for various applications, as shown by several biophysical methods, and crystal structures confirm the design (Madhurantakam et al., 2012).

However, some characteristics of designed ArmRP still have to be analyzed in detail and further optimized: An optimal and uniform curvature is essential to allow continuous binding of extended peptides over several repeats. In natural ArmRP, the longest stretch of a continuous and modular bound peptide is limited to six amino acids as found in the crystal structure of importin-α. Ideally, curvature-optimized designed ArmRPs should be able to bind longer peptides in a modular way along the whole elongated binding surface.

The optimization of the ArmRP design was an iterative process which improved the stability of the ArmRPs in a stepwise manner. Many different approaches have collectively contributed to the final design: homology and force-field based modeling, molecular dynamics simulations, NMR, X-ray crystallography, several biophysical methods and directed evolution approaches. In our opinion, this work is an example of a successful collaboration between theoretical approaches to optimize the design, experimental verification and many state-of-the-art techniques of protein engineering. We are convinced that the combination of all these methods (and more) will be needed to reach this challenging goal of creating a system of modular peptide recognition.

Acknowledgments

Funding in the author's laboratory was through the Swiss National Science Foundation (SNF), initially through the NCCR Structural Biology, later through the Sinergia project CRSI33_122686. We thank Dr. Viola Günther and Christina Ewald for critical reading of the manuscript.

References

Aasland, R., Abrams, C., Ampe, C., Ball, L.J., Bedford, M.T., et al., 2002. Normalization of nomenclature for peptide motifs as ligands of modular protein domains. FEBS Lett. 513, 141–144.
Adams, J., Kelso, R., Cooley, L., 2000. The kelch repeat superfamily of proteins: propellers of cell function. Trends Cell Biol. 10, 17–24.
Alfarano, P., Varadamsetty, G., Ewald, C., Parmeggiani, F., Pellarin, R., et al., 2012. Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. Protein Sci. 21, 1298–1314.
Almagro, J.C., 2004. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. J. Mol. Recognit. 17, 132–143.
Andrade, M.A., Bork, P., 1995. HEAT repeats in the Huntington's disease protein. Nat. Genet. 11, 115–116.
Andrade, M.A., Ponting, C.P., Gibson, T.J., Bork, P., 2000. Homology-based method for identification of protein repeats using statistical significance estimates. J. Mol. Biol. 298, 521–537.
Andrade, M.A., Perez-Iratxeta, C., Ponting, C.P., 2001a. Protein repeats: structures, functions, and evolution. J. Struct. Biol. 134, 117–131.
Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W., Bork, P., 2001b. Comparison of ARM and HEAT protein repeats. J. Mol. Biol. 309, 1–18.
Ball, L.J., Jarchau, T., Oschkinat, H., Walter, U., 2002. EVH1 domains: structure, function and interactions. FEBS Lett. 513, 45–52.
Ben-Efraim, I., Gerace, L., 2001. Gradient of increasing affinity of importin beta for nucleoporins along the pathway of nuclear import. J. Cell Biol. 152, 411–417.
Binz, H.K., Amstutz, P., Plückthun, A., 2005. Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. 23, 1257–1268.
Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., et al., 1994. A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. J. Biol. Chem. 269, 32031–32034.
Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., et al., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509–1512.
Boersma, Y.L., Plückthun, A., 2011. DARPins and other repeat protein scaffolds: advances in engineering and applications. Curr. Opin. Biotechnol. 22, 849–857.
Byeon, I.J., Li, H., Song, H., Gronenborn, A.M., Tsai, M.D., 2005. Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. Nat. Struct. Mol. Biol. 12, 987–993.
Caravella, J., Lugovskoy, A., 2010. Design of next-generation protein therapeutics. Curr. Opin. Chem. Biol. 14, 520–528.
Catimel, B., Teh, T., Fontes, M.R., Jennings, I.G., Jans, D.A., et al., 2001. Biophysical characterization of interactions involving importin-alpha during nuclear import. J. Biol. Chem. 276, 34189–34198.
Cesareni, G., Gimona, M., Sudol, M., Yaffe, M. (Eds.), 2005. Modular Protein Domains. Wiley-VCH, Weinheim.
Chen, M.H., Ben-Efraim, I., Mitrousis, G., Walker-Kopp, N., Sims, P.J., et al., 2005. Phospholipid scramblase 1 contains a nonclassical nuclear localization signal with unique binding site in importin alpha. J. Biol. Chem. 280, 10599–10606.
Chen, C.K., Chan, N.L., Wang, A.H., 2011. The many blades of the beta-propeller proteins: conserved but versatile. Trends Biochem. Sci. 36, 553–561.
Choi, H.J., Huber, A.H., Weis, W.I., 2006. Thermodynamics of beta-catenin-ligand interactions: the roles of the N- and C-terminal tails in modulating binding affinity. J. Biol. Chem. 281, 1027–1038.
Choi, H.J., Gross, J.C., Pokutta, S., Weis, W.I., 2009. Interactions of plakoglobin and beta-catenin with desmosomal cadherins: basis of selective exclusion of alpha- and beta-catenin from desmosomes. J. Biol. Chem. 284, 31776–31788.

Chook, Y.M., Blobel, G., 1999. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nature 399, 230–237.
Christinck, E.R., Luscher, M.A., Barber, B.H., Williams, D.B., 1991. Peptide binding to class I MHC on living cells and quantitation of complexes required for CTL lysis. Nature 352, 67–70.
Cingolani, G., Petosa, C., Weis, K., Müller, C.W., 1999. Structure of importin-beta bound to the IBB domain of importin-alpha. Nature 399, 221–229.
Coates, J.C., 2003. Armadillo repeat proteins: beyond the animal kingdom. Trends Cell Biol. 13, 463–471.
Collins, R.E., Northrop, J.P., Horton, J.R., Lee, D.Y., Zhang, X., et al., 2008. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. Nat. Struct. Mol. Biol. 15, 245–250.
Collis, A.V., Brouwer, A.P., Martin, A.C., 2003. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J. Mol. Biol. 325, 337–354.
Conti, E., Kuriyan, J., 2000. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure 8, 329–338.
Conti, E., Uy, M., Leighton, L., Blobel, G., Kuriyan, J., 1998. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell 94, 193–204.
Corper, A.L., Stratmann, T., Apostolopoulos, V., Scott, C.A., Garcia, K.C., et al., 2000. A structural framework for deciphering the link between I-Ag7 and autoimmune diabetes. Science 288, 505–511.
Cortajarena, A.L., Kajander, T., Pan, W., Cocco, M.J., Regan, L., 2004. Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. Protein Eng. Des. Sel. 17, 399–409.
Cortajarena, A.L., Yi, F., Regan, L., 2008. Designed TPR modules as novel anticancer agents. ACS Chem. Biol. 3, 161–166.
Cortajarena, A.L., Wang, J., Regan, L., 2010a. Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand. FEBS J. 277, 1058–1066.
Cortajarena, A.L., Liu, T.Y., Hochstrasser, M., Regan, L., 2010b. Designed proteins to modulate cellular networks. ACS Chem. Biol. 5, 545–552.
Couture, J.F., Collazo, E., Trievel, R.C., 2006. Molecular recognition of histone H3 by the WD40 protein WDR5. Nat. Struct. Mol. Biol. 13, 698–703.
Cutress, M.L., Whitaker, H.C., Mills, I.G., Stewart, M., Neal, D.E., 2008. Structural basis for the nuclear import of the human androgen receptor. J. Cell Sci. 121, 957–968.
Dalby, P.A., Hoess, R.H., DeGrado, W.F., 2000. Evolution of binding affinity in a WW domain probed by phage display. Protein Sci. 9, 2366–2376.
D'Andrea, L.D., Regan, L., 2003. TPR proteins: the versatile helix. Trends Biochem. Sci. 28, 655–662.
Daniels, D.L., Weis, W.I., 2002. ICAT inhibits beta-catenin binding to Tcf/Lef-family transcription factors and the general coactivator p300 using independent structural modules. Mol. Cell 10, 573–584.
Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., et al., 2012. Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335, 720–723.
Dias, S.M., Wilson, K.F., Rojas, K.S., Ambrosio, A.L., Cerione, R.A., 2009. The molecular basis for the regulation of the cap-binding complex by the importins. Nat. Struct. Mol. Biol. 16, 930–937.
Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., et al., 2008. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. Front. Biosci. 13, 6580–6603.
Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., et al., 1996. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85, 1067–1076.
Durocher, D., Jackson, S.P., 2002. The FHA domain. FEBS Lett. 513, 58–66.
Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., et al., 2000. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. Mol. Cell 6, 1169–1182.
Eklof Spink, K., Fridman, S.G., Weis, W.I., 2001. Molecular mechanisms of beta-catenin recognition by adenomatous polyposis coli revealed by the structure of an APC-beta-catenin complex. EMBO J. 20, 6203–6212.
Ernst, A., Gfeller, D., Kan, Z., Seshagiri, S., Kim, P.M., et al., 2010. Coevolution of PDZ domain-ligand interactions analyzed by high-throughput phage display and deep sequencing. Mol. Biosys. 6, 1782–1790.
Fahnestock, M.L., Johnson, J.L., Feldman, R.M., Tsomides, T.J., Mayer, J., et al., 1994. Effects of peptide length and composition on binding to an empty class I MHC heterodimer. Biochemistry 33, 8149–8158.
Ferrières, G., Villard, S., Pugniere, M., Mani, J.C., Navarro-Teulon, I., et al., 2000. Affinity for the cognate monoclonal antibody of synthetic peptides derived from selection by phage display. Role of sequences flanking thebinding motif. Eur. J. Biochem. 267, 1819–1829.
Fontes, M.R., Teh, T., Kobe, B., 2000. Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. J. Mol. Biol. 297, 1183–1194.
Fontes, M.R., Teh, T., Jans, D., Brinkworth, R.I., Kobe, B., 2003a. Structural basis for the specificity of bipartite nuclear localization sequence binding by importin-alpha. J. Biol. Chem. 278, 27981–27987.
Fontes, M.R., Teh, T., Toth, G., John, A., Pavo, I., et al., 2003b. Role of flanking sequences and phosphorylation in the recognition of the simian-virus-40 large T-antigen nuclear localization sequences by importin-alpha. Biochem. J. 375, 339–349.
Forrer, P., Binz, H.K., Stumpp, M.T., Plückthun, A., 2004. Consensus design of repeat proteins. ChemBioChem 5, 183–189.
Freund, C., Kühne, R., Yang, H., Park, S., Reinherz, E.L., et al., 2002. Dynamic interaction of CD2 with the GYF and the SH3 domain of compartmentalized effector molecules. EMBO J. 21, 5985–5995.
Froloff, N., Windemuth, A., Honig, B., 1997. On the calculation of binding free energies using continuum methods: application to MHC class I protein-peptide interactions. Protein Sci. 6, 1293–1301.
Giesecke, A., Stewart, M., 2010. Novel binding of the mitotic regulator TPX2 (target protein for Xenopus kinesin-like protein 2) to importin-alpha. J. Biol. Chem. 285, 17628–17635.
Graham, T.A., Weaver, C., Mao, F., Kimelman, D., Xu, W., 2000. Crystal structure of a beta-catenin/Tcf complex. Cell 103, 885–896.
Graham, T.A., Ferkey, D.M., Mao, F., Kimelman, D., Xu, W., 2001. Tcf4 can specifically recognize beta-catenin using alternative conformations. Nat. Struct. Biol. 8, 1048–1052.
Graham, T.A., Clements, W.K., Kimelman, D., Xu, W., 2002. The crystal structure of the beta-catenin/ICAT complex reveals the inhibitory mechanism of ICAT. Mol. Cell 10, 563–571.
Grove, T.Z., Cortajarena, A.L., Regan, L., 2008. Ligand binding by repeat proteins: natural and designed. Curr. Opin. Struct. Biol. 18, 507–515.
Grove, T.Z., Hands, M., Regan, L., 2010. Creating novel proteins by combining design and selection. Protein Eng. Des. Sel. 23, 449–455.
Ha, N.C., Tonozuka, T., Stamos, J.L., Choi, H.J., Weis, W.I., 2004. Mechanism of phosphorylation-dependent binding of APC to beta-catenin and its role in beta-catenin degradation. Mol. Cell 15, 511–521.
Hansen, T.H., Connolly, J.M., Gould, K.G., Fremont, D.H., 2010. Basic and translational applications of engineered MHC class I proteins. Trends Immunol. 31, 363–369.
Harrison, S.C., 1996. Peptide-surface association: the case of PDZ and PTB domains. Cell 86, 341–343.
Hatzfeld, M., 1999. The armadillo family of structural proteins. Int. Rev. Cytol. 186, 179–224.
Hiipakka, M., Saksela, K., 2007. Versatile retargeting of SH3 domain binding by modification of non-conserved loop residues. FEBS Lett. 581, 1735–1741.
Hiipakka, M., Poikonen, K., Saksela, K., 1999. SH3 domains with high affinity and engineered ligand specificities targeted to HIV-1 Nef. J. Mol. Biol. 293, 1097–1106.
Hiipakka, M., Huotari, P., Manninen, A., Renkema, G.H., Saksela, K., 2001. Inhibition of cellular functions of HIV-1 Nef by artificial SH3 domains. Virology 286, 152–159.
Hirano, H., Matsuura, Y., 2011. Sensing actin dynamics: structural basis for G-actin-sensitive nuclear import of MAL. Biochem. Biophys. Res. Commun. 414, 373–378.
Hodel, M.R., Corbett, A.H., Hodel, A.E., 2001. Dissection of a nuclear localization signal. J. Biol. Chem. 276, 1317–1325.
Hosse, R.J., Rothe, A., Power, B.E., 2006. A new generation of protein display scaffolds for molecular recognition. Protein Sci. 15, 14–27.
Huang, X., Poy, F., Zhang, R., Joachimiak, A., Sudol, M., et al., 2000. Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan. Nat. Struct. Biol. 7, 634–638.
Huang, J., Koide, A., Makabe, K., Koide, S., 2008. Design of protein function leaps by directed domain interface evolution. Proc. Natl. Acad. Sci. USA 105, 6578–6583.
Huber, A.H., Weis, W.I., 2001. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. Cell 105, 391–402.
Huber, A.H., Nelson, W.J., Weis, W.I., 1997. Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell 90, 871–882.
Hübner, S., Xiao, C.Y., Jans, D.A., 1997. The protein kinase CK2 site (Ser111/112) enhances recognition of the simian virus 40 large T-antigen nuclear localization sequence by importin. J. Biol. Chem. 272, 17191–17195.
Jackrel, M.E., Valverde, R., Regan, L., 2009. Redesign of a protein-peptide interaction: characterization and applications. Protein Sci. 18, 762–774.
Jackrel, M.E., Cortajarena, A.L., Liu, T.Y., Regan, L., 2010. Screening libraries to identify proteins with desired binding activities using a split-GFP reassembly assay. ACS Chem. Biol. 5, 553–562.
Jaulin-Bastard, F., Arsanto, J.P., Le Bivic, A., Navarro, C., Vely, F., et al., 2002. Interaction between Erbin and a Catenin-related protein in epithelial cells. J. Biol. Chem. 277, 2869–2875.
Jennings, B.H., Pickles, L.M., Wainwright, S.M., Roe, S.M., Pearl, L.H., et al., 2006. Molecular recognition of transcriptional repressor motifs by the WD domain of the Groucho/TLE corepressor. Mol. Cell 22, 645–655.
Jiang, W., Boder, E.T., 2010. High-throughput engineering and analysis of peptide binding to class II MHC. Proc. Natl. Acad. Sci. USA 107, 13258–13263.
Junqueira, D., Cilenti, L., Musumeci, L., Sedivy, J.M., Zervos, A.S., 2003. Random mutagenesis of PDZ(Omi) domain and selection of mutants that specifically bind the Myc proto-oncogene and induce apoptosis. Oncogene 22, 2772–2781.
Kajander, T., Cortajarena, A.L., Main, E.R., Mochrie, S.G., Regan, L., 2005. A new folding paradigm for repeat proteins. J. Am. Chem. Soc. 127, 10188–10190.
Kajander, T., Sachs, J.N., Goldman, A., Regan, L., 2009. Electrostatic interactions of Hsp-organizing protein tetratricopeptide domains with Hsp70 and Hsp90: computational analysis and protein engineering. J. Biol. Chem. 284, 25364–25374.
Kaneko, T., Sidhu, S.S., Li, S.S., 2011. Evolving specificity from variability for protein interaction domains. Trends Biochem. Sci. 36, 183–190.

Kay, B.K., Williamson, M.P., Sudol, M., 2000. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. 14, 231–241.
Killian, B.J., Kravitz, J.Y., Somani, S., Dasgupta, P., Pang, Y.P., et al., 2009. Configurational entropy in protein-peptide binding: computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide. J. Mol. Biol. 389, 315–335.
Kippert, F., Gerloff, D.L., 2009. Highly sensitive detection of individual HEAT and ARM repeats with HHpred and COACH. PLoS One 4, e7148.
Klug, A., 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annu. Rev. Biochem. 79, 213–231.
Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., et al., 2000. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J. Mol. Biol. 296, 57–86.
Kobe, B., 1999. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. Nat. Struct. Biol. 6, 388–397.
Kofler, M.M., Freund, C., 2006. The GYF domain. FEBS J. 273, 245–256.
Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., et al., 2009. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J. Biol. Chem. 284, 478–485.
Kuriyan, J., Cowburn, D., 1997. Modular peptide recognition domains in eukaryotic signaling. Annu. Rev. Biophys. Biomol. Struct. 26, 259–288.
Lam, M.H., Briggs, L.J., Hu, W., Martin, T.J., Gillespie, M.T., et al., 1999. Importin beta recognizes parathyroid hormone-related protein with high affinity and mediates its nuclear import in the absence of importin alpha. J. Biol. Chem. 274, 7391–7398.
Lee, C.H., Leung, B., Lemmon, M.A., Zheng, J., Cowburn, D., et al., 1995. A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. EMBO J. 14, 5006–5015.
Lee, B.J., Cansizoglu, A.E., Suel, K.E., Louis, T.H., Zhang, Z., et al., 2006. Rules for nuclear localization sequence recognition by karyopherin beta 2. Cell 126, 543–558.
Lim, W.A., Richards, F.M., Fox, R.O., 1994. Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. Nature 372, 375–379.
Löfblom, J., Frejd, F.Y., Ståhl, S., 2011. Non-immunoglobulin based protein scaffolds. Curr. Opin. Biotechnol. 22, 843–848.
London, N., Movshovitz-Attias, D., Schueler-Furman, O., 2010. The structural basis of peptide-protein binding strategies. Structure 18, 188–199.
Lott, K., Bhardwaj, A., Mitrousis, G., Pante, N., Cingolani, G., 2010. The importin beta binding domain modulates the avidity of importin beta for the nuclear pore complex. J. Biol. Chem. 285, 13769–13780.
Lu, P.J., Zhou, X.Z., Shen, M., Lu, K.P., 1999. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. Science 283, 1325–1328.
Luginbühl, B., Kanyo, Z., Jones, R.M., Fletterick, R.J., Prusiner, S.B., et al., 2006. Directed evolution of an anti-prion protein scFv fragment to an affinity of 1 pM and its structural interpretation. J. Mol. Biol. 363, 75–97.
MacCallum, R.M., Martin, A.C., Thornton, J.M., 1996. Antibody-antigen interactions: contact analysis and binding site topography. J. Mol. Biol. 262, 732–745.
Macias, M.J., Wiesner, S., Sudol, M., 2002. WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. FEBS Lett. 513, 30–37.
Madhurantakam, C., Varadamsetty, G., Grütter, M.G., Plückthun, A., Mittl, P.R., 2012. Structure-based optimization of designed Armadillo-repeat proteins. Protein Sci. 21, 1015–1028.
Mahajan, A., Yuan, C., Lee, H., Chen, E.S., Wu, P.Y., et al., 2008. Structure and function of the phosphothreonine- specific FHA domain. Sci. Signal. 1, re12.
Main, E.R., Lowe, A.R., Mochrie, S.G., Jackson, S.E., Regan, L., 2005. A recurring theme in protein engineering: the design, stability and folding of repeat proteins. Curr. Opin. Struct. Biol. 15, 464–471.
Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J., Stoddard, B.L., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335, 716–719.
Malabarba, M.G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P.G., et al., 2001. A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. Oncogene 20, 5186–5194.
Marfori, M., Mynott, A., Ellis, J.J., Mehdi, A.M., Saunders, N.F., et al., 2011. Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim. Biophys. Acta 1813, 1562–1577.
Marfori, M., Lonhienne, T.G., Forwood, J.K., Kobe, B., 2012. Structural basis of high affinity nuclear localization signal interactions with Importin alpha. Traffic 13, 532–548.
Matsuura, Y., Stewart, M., 2005. Nup50/Npap60 function in nuclear protein import complex disassembly and importin recycling. EMBO J. 24, 3681–3689.
Mayer, B.J., 2001. SH3 domains: complexity in moderation. J. Cell Sci. 114, 1253–1263.
Mintz, C.S., Crea, R., 2013. Protein Scaffolds. BioProcess International 11, 40–48.
Mondon, P., Dubreuil, O., Bouayadi, K., Kharrat, H., 2008. Human antibody libraries: a race to engineer and explore a larger diversity. Front. Biosci. 13, 1117–1129.
Morgan, C.S., Holton, J.M., Olafson, B.D., Bjorkman, P.J., Mayo, S.L., 1997. Circular dichroism determination of class I MHC-peptide equilibrium dissociation constants. Protein Sci. 6, 1771–1773.
Munson, M., Balasubramanian, S., Fleming, K.G., Nagi, A.D., O'Brien, R., et al., 1996. What makes a protein a protein? Hydrophobic core designs that specify stability and structural properties. Protein Sci. 5, 1584–1593.
Muslin, A.J., Tanner, J.W., Allen, P.M., Shaw, A.S., 1996. Interaction of 14–3-3 with signaling proteins is mediated by the recognition of phosphoserine. Cell 84, 889–897.
Mynott, A.V., Harrop, S.J., Brown, L.J., Breit, S.N., Kobe, B., et al., 2011. Crystal structure of importin-alpha bound to a peptide bearing the nuclear localisation signal from chloride intracellular channel protein 4. FEBS J. 278, 1662–1675.
Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F.B., et al., 2001. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. Nature 414, 514–521.
Neefjes, J., Jongsma, M.L., Paul, P., Bakke, O., 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. 11, 823–836.
Nieto, J.L., Rico, M., Santoro, J., Herranz, J., Bermejo, F.J., 1986. Assignment and conformation of neurotensin in aqueous solution by 1H NMR. Int. J. Pept. Protein Res. 28, 315–323.
Nikkhah, M., Jawad-Alami, Z., Demydchuk, M., Ribbons, D., Paoli, M., 2006. Engineering of beta-propeller protein scaffolds by multiple gene duplication and fusion of an idealized WD repeat. Biomol. Eng. 23, 185–194.
Nourry, C., Grant, S.G., Borg, J.P., 2003. PDZ domain proteins plug and play! Sci. STKE, re7.
Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801.
Obsil, T., Obsilova, V., 2011. Structural basis of 14-3-3 protein functions. Semin. Cell Dev. Biol. 22, 663–672.
Oliver, A.W., Swift, S., Lord, C.J., Ashworth, A., Pearl, L.H., 2009. Structural basis for recruitment of BRCA2 by PALB2. EMBO Rep. 10, 990–996.
Orlicky, S., Tang, X., Willems, A., Tyers, M., Sicheri, F., 2003. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. Cell 112, 243–256.
Panni, S., Dente, L., Cesareni, G., 2002. In vitro evolution of recognition specificity mediated by SH3 domains reveals target recognition rules. J. Biol. Chem. 277, 21666–21674.
Parmeggiani, F., Pellarin, R., Larsen, A.P., Varadamsetty, G., Stumpp, M.T., et al., 2008. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J. Mol. Biol. 376, 1282–1304.
Pawson, T., Nash, P., 2003. Assembly of cell regulatory systems through protein interaction domains. Science 300, 445–452.
Pawson, T., Scott, J.D., 1997. Signaling through scaffold, anchoring, and adaptor proteins. Science 278, 2075–2080.
Pennell, S., Westcott, S., Ortiz-Lombardia, M., Patel, D., Li, J., et al., 2010. Structural and functional analysis of phosphothreonine-dependent FHA domain interactions. Structure 18, 1587–1595.
Pershad, K., Wypisniak, K., Kay, B.K., 2012. Directed evolution of the forkhead-associated domain to generate anti-phosphospecific reagents by phage display. J. Mol. Biol. 424, 88–103.
Peterson, F.C., Volkman, B.F., 2009. Diversity of polyproline recognition by EVH1 domains. Front Biosci. 14, 833–846.
Petsalaki, E., Stark, A., Garcia-Urdiales, E., Russell, R.B., 2009. Accurate prediction of peptide binding sites on protein surfaces. PLoS Comput. Biol. 5, e1000335.
Plückthun, A., Schaffitzel, C., Hanes, J., Jermutus, L., 2000. In vitro selection and evolution of proteins. Adv. Protein Chem. 55, 367–403.
Pope, M.E., Soste, M.V., Eyford, B.A., Anderson, N.L., Pearson, T.W., 2009. Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. J. Immunol. Meth. 341, 86–96.
Posern, G., Zheng, J., Knudsen, B.S., Kardinal, C., Müller, K.B., et al., 1998. Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G. Oncogene 16, 1903–1912.
Poy, F., Lepourcelet, M., Shivdasani, R.A., Eck, M.J., 2001. Structure of a human Tcf4-beta-catenin complex. Nat. Struct. Biol. 8, 1053–1057.
Prehoda, K.E., Lee, D.J., Lim, W.A., 1999. Structure of the enabled/VASP homology 1 domain-peptide complex: a key component in the spatial control of actin assembly. Cell 97, 471–480.
Pumroy, R.A., Nardozzi, J.D., Hart, D.J., Root, M.J., Cingolani, G., 2012. Nucleoporin Nup50 stabilizes closed conformation of armadillo repeat 10 in importin alpha5. J. Biol. Chem. 287, 2022–2031.
Reimand, J., Hui, S., Jain, S., Law, B., Bader, G.D., 2012. Domain-mediated protein interaction prediction: From genome to network. FEBS Lett. 586, 2751–2763.
Reina, J., Lacroix, E., Hobson, S.D., Fernandez-Ballester, G., Rybin, V., et al., 2002. Computer-aided design of a PDZ domain to recognize new target sequences. Nat. Struct. Biol. 9, 621–627.
Riggleman, B., Wieschaus, E., Schedl, P., 1989. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes Dev. 3, 96–113.
Rini, J.M., Schulze-Gahmen, U., Wilson, I.A., 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 255, 959–965.
Rittinger, K., Budman, J., Xu, J., Volinia, S., Cantley, L.C., et al., 1999. Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. Mol. Cell 4, 153–166.
Rudolph, M.G., Speir, J.A., Brunmark, A., Mattsson, N., Jackson, M.R., et al., 2001. The crystal structures of K(bm1) and K(bm8) reveal that subtle changes in the peptide environment impact thermostability and alloreactivity. Immunity 14, 231–242.

Rudolph, M.G., Heissmann, R., Wittmann, J.G., Klostermeier, D., 2006. Crystal structure and nucleotide binding of the Thermus thermophilus RNA helicase Hera N-terminal domain. J. Mol. Biol. 361, 731–743.
Sadegh-Nasseri, S., Stern, L.J., Wiley, D.C., Germain, R.N., 1994. MHC class II function preserved by low-affinity peptide interactions preceding stable binding. Nature 370, 647–650.
Sampietro, J., Dahlberg, C.L., Cho, U.S., Hinds, T.R., Kimelman, D., et al., 2006. Crystal structure of a beta-catenin/BCL9/Tcf4 complex. Mol. Cell 24, 293–300.
Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., et al., 2000. Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell 101, 199–210.
Schlessinger, J., Lemmon, M.A., 2003. SH2 and PTB domains in tyrosine kinase signaling. Sci. STKE 2, re12.
Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., et al., 1999. Mutagenesis and selection of PDZ domains that bind new protein targets. Nat. Biotechnol. 17, 170–175.
Shapiro, L., Weis, W.I., 2009. Structure and biochemistry of cadherins and catenins. Cold Spring Harb. Perspect. Biol. 1, a003053.
Smith, T.F., Gaitatzes, C., Saxena, K., Neer, E.J., 1999. The WD repeat: a common architecture for diverse functions. Trends Biochem. Sci. 24, 181–185.
Song, J.J., Garlick, J.D., Kingston, R.E., 2008. Structural basis of histone H4 recognition by p55. Genes Dev. 22, 1313–1318.
Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., et al., 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 275, 73–77.
Stanfield, R.L., Wilson, I.A., 1995. Protein-peptide interactions. Curr. Opin. Struct. Biol. 5, 103–113.
Stiffler, M.A., Chen, J.R., Grantcharova, V.P., Lei, Y., Fuchs, D., et al., 2007. PDZ domain binding selectivity is optimized across the mouse proteome. Science 317, 364–369.
Stirnimann, C.U., Petsalaki, E., Russell, R.B., Müller, C.W., 2010. WD40 proteins propel cellular networks. Trends Biochem. Sci. 35, 565–574.
Stoevesandt, O., Taussig, M.J., 2012. Affinity proteomics: the role of specific binding reagents in human proteome analysis. Exp. Rev. Proteomics 9, 401–414.
Sun, J., Weis, W.I., 2011. Biochemical and structural characterization of beta-catenin interactions with nonphosphorylated and CK2-phosphorylated Lef-1. J. Mol. Biol. 405, 519–530.
Sundberg, E.J., 2009. Structural basis of antibody-antigen interactions. Meth. Mol. Biol. 524, 23–36.
Takeda, A.A., de Barros, A.C., Chang, C.W., Kobe, B., Fontes, M.R., 2011. Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80. J. Mol. Biol. 412, 226–234.
Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., et al., 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat. Struct. Mol. Biol. 14, 229–233.
Tewari, R., Bailes, E., Bunting, K.A., Coates, J.C., 2010. Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol. 20, 470–481.
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, 2631–2637.
Tonikian, R., Zhang, Y., Sazinsky, S.L., Currell, B., Yeh, J.H., et al., 2008. A specificity map for the PDZ domain family. PLoS Biol. 6, e239.
Tripp, K.W., Barrick, D., 2007. Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats. J. Mol. Biol. 365, 1187–1200.
Uhlik, M.T., Temple, B., Bencharit, S., Kimple, A.J., Siderovski, D.P., et al., 2005. Structural and evolutionary division of phosphotyrosine binding (PTB) domains. J. Mol. Biol. 345, 1–20.
Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., et al., 2010. Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats. J. Mol. Biol. 404, 307–327.
Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F., Plückthun, A., 2012. Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. J. Mol. Biol. 424, 68–87.
Vyas, J.M., Van der Veen, A.G., Ploegh, H.L., 2008. The known unknowns of antigen processing and presentation. Nat. Rev. Immunol. 8, 607–618.
Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D., Kuriyan, J., 1993. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. Cell 72, 779–790.
Wetzel, S.K., Settanni, G., Kenig, M., Binz, H.K., Plückthun, A., 2008. Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. J. Mol. Biol. 376, 241–257.
Wilson, I.A., Ghiara, J.B., Stanfield, R.L., 1994. Structure of anti-peptide antibody complexes. Res. Immunol. 145, 73–78.
Wolf, G., Trub, T., Ottinger, E., Groninga, L., Lynch, A., et al., 1995. PTB domains of IRS-1 and Shc have distinct but overlapping binding specificities. J. Biol. Chem. 270, 27407–27410.
Xing, Y., Clements, W.K., Le Trong, I., Hinds, T.R., Stenkamp, R., et al., 2004. Crystal structure of a beta-catenin/APC complex reveals a critical role for APC phosphorylation in APC function. Mol. Cell 15, 523–533.
Xu, W., Kimelman, D., 2007. Mechanistic insights from structural studies of beta-catenin and its binding partners. J. Cell Sci. 120, 3337–3344.
Xu, C., Jin, J., Bian, C., Lam, R., Tian, R., et al., 2012. Sequence- specific recognition of a PxLPxI/L motif by an ankyrin repeat tumbler lock. Sci. Signal. 5, ra39.
Yaffe, M.B., 2002. Phosphotyrosine-binding domains in signal transduction. Nat. Rev. Mol. Cell Biol. 3, 177–186.
Yaffe, M.B., Smerdon, S.J., 2004. The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function. Annu. Rev. Biophys. Biomol. Struct. 33, 225–244.
Yaffe, M.B., Rittinger, K., Volinia, S., Caron, P.R., Aitken, A., et al., 1997. The structural basis for 14–3-3:phosphopeptide binding specificity. Cell 91, 961–971.
Yaneva, R., Schneeweiss, C., Zacharias, M., Springer, S., 2010. Peptide binding to MHC class I and II proteins: new avenues from new methods. Mol. Immunol. 47, 649–657.
Yang, S.N., Takeda, A.A., Fontes, M.R., Harris, J.M., Jans, D.A., et al., 2010. Probing the specificity of binding to the major nuclear localization sequence-binding site of importin-alpha using oriented peptide library screening. J. Biol. Chem. 285, 19935–19946.
Zahnd, C., Spinelli, S., Luginbühl, B., Amstutz, P., Cambillau, C., et al., 2004. Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. J. Biol. Chem. 279, 18870–18877.
Zarrinpar, A., Bhattacharyya, R.P., Lim, W.A., 2003. The structure and function of proline recognition domains. Sci. STKE 2, re8.
Zeytuni, N., Zarivach, R., 2012. Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module. Structure 20, 397–405.
Zhou, M.M., Huang, B., Olejniczak, E.T., Meadows, R.P., Shuker, S.B., et al., 1996. Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. Nat. Struct. Biol. 3, 388–393.