Curr. Issues Mol. Biol. (2003) 5: 129-146.

Exploring Biochemistry and Cellular Biology with Protein Libraries

Juan E. Diaz[2], Brett E. Howard[1], Marc S. Neubauer[1], Allison Olszewski[1] and Gregory A. Weiss[1,2]*

[1]Department of Chemistry and [2]Department of Molecular Biology and Biochemistry, University of California, Irvine
Irvine, CA 92697-2025, USA

Abstract

Polypeptide libraries cast a broad net for defining enzyme and binding protein specificities. In addition to uncovering rules for molecular recognition, the binding preferences and functional group tolerances from such libraries can reveal mechanisms underlying biochemical and cellular processes. Ligands obtained from protein libraries can also provide pharmaceutical lead compounds and even reagents to further explore cell biology. Here, we review selected recent examples of protein libraries demonstrating these principles. In particular, we focus on combinatorial libraries composed of randomized peptides or variations of a single protein. The characteristics of various techniques for library constructions and screening are also briefly surveyed.

Introduction

Essentially all processes in biology involve some aspect of molecular recognition – one molecule (often a protein) specifically binding to another. The rules for molecular recognition remain unpredictable. Current technologies, for example, usually cannot accurately predict binding partners from sequence information. Protein libraries, diverse collections of polypeptides, offer a powerful method to fill-in the gaps in our predictive abilities, such as identification of binding partners, elucidation of key contributors to recognition or obtaining new binding interactions. This review focuses on the first two aspects of protein libraries – identification of binding partners and dissection of key residues at receptor-ligand interfaces. In addition, we focus upon recent examples that illustrate the power of protein libraries to elucidate principles of molecular recognition and enzymatic catalysis. Due to space limitations, some superb contributions to the field are omitted, and we apologize to those researchers.

The quintessential library approach to harnessing molecular recognition can be found in the immune system, which leverages large collections of circulating antibodies for repelling invaders. Such natural libraries illustrate two principles considered by every practitioner of protein library techniques. First, antibodies are quite large (150+ kD) and cannot be fully randomized at every position. Thus, the immune system focuses its diversity generation apparatus on genes encoding flexible antigen recognition loops. Site-directed mutagenesis offers researchers an analogous process. Second, different antibody formats, such as the decavalent IgM or bivalent IgG, offer advantages and disadvantages that effect the outcome of the attempt to identify binding partners. For example, the decavalent format of IgM can provide a velcro or avidity effect to boost the apparent binding affinity of a weak initial binding partner; however, the absolute affinity of the binding partner obtained, when removed from the decavalent context of an IgM scaffold, can be quite weak. Since consideration of library format is necessary for successful protein library experiments, we will first review the most common formats used, before surveying applications of these formats in dissecting biochemistry and cell biology.

Methods for polypeptide libraries

Using combinatorial approaches, one can readily generate vast polypeptide libraries with sizes around a trillion (10^{12}) different molecules (Sidhu and Weiss, 2002). As reviewed here, such libraries can identify sidechains critical to protein-protein interactions and delineate protein domain and enzyme specificity.

Although many combinatorial protein library formats have been employed, libraries are constructed either by chemical or biological synthesis. Among the many examples of biologically derived library methods, one of the most common is phage display whereby members of a peptide library are fused to ("displayed" on) bacteriophage coat proteins. Other biologically based methods also rely on display of fusion-peptides, such as yeast surface display (Kieke *et al.*, 1999; Yeung and Wittrup, 2002), *E. coli* surface display (Daugherty *et al.*, 1999), ribosome display (Hanes and Pluckthun, 1997) and mRNA display (Wilson *et al.*, 2001) (Table 1). The key feature of these biological libraries is a direct link between the phenotype (the displayed peptide) and the genotype (DNA encoding the peptide) that allows for selection and amplification of desired peptides. Synthetic peptide libraries on the other hand do not require cellular components for construction, but instead rely on chemical techniques for the synthesis of peptides, through SPOT, parallel and split-pool synthesis.

Phage Display

Since its introduction in the mid-1980's by George Smith (Smith, 1985), phage display has developed into a powerful tool for the display and screening of large combinatorial peptide libraries with diversities up to 10^{12} members in size (Sidhu *et al.*, 2000). Most peptide libraries are constructed using the well-established non-lytic M13 filamentous phage as a scaffold. Other display scaffolds such as the lytic phages λ, T4, and T7 have also been used. Many methods for phage display rely on the use of phagemid vectors to

*For correspondence. Email gweiss@uci.edu.

© 2003 Caister Academic Press

Table 1. Comparision of selected methods for biological synthesis of polypeptide libraries.

Library Method	Advantages	Caveats	Potential Diversity
Phage display	• Direct link between phenotype and genotype • Rapid amplification of displayed peptide • *In vitro* selection	• Transformation step required • Typically only naturally occurring amino acids incorporated • Selections usually limited to binding events	$\leq 10^{12}$
Yeast two-hybrid	• Identification of protein-protein interactions • Multiple reporter genes available for selection	• *In vivo* selection • Transformation step required • No direct control over binding conditions • Need to screen for false positives • Proteins must be transported to nucleus	$10^5 - 10^7$
Ribosome display	• No transformation step required • *In vitro* transcription and translation	• Ternary complex stable only at 4 °C • Possibility for unwanted truncation products • Dependent upon RNA stability	$> 10^{13}$
Messenger RNA display	• *In vitro* transcription and translation • Possible to incorporate unnatural amino acids • Covalent linkage between mRNA and displayed polypeptide • Improved display complex stability	• RNA stability • Enzymatic ligation of mRNA to DNA linker required • Puromycin incorporation requires additional step	$> 10^{13}$

encode the protein of interest fused to a viral coat protein displayed on the surface of the phage particle. The fusion protein is packaged into phage particles by a co-infected helper phage that provides all of the components necessary for viral assembly. As described below, an equivalent strategy of mutating one copy of a coat protein gene fused to the protein of interest can also feature rescue and assembly of the virions by a special strain of *E. coli;* however, the principle of altering just a few copies of the viral coat protein provides a more stable platform for library construction and screening that is especially important for larger polypeptides.

Filamentous M13 phage display

M13, in addition to the closely related fd and f1 bacteriophage, infects male *E. coli* via the reproductive bacterial F pilus. Due to its slender morphology (1μm in length by 8nm in diameter), M13 is classified as a filamentous bacteriophage. A single-stranded closed circular DNA genome is encapsulated within a viral coat composed of five viral coat proteins (P3, P6, P7, P8, P9). At one end, five copies each of the minor viral coat proteins P7 and P9 cap the virus; the "head" consists of five copies each of minor coat proteins P3 and P6. The major viral coat protein P8 covers the entire length of the virus and about 2700 copies cover the viral suface (Feng *et al.,* 1999). Because P8 comprises most of the viral coat, peptides fused to P8 are often displayed in a polyvalent format; fusion to the minor coat proteins (P3, P6, P7, P9) can yield a low-valent or monovalent display level. Polyvalent protein display can identify low-affinity interactions, through an avidity or "velcro" effect. By contrast, monovalent display can identify high-affinity interactions. Polypeptides can either be displayed as attachments to the amino-terminus of P3, P7, P8 and P9, or as a carboxyl-terminal fusion to P6 and P8 (Figure 1a).

T7 phage display

Bacteriophage T7, a lytic double-stranded DNA virus, infects *E. coli* and has been employed as a scaffold for phage display. T7 DNA is packaged into an icosahedral particle (~55nm in diameter) that is composed entirely of the gene 10 viral capsid protein arranged in hexamers (icosahedral faces) and pentamers (icosahedral vertices). The gene 10 capsid protein is generated as a full-length 10B protein (397 residues) and a truncated 10A form (344 residues).

Fusion proteins are displayed as C-terminal fusions to the 10B major coat protein, with peptides (<50 residues) being displayed in high copy numbers (415 copies per phage) and larger polypeptides (~1200 residues) displayed in low copy numbers, usually between 0.1-1 copies per phage (Rosenberg *et al.*, 1996). In theory, a greater variety of proteins and peptides can be displayed on the surface of the T7 phage, due to the lytic life cycle of the virus whereby viral particles assemble within the cytoplasm of the cell and are released to the external environment upon lysis. By contrast, polypeptides displayed on M13 phage must tolerate extrusion through the bacterial inner membrane during the viral assembly, which could limit the types of proteins that can be fused to and displayed on the virus. In a recent example, RNA-binding proteins have been successfully screened and selected using this T7 phage display approach (Danner and Belasco, 2001).

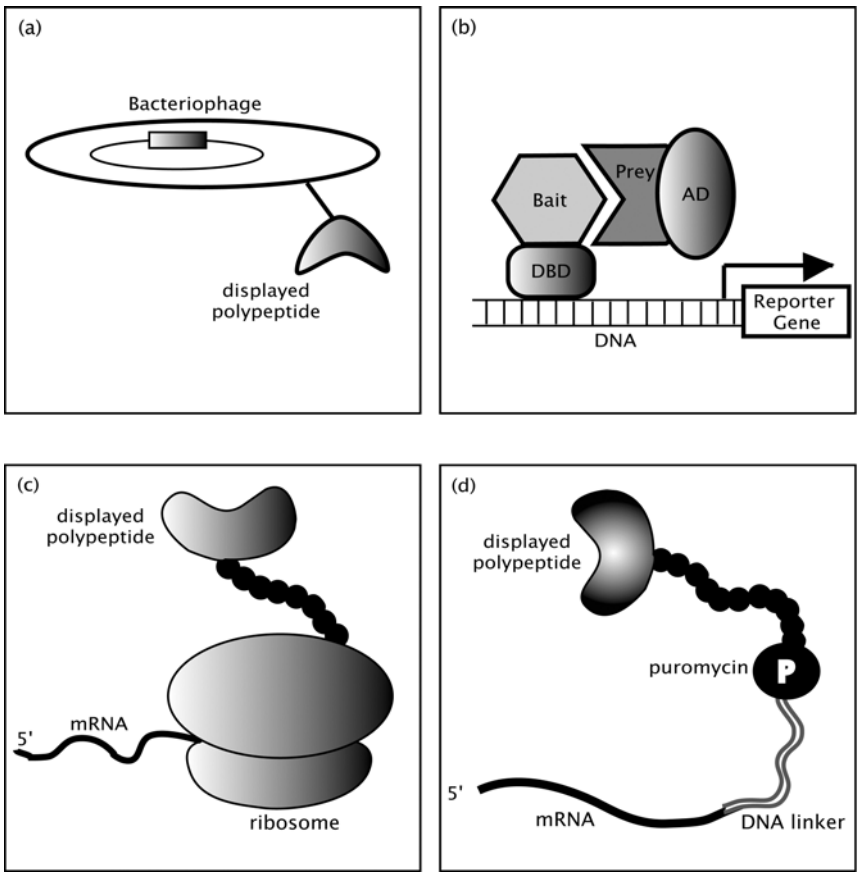


Figure 1. Selected methods for biological synthesis of polypeptide libraries. (a) M13 phage display. Polypeptide library members are fused to a coat protein, encoded by a phagemid vector. Fusion proteins and the phagemid DNA are packaged into phage particles by a helper phage that provides all of the components necessary for viral assembly. Selections often feature binding to solid target ligands fixed to a solid support. (b) Yeast two-hybrid. Fusion peptides are expressed within yeast cells. Library proteins (prey) are fused to a transcription activation domain (AD). The target protein under study is fused to a DNA-binding domain (DBD) and used as "bait". Binding of the target protein to the library member brings the DBD and AD domains into close proximity, which activates transcription of a reporter gene (Figure adapted from Uetz and Hughes, 2000). (c) Ribosome display. Members of the peptide library are fused directly to a C-terminal spacer, which lacks an encoded stop codon. After *in vitro* transcription and translation, the ribosomes stalls to result in formation of a stable ternary complex (mRNA, ribosome, peptide). The C-terminal spacer allows the displayed protein to exit the ribosome, fold into its correct conformation and remain associated with its encoding mRNA. Ternary complexes, which are stable at 4 °C can be directly used for binding-based selections. (d) mRNA display. mRNA encoding the displayed peptide is fused to a DNA linker, which is terminated with a covalently attached puromycin. After *in vitro* translation, the ribosome stalls at the RNA-DNA junction, allowing the puromycin to enter the ribosome and form a stable covalent bond with the nascent protein. The resulting mRNA-peptide fusion can be utilized for selection experiments.

Yeast two-hybrid

Unlike phage display, which physically links a peptide to the surface of phage for *in vitro* selection and screening, yeast two-hybrid libraries remain entirely *in vivo* (usually *Saccharomyces cerevisiae* cells). This method of library screening is best suited for studying protein-protein interactions (reviewed by Uetz and Hughes, 2000; Toby and Golemis, 2001). In the two-hybrid system (Figure 1b), the protein under study is fused to a DNA binding domain (DBD) and used as "bait" to identify novel binding partners. Library proteins (e.g., cDNA products) are fused to a transcription activation domain (AD) to generate the "prey" and then coexpressed with the "bait." Neither bait, nor prey alone, should activate the transcription of the reporter gene. When binding of the target protein to the library member brings the DBD and AD domains into close proximity, transcription occurs. Commonly used reporter genes include *lacZ*, *HIS3*, *LEU2* (Toby and Golemis, 2001) and *ADE2*. Transcription of *lacZ* is detectable by addition of X-Gal substrate to produce a color change, while the *HIS3* and *LEU2* genes allow for growth of yeast on media deficient in histidine or leucine, respectively. *ADE2* transcription can produce a color change, due to an enzymatic mutation in the adenine biosynthetic pathway. Variants of this technique include the three-hybrid (three binding partners), the reverse two-hybrid (association results in death), the reverse three-hybrid system and others.

Ribosome Display

The use of *in vitro* techniques for peptide library synthesis can produce libraries with diversities exceeding 10^{13} unique members (reviewed by Roberts, 1999). The large diversity is a result of circumventing the cellular limitations of DNA

transformation required for construction of phage, yeast, and *E. coli* display libraries. Ribosome display relies on a direct linkage between the peptide being displayed and the mRNA that encodes it via a stable ternary complex (Figure 1c). This technique has been used for investigating binding affinities and catalytic activity of enzymes (Amstutz *et al.*, 2002). In this technique, *in vitro* transcription and translation are initiated upstream of the library gene by a T7 promoter and ribosome binding sequence, respectively. The lack of a stop codon terminating the construct prevents dissociation of the ribosome from the mRNA template, allowing the newly synthesized protein to fold and remain tethered to the ribosome via a C-terminal linker. At 4°C, this ternary complex of ribosome, mRNA and displayed protein is surprisingly stable. Selections can be performed on the displayed polypeptide; after which, the encoding mRNA can be reverse-transcribed for amplification, further rounds of selection, and eventual sub-cloning into a vector for sequencing.

Ribosome-inactivation display system (RIDS)
A recently described variation to ribosome display features a room temperature stable ternary complex of the displayed polypeptide, encoding mRNA and ribosome by use of the ricin A chain (RTA), a eukaryotic plant toxin that inhibits protein synthesis (Zhou *et al.*, 2002). The ricin A chain hydrolyzes a N-glycosidic bond within the large ribosomal subunit; glycoside hydrolysis prevents dissociation of the nascent protein and mRNA from the ribosome. The ribosome-inactivation display construct consists of a single ORF with the displayed polypeptide gene fused to the 5'end of the RTA gene by a linker. A C-terminal spacer following the RTA gene allows the ricin A to exit the ribosome, fold properly, bind and inactivate the ribosome.

mRNA Display

Messenger RNA display (Nemoto *et al.*, 1997; Roberts and Szotak, 1997), an improved variation of ribosome display, is an *in vitro* selection technique capable of generating polypeptide diversities exceeding 10^{13} (Cho *et al.*, 2000; Barrick *et al.*, 2001; Li *et al.*, 2002). A more stable linkage between the displayed polypeptide to its encoding mRNA is accomplished by the introduction of a puromycin antibiotic to the tail of the mRNA (Figure 1d). Originally, puromycin molecules were enzymatically ligated to the 3' ends of mRNA (encoding library peptides) by a DNA linker, but now photo-crosslinking of puromycin to mRNA using a psoralen-based technique (Kurz *et al.*, 2000) is possible. Also, replacement of mRNA molecules with double-stranded cDNA (Kurz *et al.*, 2001) significantly increases the stability of fusion proteins to their transcripts. Ribosomes begin translating the mRNA transcripts and stall upon reaching the RNA-DNA junction, which allows the puromycin to enter the ribosome (peptidyltransferase site) to form a stable covalent amide bond to the nascent protein. The resulting mRNA-peptide fusion can then be utilized in selection experiments.

SPOT-synthesis

SPOT-synthesis is an efficient method for the synthesis and screening of potentially thousands of peptides and small molecules in a spatially addressable array. This technique features chemical synthesis at defined "spots" on a solid support, such as cellulose (e.g., filter paper). Polypropylene membranes have also been used to increase the chemical and mechanical stability between the attached molecule and the membrane (Scharn *et al.*, 2000). For example, for SPOT synthesis of peptide libraries, the activated amino acids are dotted at specific positions onto a functionalized membrane forming a "spot" within which miniscale coupling reactions can proceed. Conventional solid phase peptide synthesis proceeds with each additional amino acids added directly to the defined spots. Linkers can also be included on the functionalized membrane to allow cleavage and recovery of the synthesized product (Wenschuh *et al.*, 2000; Scharn *et al.*, 2000). This method has been employed for a number of applications ranging from the identification of protein-protein interactions, development of protein mimics (Bracci *et al.*, 2001), optimization of peptide spotting densities, improved reproducibility, increased signal intensities (Kramer *et al.,* 1999) and enzyme-substrate recognition (Reineke *et al.*, 2001). Chemical synthesis of combinatorial peptide libraries is also used in examples cited throughout this review to produce solution phase peptide libraries (often by split-pool combinatorial methods), which are sequenced by conventional peptide sequencing methods.

Protein Library Techniques for Mapping Molecular Recognition

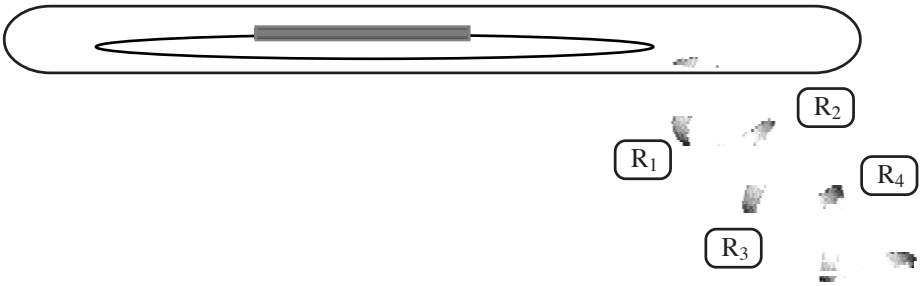
Single point, site-directed mutagenesis is a powerful way to identify key residues contributing energy to receptor-ligand interactions. However, the techniques for site-directed mutagenesis, followed by protein expression and purification can be laborious. Other methods, such as alanine and homolog shotgun scanning can turbocharge determination of key residues at the receptor-ligand interface. These methods query simultaneously multiple amino acids for their contributions to receptor binding. High throughput quantification of energetic contributions by multiple sidechain residues can be used to identify residues critical to molecular recognition.

Before library-based methods were used for mapping receptor-ligand interactions, single point mutagenesis elucidated many key principles. For example, analysis of human growth hormone (hGH) illustrated the key role played by a small number of hGH residues interacting with the hGH binding protein extracellular domain (hGHbp) (Cunningham and Wells, 1993). By mutating surface residues of hGH to alanine, sidechain atoms past the β-carbon were truncated. Comparison of binding by wild-type hGH with the hGH alanine mutant quantified contributions to binding by atoms past the sidechain β-carbon. A small cluster of hGH residues were found to contribute most of the binding energy to the interaction with hGHbp; these residues were termed the "functional epitope". While the structural epitope consists of the 19 hGH residues buried

upon binding to hGHbp, only the functional epitope residues contribute binding energy to the hGH-hGHbp interaction. Similar studies with hGHbp identified a complementary functional epitope on the receptor (Clackson and Wells, 1995). To further understand the composition of the hGH functional epitope, approximately one million hGH mutants were synthesized by phage display (Lowman *et al.*, 1991). Following selections and screens, a consensus sequence for hGH mutants binding to hGHbp was obtained. The consensus residues identified as important in this protein library experiment were similar to the residues highlighted by alanine mutagenesis analysis.

Recent studies applying alanine shotgun scanning have demonstrated more rapid ways to analyze functional epitopes (Figure 2). In the first example of alanine shotgun scanning, hGH was displayed on phage, and a library of hGH variants constructed such that the 19 residues of the structural epitope were substituted with either alanine, the wild-type residue, or, in some cases, up to two other amino acids. By selecting and screening for hGHbp binding to the phage-displayed hGH variants, the contributions to binding by the 19 sidechains were analyzed simultaneously. Selectants from the hGH shotgun scanning were analyzed to determine the distribution of wild-type and alanine in the mutated positions. When shotgun scanning, selection for a high percentage of a wild-type sidechain in a particular position indicates a key contribution to protein function by that sidechain; conversely, tolerance for alanine demonstrates little contribution to protein function by atoms past the β-carbon in that position. The same functional epitope residues determined previously by conventional and alanine scanning were highlighted by hGH shotgun scanning, which validates the technique (Weiss *et al.*, 2000).

Shotgun alanine scanning has also been used to map protein-small molecule interactions. The streptavidin-biotin interaction has femtomolar dissociation constant and is a model for high affinity receptor-ligand interactions (Green, 1990). Streptavidin residues critical for binding to biotin were analyzed by phage-displayed alanine shotgun scanning. Residues identified by shotgun scanning extended understanding of the streptavidin-biotin interaction to include residues not in direct contact with biotin (Avrantinis *et al.*, 2002) (Figure 3). Phage display with shotgun alanine scanning and protein structure analysis has also been used to determine key residues of the antigen binding site of the Fab2C4 antibody, which binds to the ErbB2 oncogene product extracellular domain (Vajdos *et al.*, 2002). This study also introduced homolog shotgun scanning, which applied libraries of Fab2C4 substituted with wild-type or homologous residues in specific positions (Figure 2). Homologous residues have similar charge and structure to the wild-type residue (e.g., aspartic acid and glutamic acid), and identify subtle features of sidechain geometries critical to binding. Homolog shotgun scanning identified a Fab2C4 functional epitope that contributes specificity to the interaction with antigen, but the alanine shotgun scanning identified a larger functional epitope encompassing the homolog shotgun scanning epitope. Sidechains with indirect contributions to the binding interaction, as shown by Fab-antigen co-crystal structure determination, were identified through both homolog and alanine shotgun scanning experiments (Vajdos *et al.*, 2002).



Position	Alanine	Homolog
R_1	A/Y/D/S	Y/F
R_2	A/V	V/I
R_3	A/E	E/D
R_4	A/T	T/S

Figure 2. Alanine and homolog shotgun scanning. Limited mutagenesis with the substitutions shown is used to vary specific positions of a phage-displayed protein library. Following selection for binding to the cognate receptor, enrichment for wild-type is observed in key positions (alanine shotgun scanning). Homolog shotgun scanning is used to determine structure activity relationship information and/or optimize binding affinity.

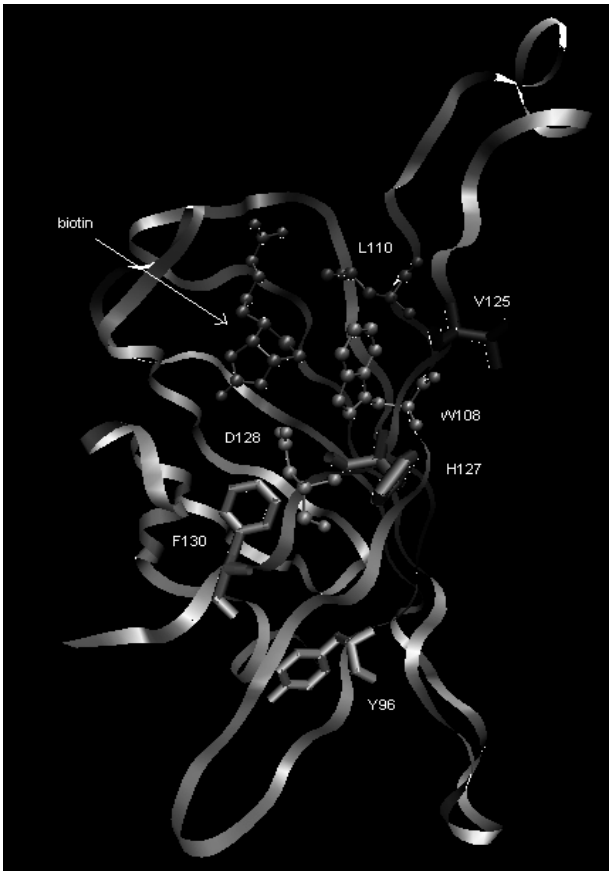


Figure 3. Shotgun scanning the streptavidin-biotin interaction. Biotin and residues highlighted by shotgun scanning in direct interaction are represented as ball-and-stick sidechains, and residues which contribute indirectly to the interaction are shown as tubes (Avrantinis *et al.*, 2002). Figure produced using Visual Molecular Dynamics. (PDB accession code: 1STP; Weber *et al.*, 1989).

In addition to elucidating the role of sidechains buried at receptor-ligand interfaces, protein libraries can identify residues critical for structure formation. For example, residues of a minimal binding protein motif have also been analyzed for their effect on binding affinity (Li *et al.*, 2002). A turn between the C and D helices (CD turn) of interleukin-5 (IL-5) was previously found to be integral for IL-5 receptor binding, and a peptide library based on this turn sequence within a coiled coil stem loop miniprotein was displayed on phage. Selection for binding to IL-5 receptor α-chain resulted in sequences with sidechain charge patterns analogous to the endogenous turn sequence.

Shotgun scanning has also been used to systematically dissect the contribution to protein functionality by every residue of the M13 major coat protein (P8) (Roth *et al.*, 2002). Three functional epitopes were identified, through a selection for incorporation of mutant P8 into a phage composed of wild-type coat proteins provided by trans infection with helper phage. First, a basic patch near the P8 C-terminus most likely interacts with the negatively charged DNA running through the core of the virus. Two additional hydrophobic functional epitopes near the N- and C-termini of P8 were identified that could lock into each other from neighboring P8 molecules, providing a tough viral coat held together by spot wields. This experiment illustrates the potential for powerful library techniques to comprehensively address questions that would be perhaps too daunting for conventional single point methods to tackle, such as quantifying relative contributions by every sidechain in a protein to functionality.

Protein domains and ligand specificity

The techniques for mapping receptor-ligand interactions, described in the previous section, have found perhaps their most important application in mapping the specificities of protein domains, which can fold independently of a longer polypeptide chain. Such domains (or truncated, sub-domain fragments) can often recognize short peptide sequences of three to nine amino acids in length. Many protein-protein interactions in eukaryotic cells require such interaction modules and specific binding to cognate ligands. Phage display and other protein library techniques have been used to identify peptide-binding motifs for various protein domains. Often peptide ligands isolated from *in vitro* selection of these libraries have been used to identify peptides that are identical or similar in sequence to the *in vivo* interacting protein, exemplifying *in vitro/ in vivo* "convergent evolution" (Kay, *et al.,* 2000). This section focuses on peptide ligand specificity determined by the use of protein library methods for six well-characterized protein domains. Polypeptide library experiments work especially well with protein domains that bind peptides in extended conformations (i.e., ligands not dependent on a larger protein to organize a particular conformation).

EH Domains

The Eps homology (EH) domain (approximately 100 amino acids in length) mediates protein-protein interactions that coordinate endocytosis, actin remodeling, and intracellular transduction of signals. The EH domain was first identified as a motif in three copies of the N-terminus of the epidermal growth factor receptor substrate Eps15 and the related protein Eps15R (reviewed by Di Fiore *et al.*, 1997). All EH domains consist of two helix-loop-helix motifs termed EF hands, connected by a short anti-parallel β-sheet. Phage-displayed libraries (Paoluzi *et al.,* 1998) and filter binding assays (Fazioli *et al.,* 1993) were used to determine the ligand specificities for several EH domains. Results from phage display experiments demonstrate that the majority of EH domains bind to peptides containing the amino acid sequence NPF. Three classes of binding peptides have been identified (Table 2) (Paoluzi *et al.*, 1998). Some EH domains (i.e Eps 15) exhibited low affinity for NPF peptides, but higher affinity for peptides with the sequence FW or WW. In addition, a cDNA product expression library on the surface of lambda phage was screened with the same EH domains and resulted in several novel proteins with multiple NPF motifs (Salcini *et al.*, 1997). Further experiments have shown that these proteins are cellular ligands of Eps1 and intersectin (reviewed by Kay *et al.*, 2000). These results demonstrate the power of polypeptide libraries to identify peptides that bind to proteins with similar affinity, specificity and primary sequence as the natural interacting proteins.

Table 2. Peptide binding motifs for EH, PDZ, PTB, SH2, SH3, and WW domains*.

Class	Class consensus	Protein	Ref
EH Domain			
I	NPF	Eps15	(Paoluzi, Castagnoli *et al.* 1998)
		Eps15r	
II	FW, WW, SWG	YBL47c	(Paoluzi, Castagnoli *et al.* 1998)
		Eps15-EH3	
III	H(S/T)F	End3p	(Paoluzi, Castagnoli *et al.* 1998)
PDZ Domain			
I	(S/T)XV	PSD-95	(Songyang, Fanning *et al.* 1997)
		Dlg	
II	#/@X#	LIN-2	(Songyang, Fanning *et al.* 1997)
		p55	
III	(E/D)XW(C/S)	Mint1-1	(Maximov, Sudhof *et al.* 1999)
IV	C-terminal acidic resid.	INADL	(Vaccaro, Brannetti *et al.* 2001)
PTB Domain			
I	NPXpY	Shc	(Kavanaugh and Williams 1994)
		IRS-1	(Gustafson, He *et al.* 1995)
II	NPXY	X11	(Borg 1996)
		Fe65	(Borg 1996)
III	YIGPY	dNUMB	(Li, Zong *et al.* 1997)
SH2 Domain			
I	pY%%I/P	Src	(Songyang, Shoelson *et al.* 1993)
		Fyn	(Songyang, Shoelson *et al.* 1993)
II	pYMEP	Vav	(Songyang 1994)
III	PY#X#	SHC	(Songyang, Shoelson *et al.* 1993)
		HCP	(Songyang, Shoelson *et al.* 1993)
IV	No peptides selected	SHP-2-C	(Songyang, Shoelson *et al.* 1993)
		113TF	(Songyang, Shoelson *et al.* 1993)
SH3 domain			
I	RX#PXXP	Src	(Rickles, Botfield *et al.* 1994)
		Yes	(Sparks, Rider *et al.* 1996)
II	PXXPXR	PLCg	(Sparks, Rider *et al.* 1996)
		Ysc84	(Tong, Drees *et al.* 2002)
III	+XXPXXP	sho1	(Tong, Drees *et al.* 2002)
		Bzz1-1	(Tong, Drees *et al.* 2002)
IV	PXXPX+	Crk N	(Sparks, Rider *et al.* 1996)
		Cortactin	(Sparks, Rider *et al.* 1996)
V	PX@XXPXXP	Abl	(Rickles, Botfield *et al.* 1994)
		Myo3	(Tong, Drees *et al.* 2002)
VI	PXXDY	Eps8	(Mongiovi, Romano *et al.* 1999)
VII		Amph	(Cestra, Castagnoli *et al.* 1999)
		End	(Cestra, Castagnoli *et al.* 1999)
VIII	No peptide selected	cdc25	(Tong, Drees *et al.* 2002)
		Yar014c	(Tong, Drees *et al.* 2002)
WW Domain			
I	PPXY	YAP65	(Chen and Sudol 1995)
		Nedd4	(Schild 1996)
II	PPLP	FBP	{Bedford, 1997 #2}
		FE65	(Ermekova 1997)
III	PXXGMXPP	FBP11	(Bedford, Reed *et al* 1998)
		FBP21	(Bedford, Reed *et al.* 1998)
IV	(p-S/T)P	Pin1	(Lu, Zhou *et al.* 1999)
		Nedd-4	(Lu, Zhou *et al.* 1999)
V	RXPPGPPXR	Npw38	(Komuro, Saeki *et al.* 1999)
		NpwBP	(Komuro, Saeki *et al.* 1999)

*The SH3 domains follow the classification system outlined by Cesareni and coworkers (Cesareni *et al.*, 2002). The class VII SH3 domains bind to peptide ligands that do not resemble any previously defined class. Residues labeled #, +, @, % represent hydrophobic, basic, aromatic and hydrophilic residues respectively. PS, pT and pY refer to phosphoserine, phosphothreonine and phosphotyrosine, respectively. Amino acids in parentheses indicate equivalent substitutions.

PDZ Domains

PDZ domains (≈100 amino acids in length) can mediate protein-protein interactions of cytosolic proteins. The name derives from the three original proteins that were found to contain PDZ domains: mammalian PSD-95 (postsynaptic density), *Drosophila* Dgl (disc-large tumor suppressor), and mammalian ZO (zonula occludans). PDZ domains bind to the C-termini of specific membrane proteins. The C-terminal carboxylate fits into a highly conserved hydrophobic pocket on the PDZ domain surface and three to eight C-terminal residues confer specificity. PDZ domains have been classified into four groups based upon different C-terminal ligand binding sequences, as determined by polypeptide library methods (Table 2). For example, a chemically synthesized oriented library with randomization around a few fixed positions identified consensus motifs for class I and class II PDZ ligands that could be rationalized by structural analysis (Songyang *et al.*, 1997). From a yeast two-hybrid assay, class III ligands were defined as demonstrated by the PDZ Mint-1 binding to Ca^{2+} channel pore forming α1b subunit (Maximov *et al.*, 1999). PDZ class IV ligands are characterized by the presence of an acidic residue at the carboxy-terminal position (Vaccaro *et al.*, 2001).

Since PDZ domains bind to specific C-terminal sequences of target proteins and/or dimerize to other PDZ domains, phage display libraries, which usually display peptides as N-terminal coat fusions, have been of limited use for determining PDZ specificity. Investigators have responded with novel library display systems. For example, PDZ specificity has been determined using chemically synthesized combinatorial peptide libraries or C-terminal fusion to the *Escherichia coli* Lac repressor (Stricker *et al.*, 1997). Sidhu and co-workers also demonstrated that the M13 P8 protein could tolerate peptide fusion to its carboxy termini ("C-terminal phage display"), and screened a repertoire of peptides with free C-termini to identify ligands for two PDZ domains of the protein MAGI (Fuh *et al.*, 2000). Dente and coworkers designed a protein library with peptides displayed at high density on the surface of λ phage by fusion to the carboxy terminus of the D-capsid protein. High density peptide display can leverage avidity effects to identify weak binding ligands. This approach is especially well-suited for PDZ ligand identification, since the highest affinity PDZ ligand is not necessarily the endogenous binding partner. This approach yielded ligands that bind to class I and II PDZ domains and identified a new specificity class (class IV) (Vaccaro *et al.*, 2001).

In a protein engineering approach that examined the receptor instead of the ligand, Moelling and coworkers engineered artificial PDZ domains to obtain PDZ variants with unnatural binding specificities (Schneider *et al.*, 1999). Using a yeast two-hybrid selection system, PDZ domains were isolated that bound their artificial ligands specifically and with affinities of approximately 100 nM Kd. In addition, the use of green fluorescent protein (GFP) fusion proteins and confocal laser scanning microscopy demonstrated that the artificial PDZ domain variants direct target proteins to different subcellular compartments *in vivo*.

PTB Domains

The phosphotyrosine interaction domain (PTB) is a protein module of ≈100 to 170 amino acids that contributes to signal transduction and protein trafficking. PTB domains recognize phosphotyrosine-containing ligands with specificity conferred by the amino-terminal residues, as opposed to the C-terminal specificity determination of SH2 domains to pY (reviewed by van der Geer and Pawson, 1995). PTB domains generally share low sequence homology and feature high ligand binding selectivity. The PTB domain was first identified in the adaptor proteins Shc through co-immunnoprecipitation experiments (Kavanaugh and Williams, 1994) and the autophosphorylated epidermal growth factor receptor via screening a cDNA expression library (Blaikie *et al.*, 1994). A yeast two-hybrid binding assay has also been used to identify a PTB domain in the insulin receptor substrate 1 (IRS-1) (Gustafson *et al.*, 1995).

The ligand recognition of the PTB domain generally falls into two groups: group I PTB domains bind peptides with an NPXpY motif (van der Geer and Pawson, 1995) and group II PTB domains bind peptides with the sequence NPXY. In group II PTB domains, the tyrosine of the ligand is not necessarily phosphorylated. In a few cases, ligands that do not contain the NPXY motif have also been found using yeast two-hybrid libraries (Chien *et al.*, 1998) and tyrosine-oriented synthetic peptide libraries (Table 2). For example, the dNumb PTB domain was found to bind to peptides containing a YIGPY# motif (# denotes a hydrophobic residue) (Li *et al.*, 1997). Phage display to determine the binding partners of the Shc PTB domain was made possible from a peptide library constructed with phosphotyrosines (Dente *et al.*, 1997). In this experiment, a standard P8-displayed randomized peptide library was incubated in the presence of a specific protein tyrosine kinase (PTK). The library was enriched for phosphotyrosine by selection with anti-phosphotyrosine antibodies, before biopanning with a Shc-GST fusion protein to identify Shc PTB ligands.

SH2 domain

The Src homology 2 domain (SH2), approximately 100 amino acids in length, was one of the first intracellular domains to be discovered and characterized. A number of research groups in the early 1990's demonstrated SH2 domain specific binding to tyrosine-phosphorylated cellular proteins involved in signaling pathways. In a now classic series of experiments, optimal binding ligands were determined for class I, II, and IV SH2 domains through screening of synthetic phosphopeptide libraries (Songyang *et al.*, 1993). These experiments demonstrate SH2 domains bind ligands containing phosphorylated tyrosines and that amino acids on the carboxy-terminal side of the pY confer specificity.

Comparison of several SH2 crystal structures reveals that amino acid residues at five positions interact with the phosphotyrosine containing peptide: βD5, βE4, EF1, BG2, and BG3 (positions as designated in Eck *et al.*, 1993). Using NMR structure determination and repertoires of chemically synthesized phosphopeptides, SH2 domains have been

classified into four groups based upon the amino acid at the βD5 position of the SH2 domain and the amino acid at the pY + 1 (C-terminal side) position of the binding peptide (Table 2), (Songyang *et al*., 1993). As in other examples reviewed here, a powerful ability of library techniques is to subclassify a broad set of domain ligands.

Phosphopeptide selectivity can be adjusted by mutagenesis of the five critical SH2 domain positions. For example, a Src SH2 domain mutant with W replacing wild-type T at the EF1 position bound a Grb2 SH2-binding peptide motif (Marengere *et al*., 1994). In another example, p85 and PLCγ SH2 domains recognize a Src SH2 ligand after mutation at the βD5 position (Songyang *et al*., 1995). Synthetic SH2 domains with altered binding specificities were obtained from a SH2 domain library constructed by mutagenesis of the five critical amino acid positions of the PLCγ C-terminal SH2 domain (Malabarba *et al*., 2001). Such synthetic receptors could be valuable as reagents for sequestering ligands and altering cell fates.

SH3 domains

Src homology 3 domains (SH3) are perhaps the most widespread protein recognition domains in the proteome with over 1500 different SH3 domains found in protein database (reviewed by Cesareni *et al*., 2002). Between 50 to 70 amino acids in length, SH3 domains are present in a number of eukaryotic signal transduction, cytoskeletal organization, and membrane traffic proteins. SH3 domains bind to proline-rich peptides that fold into a polyproline type II helices. All SH3 domains share a highly conserved fold, characterized by a sandwich of two β sheets composed of three strands. One side of the β-sheet is typically hydrophobic and constitutes the ligand-binding surface. The SH3 domain surface features three shallow pockets, two "LP dipeptide pockets" and one specificity pocket. The LP dipeptide pockets are 25 Å long and 10 Å wide, large enough to accommodate each of the PXXP prolines and a hydrophobic residue. Typical binding affinities are 1-100 μM for synthetic peptide ligands/SH3 domains and Kd of 250 nM for full length protein/SH3 domain interactions.

In general, SH3 domains bind proline rich peptides (Table 2) with the motif PXXP. Phage display libraries and combinatorial chemistry have defined the specificity of individual SH3 domains such as the SH3 domains of Src, Lyn, Abl, p53bp2, and Grb (Rickles *et al*.,1994; Sparks *et al*., 1994; Sparks *et al*., 1996). Initial experiments applied randomized peptide libraries expressed as N-terminal fusions to P3 of bacteriophage M13 with affinity selection for binding to the targeted SH3 domain fused to GST. As consensus for the PXXP motif emerged, biased peptide libraries were screened to optimize ligand-binding sequences. More recently, comprehensive library-based screening of SH3 domain binding preferences by Cesareni and coworkers was used to classify SH3 domains into eight classes based on peptide recognition specificity with six of the eight groups containing the PXXP motif (Cesareni *et al*., 2002). In addition, protein library methods have identified a ninth class of SH3 domains that recognize a RXXK motif (Berry *et al.,* 2002; Kato *et al.,* 2000). Most SH3 domains bind to peptides that are categorized as class I or class II ligands. Class I ligands bind in an N- to C-terminal orientation relative to the SH3 domain with a RX#PXXP motif (# denotes a hydrophobic residue). Class II SH3 ligands bind in the opposite orientation, C- to N-terminal, with a PXXPXR consensus sequence (Figure 4). The orientation of the peptide is dictated by the location of a positively charged residue, which forms a salt bridge from the ligand to an acidic residue in the SH3 domain (reviewed by Kay *et al*., 2000).

There are many SH3 domains that do not comply with the polyproline-binding model. For example, Di Fiore and coworkers used a phage-displayed random peptide library to show that the SH3 domain of Eps8 binds to a PXXDY motif (Mongiovi *et al*., 1999). The second SH3 domain of FYB/SLAP, an immune cell adaptor, was recently shown to form a complex with proteins containing a tyrosine-based RKXXYXXY motif (Kang *et al*., 2000). Additionally, a yeast two-hybrid screen identified the interaction of the p53BP2 SH3 domain with a VPMRLR motif of the YAP protein (Espanel and Sudol, 2001). Thus, while many SH3 domains evolved to recognize proline-rich sequences, other modes of binding to the rather shallow SH3 binding groove are possible.

Other protein library-based efforts have focused upon identifying endogenous binding partners for SH3 domains. A mouse embryo cDNA expression library identified the Wiskott-Aldrich syndrome protein (WASP, a putative CDC42 effector) and a serine/threonine protein kinase (PKR2, a homolog of the Rho effector PKN) as possible

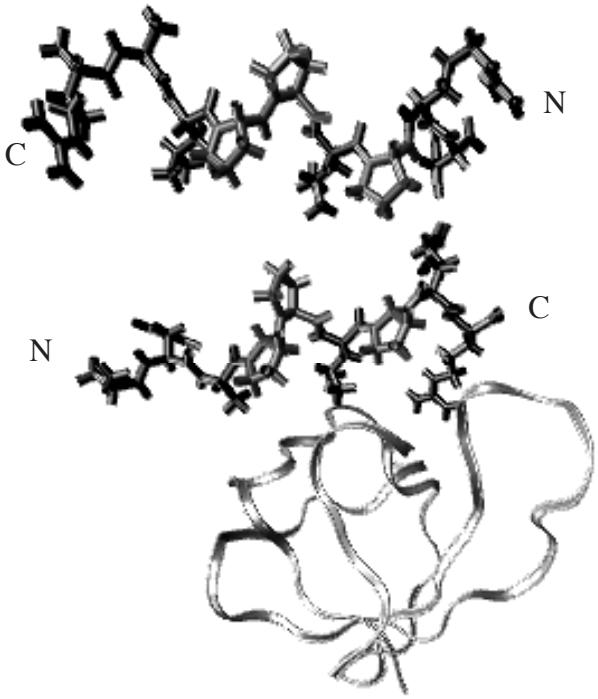


Figure 4. A general model for the two binding interactions of the SH3 (Src domain) with class I (PDB accession code: 1RLQ (Feng *et al*., 1994), top peptide) and class II (PDB accession code: 1PRM (Feng *et al*., 1994), bottom peptide). The ligand c-Src SH3 complex solution structures were determined by multidimensional nuclear magnetic resonance spectroscopy (Feng *et al*., 1994).

binding ligands to the three consecutive SH3 domains of the NCK adaptor protein (Quillman *et al.*, 1996). In another example, phage display libraries demonstrated that the SH3 domains of endophilin and amphiphysin bind to distinct proline-rich regions of synaptojanin 1 (Cestra *et al.*, 1999). Cesareni and coworkers have also displayed a cDNA expression library on the capsid of bacteriophage lambda to identify potential receptors for synaptojanin 1. The cDNA expression libraries were screened with proline-rich fragment of synaptojanin 1, and seven ligands containing SH3 or WW domains were identified, the physiological relevance of three putative ligands remains to be established (Zucconi *et al.*, 2001).

In an effort to define the protein interaction network of all cellular SH3 domains, Cesareni and coworkers characterized the binding potential of the entire SH3 repertoire of *Saccharomyces cerevisiae*, using a combination of phage display libraries and large-scale yeast two-hybrid binding assays. From phage-display, a network containing 394 interactions among 206 proteins was identified. The two-hybrid yeast analysis identified 233 interactions among 145 proteins with 59 interactions common to both analyses. Multiple SH3 interactions were found for Las 17 protein, a member of the WASP family of actin-assembly proteins (Tong *et al.*, 2002). The richness of biological data obtained from these experiments demonstrates the effectiveness of library approaches for unraveling complex systems in cells.

Libraries of mutant SH3 domains have been used to identify SH3 domain variants with altered specificity. For example, phage-displayed libraries of artificial Hck-derived SH3 domains were used to identify inhibitors of the HIV accessory protein Nef (Hiipakka *et al.*, 2001). These libraries of the Hck SH3 domain focused randomization on the RT loop of the protein, which is one determinant of SH3 domain specificity. SH3 domains, selected from the library and shown to bind Nef with high affinity (nM dissociation constants), were potent *in vivo* inhibitors of Nef, which contributes to HIV pathogenicity. This study provides proof-of-concept for the potential useful of expanding the anti-HIV arsenal to include Nef inhibitors. This research also demonstrates the efficacy of using phage-displayed domain libraries to derive "pertubagens" that alter signal transduction, for use as cell biology reagents and as potential therapeutic proteins. In a different study that also resulted in a library of potential perturbagens, phage-displayed libraries of SH3 domains based upon the Abl scaffold were used for an ambitious attempt to elucidate predictive rules for molecular recognition by SH3 domains (Panni *et al.,* 2002). The authors found that just a few amino acid substitutions were sufficient to radically shift target binding preferences in unpredictable ways, and, thus, soberly conclude that it is not possible to define a simple set of rules for target recognition by SH3 domains. These studies contribute substantially to our understanding of protein evolution, through the insight that small numbers of mutations to protein domains can dictate dramatic changes in binding specificities.

WW domains

The WW domain is one of the smallest known naturally occurring protein modules, consisting of 38 to 40 amino acids. The WW label refers to two conserved W residues spaced 20 –22 amino acids apart that contribute to both domain structure and function. WW domains contain an anti-parallel three-stranded β-sheet that forms a shallow binding pocket for various polyproline peptide motifs. Although WW domains share a similar preference for polyproline motifs with SH3 domains, the two domains feature distinct structures.

The specificity of the WW domain was first revealed by a yeast two-hybrid assay that uncovered two proteins (WBP1 and WBP2) with affinity for the YAP WW domain (Chen and Sudol, 1995). Alanine scanning experiments confirmed that a PPXY peptide motif mediates binding between the two proteins. Phage display libraries screened for binding to WW domains further indicate the importance of the PPXY motif (Linn *et al.*, 1997; Kasanov *et al.*, 2001), which is designated as one of the major ligand binding groups (class I) of WW domains. From cDNA expression libraries and yeast two-hybrid libraries ligands grouped into classes II, III, IV and V have been defined (Table 2) (Bedford *et al.*, 1998; Lu *et al.*, 1999; Komuro *et al.*, 1999).

WW domains mediate signal complexes linked to several human diseases including Liddle's syndrome of hypertension, muscular dystrophy, Huntington's and Alzeheimer's diseases, and cancer. Thus, based on peptide ligands derived from phage display, efforts have been directed at determining the natural ligands of WW domains. One successful example is the association of PPXY ligands isolated from phage display libraries with sequences that correspond to the β-subunit of human epithelial sodium channels (Kay *et al.*, 2000). This interaction is physiologically important because truncations or substitutions within the PPXY motif result in Liddle's syndrome.

Enzyme dissection with polypeptide libraries

Dissection of key residues used by enzymes to catalyze chemical reactions with remarkable specificity and efficiency can apply techniques similar to the ones used for binding epitope mapping described in the previous sections. Three aspects of enzyme function can be investigated by protein libraries – catalysis, specificity and regulation. The information obtained from such libraries can then provide a foundation to construct novel catalysts, design inhibitors and explore biological pathways.

Enzyme libraries for investigating catalysis and specificity

Investigation of enzyme catalysis can apply libraries of enzymes with substituted active site residues, which can also provide enzymes with improved catalytic activity. For example, a yeast surface display library fashioned by the Ueda group explored mutations to six amino acids comprising the lid domain of *Rhizopus oryzae* lipase (ROL) to observe their effect on soybean oil hydrolysis (Shiraga

et al., 2002). The lid domain of lipases composes the top portion of the active site and directly contacts the substrate. ROL responds to increased micellar concentration of the substrate, through changes in the conformation of the lid domain conformation, which in turn regulates enzyme activity. The ROL enzyme library was screened using a yeast display method, and several mutants were found with increased lipase activity. Activity of ROL was found to favor three sequential residues with basic, followed by polar, and then non-polar sidechains. This sequence of sidechain functionalities aligns directly with the substrate and the position of the sequence in the lid domain can also influence substrate specificity.

In addition to improving the catalytic activity of enzymes, information about motifs important to an active site can also be gained via enzyme libraries. In one example from the Craik laboratory, an enzyme library, constructed by site-directed mutagenesis, was combined with metabolic selection to probe the specificity of rat anionic trypsin (Evnin *et al.,* 1990). Trypsin, a proteolytic enzyme, cleaves amide bonds following arginine and lysine residues. Crystal structures of the protein had suggested that trypsin residues 189 and 190 participate heavily in substrate specificity. Upon enzyme library screening with substitutions at residues 189 and 190, a preference for a negative charge in either position was demonstrated, and the highest level of catalytic activity was obtained from an aspartic acid in position 189. Thus, critical residues for substrate specificity were identified by a library of trypsin active site variants.

Libraries of enzyme substrates and regulatory motifs

While mutations to an enzyme are a useful tool for understanding enzyme specificity, libraries of enzyme substrates and binding partners can also provide detailed mechanistic information. Three different types of peptide libraries have been utilized to analyze enzyme specificity and binding preferences – random, oriented and positional. For example, randomized peptide libraries have been used to probe the activities of the *Leishmania mexicana* cysteine protease (CPB). *L. mexicana* can cause lesions in humans, and CPB is expected to offer a good target for therapeutic intervention. Meldal and colleagues have reported detailed CPB substrate specificities, which include a distinctive preference for basic residues at alternating subsites (St. Hilaire *et al.,* 2000). CPB specificities were identified from a library chemically synthesized on resin with randomized 7-mer peptides tethered to a fluorescence quencher to aid library screening. In this experiment, beads with substrates for the protease become fluorescent, due to removal of the quencher. Building upon these results, the authors have also used a one-bead-two compound library to identify a substrate inhibitor that binds to *Leishmania* CPB with antiparasite efficacy (IC_{50} around 50 μM) (St. Hilaire *et al.,* 2002). In their one-bead-two compound format, potential enzyme inhibitors and an enzyme substrate (tethered to a fluorescence quencher) were attached to the same bead. Thus, upon each bead, a resin-bound substrate competes against an inhibitor with fluorescence indicating lack of inhibitory activity (and vice versa). This example illustrates how library approaches to deciphering enzymatic activity can guide medicinal chemistry efforts.

While combinatorial libraries of randomized peptides have been successful for elucidating enzyme specificities, some enzymes require a highly conserved motif for function. Understanding such enzymes can be expedited by an oriented library approach. Oriented libraries have specific positions with fixed residues. For instance, AKT protein kinase (also called Rac-protein kinase or protein kinase B) has an active site similar in structure to protein kinase C. AKT has been shown to be involved in many cellular activities including cell growth, differentiation, transcription, and translation. Substrates for AKT are numerous, and a minimal binding motif has been identified as RXRXX(S/T) (X denotes any amino acid). With this information, Cantley and coworkers used AKT substrate libraries to investigate the importance of residues neighboring the site the S/T phosphorylation site (Obata *et al.,* 2000). By employing a solution-phase peptide library approach the optimal AKT substrate was identified and fit to a structural motif for AKT substrate binding (Figure 5). In this example, a chemically synthesized library was screened, phosphorylated peptides collected and Edman degradation peptide sequencing performed.

In addition to identification of optimal substrates and inhibitors, oriented peptide libraries have also been used to suggest possible targets for cell signaling pathways. For example, DNA damage activates a number of repair mechanisms, including the mammalian kinases Chk1 and Chk2. The two proteins serve as checkpoints in DNA repair and regulate several other cellular processes. To identify potential phosphorylation targets of these enzymes, the Rathbun group used a chemically synthesized peptide library to find substrate motifs (O'Neil *et al.,* 2002). When

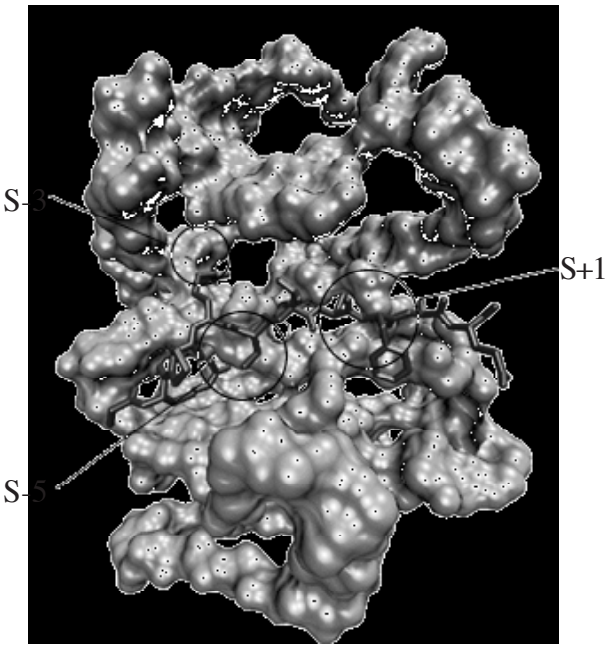


Figure 5. AKT catalytic cleft bound to a peptide inhibitor Gsk3 (PDB accession code: 1O6K (Yang *et al.,* 2002)). Regions that correspond with substrate positions as identified by substrate libraries (Toshiyuki *et al.*, 2000.) have been highlighted.

Table 3.*

Chk 1	
Sequence	Target
SGLYRSPS^{216}MPENLNR	Human Cdc25C
QRLFRSPS^{309}MPCSVIR	Human Cdc25B
ELLERLES^{714}FISSVRG	Son of sevenless protein homologue 2 (SOS-2)
HRLGRCRS^{35}VKEFEKL	Cell division protein kinase 10 (EC 2.7.1)
LHLERHDS^{327}VFVAAHT	Helicase SKI2W (helicase-like protein) (HLP)
FGLARIYS^{171}FQMALTS	CDK6
FKLFRVRS^{37}FEKAPEE	V(D)J recombination activating protein 1(RAG)
KDLERYNS^{633}QMKRAIE	Protein homologue 1 (DNA mismatch repair protein PMS1)
QQVERHFT^{2083}MNELTEL	Transcriptional regulator ATRX (*X*-linked helicase II)
SDIFRYSS^{357}MEDHLEI	DNA polymerase delta small subunit
VLQRIES1^{94}LEEEIISK	DNA repair protein RAD51 homologue 2 (R51H2-alternative splice)
GDVYKYFS^{225}RNGYNYT	DNA ligase IV
Chk 2	
Sequence	Target
RSLSRSNS^{472}LPHPAGG	Mouse DMR-N9 Protein
APVVRTPS^{1023}IQPSLLP	Human nuclear pore complex protein NUP214 (nucleoporin NUP214)
SNLERTHS^{1076}YKVPEK	Mouse Werner syndrome helicase homologue
TRRSRTES^{307}ITATSPA	Human insulin receptor substrate-1 (IRS-1)
NRLKRS^{395}FSLDIKSAY	Human dual specificity protein phosphatase 8 (h-VH-5)
ERLMRLDS^{747}FLERPST	Human kinesin-like protein KIF3C
ELLSRTSS^{426}LLPKMSR	Human DNA repair protein XRCC9
ENIKRTLS^{1075}FRKKRSH	Human DNA polymerase ζ catalytic subunit
PALGRSWS^{269}FVPSTRI	Human DNA repair protein RAD51 homologue 4 (R51H3) (TRAD)
QSVSRQPS^{162}FTYSEWM	Human protein-tyrosine phophatase striatum-enriched
KKIKRQLS^{12}MTLRGGR	
FGLARAKS^{312}IPTKTYS	Human serine-threonine-protein kinase PCTAIRE-1
KKLERNLS^{435}FEIKKVP	Human protein-tyrosine phosphatase G1
CSFYRTRT^{46}LQTLSSE	Human DCAMKL1
RSLSRSNS^{472}LPHPAGG	Mouse DMR-N9 Protein
APVVRTPS^{1023}IQPSLLP	Human nuclear pore complex protein NUP214 (nucleoporin NUP214)
SNLERTHS^{1076}YKVPEK	Mouse Werner syndrome helicase homologue
TRRSRTES^{307}ITATSPA	Human insulin receptor substrate-1 (IRS-1)

* Data from O'Neill *et al.*, 2002.

this data was searched using SCANSITE (software designed to identify functional sites from protein databases) against a mammalian genome database, a multitude of possible cellular targets were found (Table 3). Information obtained from such approaches has potential applications for mapping cell signaling pathways, a critical challenge in the post-genomic era.

Positional-scanning synthetic combinatorial libraries use an iterative approach to analyze substrate affinity; each position is optimized individually before moving to the next residue. The Craik group has reported using this technique to compare the substrate specificity of schistosome and human legumain enzymes (Mathieu *et al.,* 2002). Legumains, also called asparaginyl endopeptidases, are members of the cysteine protease family and are specific for hydrolysis of the carboxyl terminal asparagine amino acids. The group initially created a diverse substrate library at the P1 position to verify the preference of legumain for an Asn residue. After confirming this preference, P2 and P3 libraries were synthesized on solid support and screened using a fluorescence-based assay. The results demonstrate that schistosome legumain has an optimal substrate preference for the peptide sequence TAN while the human legumain prefers the sequence PTN. This data could potentially be used to selectively inhibit either of the enzymes.

Corey and colleagues have reported another example of iterative optimization of substrate activity towards prostate-specific antigen (PSA) (Coombs *et al.,* 1998). Elevated serum levels of PSA, a serine protease, are associated with metastatic cancer. PSA has been shown to cleave laminin, fibronectin, and growth-factor-binding protein-3. Inhibition of PSA may prove effective in fighting prostate cancers; however, PSA shares 62% identity with human pancreatic kallikrein (hk1) and 78% identity with human glandular kallikrein (hk2) making determination of specificity critical to eventual inhibition. The researchers identified that the optimal substrate for PSA as SS(Y/F)↓S(G/S) (down arrow indicates site of cleavage), through a combination of substrate phage display and iterative optimization of substrates. In substrate phage display, the combinatorial library of substrate peptides (in this case, a library of octameric peptides) are placed between an antibody-binding epitope and the phage surface. Substrates for PSA were cleaved off phage bound to the antibody attached to a solid support. PSA cleaved the optimized substrates with efficiencies up to a thousand fold higher than peptides based upon the likely physiological targets of the enzyme. Although optimized PSA substrates

could also be cleaved by the protease chymotrypsin, the authors suggest that their approach could be adapted to identify PSA-specific substrates.

Summary and Outlook

In this review, we have highlighted the importance of using library approaches (phage display, yeast two-hybrid and others) to identify the cellular binding partners of protein domains. In many cases the peptide motifs obtained from these library methods have similar affinity, specificity and primary structure as the endogenous interacting protein. Thus, a valuable approach to decipher the functions and interactions of the proteome can be achieved in two steps. First, peptide ligands to a target protein are isolated, from large libraries of potential ligands. Second, database mining of ligands uncovered from peptide libraries can suggest possible interacting proteins from a sequenced genome. Since peptide library members focus exclusively on receptor ligand affinity, other factors such as local protein concentration could play a significant role in determining whether a particular interaction forms *in vivo*.

It should be emphasized that not all sequences selected from peptide libraries correspond to an endogenous ligand. However, success in this field illustrates the power of protein libraries to uncover an initial lead, which can be used for further studies. Information obtained from peptide library approaches also provides useful information and tools for drug discovery. For example, binding information derived from phage display methods has been used to validate particular proteins as potential drug targets *in vivo* (Stauffer *et al*., 1997; Tao *et al*., 2000). In addition, phage display ligands have been used to guide the design of peptidomimetics and for development of high-throughput small molecule inhibitor screens (Kay *et al*., 1998; Gron *et al*., 2000).

The examples reviewed here illustrate the power for polypeptide libraries to identify binding partners and key residues for receptors. These abilities complement ongoing efforts in structural proteomics and genomic sequencing. In the future, we can envision more comprehensive applications of polypeptide libraries. For example, the tour de force identification of ligands for every SH3 domain in the yeast cell described above (Tong *et al*., 2002) illustrates the tremendous potential for mapping entire signaling networks with the techniques described here. In mapping key residues, comprehensiveness is exemplified by recent applications of whole protein shotgun scanning to map the contributions to protein functionality from every sidechain in a protein (Roth *et al*., 2002). Enzyme libraries to alter substrate and product specificities have often applied techniques to introduce variation throughout the protein (e.g., error-prone PCR). With improved high throughput bioinformatic and structure determination methods, new tools to rationalize the results from enzyme engineering could be gained. Such efforts are expected to uncover new mechanistic understanding for how proteins function *in vitro* and in cells.

Acknowledgements

We gratefully acknowledge support by the NIH-funded Post-Baccalaureate Research Education Program (training grant to J.E.D.) and the Arnold and Mabel Beckman Foundation.

References

Amstutz, P., Pelletier, J.N., Guggisberg, A., Jermutus, L., Cesaro-Tadic, S., Zahnd, C., and Pluckthun, A. 2002. *In vitro* selection for catalytic activity with ribosome display. J. Am. Chem. Soc. 124: 9396-9403.

Avrantinis, S.K., Stafford, R.L., Tian, X., and Weiss, G.A. 2002. Dissecting the streptavidin-biotin interaction by phage-displayed shotgun scanning. ChemBioChem. 3: 1229-1234.

Barrick, J.E., Takahashi, T.T., Balakin, A., and Roberts, R.W. 2001. Selection of RNA-binding peptides using mRNA-peptide fusions. Methods. 23: 287-293.

Bedford, M.T., Chan, D.C., and Leder, P. 1997. FBP WW domains and the Abl SH3 domain bind to a specific class of proline-rich ligands. EMBO J. 16: 2376-2383.

Bedford, M.T., Reed, R., and Leder, P. 1998. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif. Proc. Natl. Acad. Sci. USA. 95: 10602-10607.

Berry, M.D., Nash, P., Liu, S.K.W., Pawson, T., and McGlade, C.J. 2002. A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling. Curr. Biol. 12: 1336-1341.

Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B. 1994. A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. J. Biol. Chem. 269: 32031-32034.

Borg, J.P., Ooi, J., Levy, E., and Margolis, B. 1996. The phosphotyrosine interaction domains of X11 and FE65 bind to distinct sites on the YENPTY motif of amyloid precursor protein. Mol. Cell. Biol. 16: 6229-6241.

Bracci, L., Lozzi, L., Lelli, B., Pini, A., and Neri, P. 2001. Mimotypes of the nicotinic receptor binding site selected by a combinatorial peptide library. Biochemistry. 40: 6611-6619.

Cesareni, G., Panni, S., Nardelli, G., and Castognoli, L. 2002. Can we infer peptide recognition specificity mediated by SH3 domains? FEBS Lett. 513: 38-44.

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

Chen, H.I., and Sudol, M. 1995. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. Proc. Natl. Acad. Sci. USA. 92: 7819-7823.

Chien, C.T., Wang, S., Rothenberg, M., Jan, L.Y., and Jan, Y.N. 1998. Numb-associated kinase interacts with the phosphotyrosine binding domain of Numb and antagonizes the function of Numb *in vivo*. Mol. Cell. Biol. 18: 598-607.

Cho, G., Keefe, A.D., Liu, R., Wilson, D.S., and Szostak, J.W. 2000. Constructing high complexity synthetic libraries of long ORFs using *in vitro* selection. J. Mol. Biol. 297: 309-319.

Clackson, T. and Wells, J.A. 1995. A hot spot of binding energy in a hormone-receptor interface. Science. 267: 383-386.

Coombs, G.S., Bergstrom, R.C., Pellequer, J.L., Baker, S.I., Navre, M., Smith, M.M., Tainer, J.A., Madison, E.L., and Corey, D.R. 1998. Substrate specificity of prostate-specific antigen (PSA). Chem. Biol. 5: 475-488.

Cunningham, B.C. and Wells, J.A. 1993. Comparison of a structural and a functional epitope. J. Mol. Biol. 234: 554-563.

Danner, S. and Belasco, J.G. 2001. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. Proc. Natl. Acad. Sci. USA. 98: 12954-12959.

Daugherty, P.S., Olsen, M.J., Iverson, B.L., and Georgiou, G. 1999. Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface. Protein Eng. 12: 613-621.

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

Di Fiore, P.P., Pelicci, P.G., and Sorkin, A. 1997. EH: a novel protein-protein interaction domain potentially involved in intracellular sorting. Trends Biochem. Sci. 22: 411-413.

Eck, M.J., Shoelson, S.E., and Harrison, S.C. 1993. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56lck. Nature. 362: 87-91.

Ermekova, K.S., Zambrano, N., Linn, H., Minopoli, G., Gertler, F., Russo, T., and Sudol, M. 1997. The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of Drosophilia enabled. J. Biol. Chem. 272: 32869-32877.

Espanel, X., and Sudol, M. 2001. Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. J. Biol. Chem. 276: 14514-14523.

Evnin, L.B., Vasquez, J.R., and Craik, C.S. 1990. Substrate specificity of trypsin investigated by using a genetic selection. Proc. Natl. Acad. Sci. USA. 87: 6659-6663.

Fazioli, S., Minichiello, L., Matoskova, B., Wong, W.T., and Di Fiore, P.P. 1993. Eps15, a novel tyrosine kinase substrate, exhibits transforming activity. Mol. Cell. Biol. 13: 5814-5828.

Feng, J.N., Model, P., and Russel, M. 1999. A trans-envelope protein complex needed for filamentous phage assembly and export. Mol. Microbiol. 34: 745-755.

Feng, S., Chen, J.K., Yu, H., Simon, J.A., and Schreiber, S.L. 1994. Two binding orientations for peptides to the Scr SH3 domain: development of a general model for SH3-ligand interactions. Science. 266: 1241-1247.

Frangioni, J.V., LaRiccia, L.M., Cantley, L.C., and Montminy, M.R. 2000. Minimal activators that bind to the KIX domain of p300/CBP identified by phage display screening. Nat. Biotechnol. 18: 1080-1085.

Fuh, G., Pisabarro, M.T., Li. Y., Quan, C., Lasky, L.A., and Sidhu, S.S. 2000. Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. J. Biol. Chem. 275: 21486-21491.

Green, N.M. 1990. Avidin and streptavidin. Methods Enzymol. 184: 51-67.

Gron, H., and Hyde-DeRuyscher, R. 2000. Peptides as tools in drug discovery. Curr. Opin. Drug Disc. Dev. 3: 636-645.

Gustafson, T.A., He, W., Craparo, A., Schaub, C.D., and O'Neil, T.J. 1995. Phosphotyrosine-dependent interaction of SHC and insulin receptor substrate 1 with the NPEY motif of the insulin receptor via a novel non-SH2 domain. Mol. Cell. Biol. 15: 2500-2508.

Hanes, J. and Pluckthun, A. 1997. *In vitro* selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA. 94: 4937-4942.

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.

Huang, W., Zhang, Z., and Palzkill, T. 2000. Design of potent beta-lactamase inhibitors by phage display of beta-lactamase inhibitory protein. J. Biol. Chem. 275: 14964-14968.

Kang, H., Freund, C., Duke-Cohan, J.S., Musacchio, A., Wagner, G., and Rudd, C.E. 2000. SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55. EMBO J. 19: 2889-2899.

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

Kato, M., Miyazawa, K., and Kitamura, N. 2000. A deubiquitination enzyme UBPY interacts with the Src homology 3 domain of Hrs-binding protein via a novel binding motif PX(V/I)(D/N)RXXKP*. J. Biol. Chem. 275: 37481-37487.

Kavanaugh, W.M., and Williams, L.T. 1994. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. Science. 266: 1862-1865.

Kay, B.K., Kasanov, J., Knight, S., and Kurakin, A. 2000. Convergent evolution with combinatorial peptides. FEBS Lett. 480: 55-62.

Kay, B.K., Kurakin, A., and Hyde-DeRuyscher, R. 1998. From peptides to drugs via phage display. Drug Discov Today. 3: 370-378.

Kay, B.K., Williamson, M.P., and 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.

Kieke, M.C., Shusta, E.V., Boder, E.T., Teyton, L., Wittrup, K.D., and Kranz, D.M. 1999. Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl. Acad. Sci. USA. 96: 5651-5656.

Komuro, A., Saeki, M., and Kato, S. 1999. Association of two nuclear proteins, Npw38 and NpwBP, via the interaction between the WW domain and a novel proline-

rich motif containing glycine and arginine. J. Biol. Chem. 274: 36513-36519.

Kramer, A., Reineke, U., Dong, L., Hoffman, B., Hoffmüller, U., Winkler, D., Volkmer-Engert, R., and Schneider-Mergener, J. 1999. Spot synthesis: observations and optimizations. J. Peptide Res. 54: 319-327.

Kurz, M., Gu, K., Al-Gawari, A., and Lohse, P.A. 2001. cDNA-protein fusions: covalent protein-gene conjugates for the *in vitro* selection of peptides and proteins. ChemBioChem. 2: 666-672.

Kurz, M., Gu, K., and Lohse, P.A. 2000. Psoralen photo-crosslinked mRNA-puromycin conjugates: a novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Res. 28: E83.

Li, C., Plugariu, C.G., Bajgier, J., White, J.R., Liefer, K.M., Wu, S.J., and Chaiken, I. 2002. Coiled coil miniprotein randomization on phage leads to charge pattern mimicry of the receptor recognition determinant of interleukin 5. J. Mol. Recognit. 15: 33-43.

Li, S., Millward, S., and Roberts, R. 2002. *In vitro* selection of mRNA display libraries containing an unnatural amino acid. J. Am. Chem. Soc. 124: 9972-9973.

Li, S.C., Songyang, Z., Vincent, S.J., Zwahlen, C., Wiley, S., Cantley, L., Kay, L.E., Forman-Kay, J., and Pawson, T. 1997. High-affinity binding of the Drosophilia Numb phosphotyrosine-binding domain to peptides containing a Gly-Pro-(p)Tyr motif. Proc. Natl. Acad. Sci. USA. 94: 7204-7209.

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

Lowman, H.B., Bass, S.H., Simpson, N., and Wells, J.A. 1991. Selecting high-affinity binding proteins by monovalent phage display. Biochemistry. 30: 10832-10838.

Lu, P.J., Zhou, X.Z., Shen, M., and Lu, K.P. 1999. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. Science. 283: 1325-1328.

Malabarba, M.G., Milia, E., Faretta, M., Zamponi, R., Pelicci, P.G., and Di Fiore, P.P. 2001. A repertoire library that allows the selection of synthetic SH2s with altered binding specificities. Oncogene. 20: 5186-5194.

Marengere, L.E., Songyang, Z., Gish, G.D., Schaller, M.D., Parsons, J.T., Stern, M.J., Cantley, L.C., and Pawson, T. 1994. SH2 domain specificity and activity modified by a single residue. Nature. 369: 502-505.

Mathieu, M.A., Bogyo, M., Caffrey, C.R., Choe, Y., Lee, J., Chapman, H., Sajid, M., Craik, C.S., and McKerrow, J.H. 2002. Substrate specificity of schistosome versus human legumain determined by P1–P3 peptide libraries. Mol. Biochem. Parasitol. 121: 99-105.

Maximov, A., Sudhof, T.C., and Bezprozvanny, I. 1999. Association of neuronal calcium channels with modular adaptor proteins. J. Biol. Chem. 274: 24453-24456.

Mongiovi, A.M., Romano, P.R., Panni, S., Mendoza, M., Wong, W.T., Musacchio, A., Cesareni, G., and Di Fiore, P.P. 1999. A novel peptide-SH3 interaction. EMBO J. 18: 5300-5309.

Nemoto, N., Miyamoto-Sato, E., Husimi, Y., and Yanagawa, H. 1997. *In vitro* virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome *in vitro*. FEBS Lett. 414: 405-408.

Obata, T., Yaffe, M.B., Leparc, G.G., Piro, E.T., Maegawa, H., Kashiwagi, A., Kikkawa, R., and Cantley, L.C. 2000. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. J. Biol. Chem. 275: 36108-36115.

O'Neil, T., Giarrantani, L., Chen, P., Iyer, L., Lee, C.H., Bobiak, M., Kanai, F., Zhou, B.B., Chung, J.H., and Rathbun, G.A. 2002. Determination of substrate motifs for human Chk1 and hCds1/Chk2 by the oriented peptide library approach. J. Biol. Chem. 277: 16102-16115.

Panni, S., Dente, L., and Cesarini, G. 2002. *In vitro* evolution of recognition specificity mediated by SH3 domains reveals target recognition rules. J. Biol. Chem. 277: 21666-21674.

Paoluzi, S., Castagnoli, L., Lauro, I., Salcini, A.E., Coda, L., Fre, S., Confalonieri, S., Pelicci, P.G., Di Fiore, P.P., and Cesareni, G. 1998. Recognition specificity of individual EH domains of mammals and yeast. EMBO J. 17: 6541-6550.

Quilliam, L.A., Lambert, Q.T., Mickelson-Young, L.A., Westwick, J.K., Sparks, A.B., Kay, B.K., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., and Der, C.J. 1996. Isolation of a NCK-associated kinase, PRK2, an SH3-binding protein and potential effector of Rho protein signaling. J. Biol. Chem. 271: 28772-28776.

Reineke, U., Volkmer-Engert, R., and Schneider-Mergener, J. 2001. Applications of peptide arrays prepared by the SPOT-technology. Curr. Opin. Biotechnol. 12: 59-64.

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

Roberts, R.W. 1999. Totally *in vitro* protein selection using mRNA-protein fusions and ribosome display. Curr. Opin. Chem. Biol. 3: 268-273.

Roberts, R.W. and Szostak, J.W. 1997. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. Proc. Natl. Acad. Sci. USA. 94: 12297-12302.

Rosenberg, A., Griffin, K., Studier, F.W., McCormick, M., Berg, J., Novy, R. and Mierendorf, R. 1996. T7 Select Phage Display System: A powerful new protein display system based on bacteriophage T7. inNovations (newsletter Novagen, Inc). 6: 1-6.

Roth, T.A., Weiss, G.A., Eigenbrot, C., and Sidhu, S.S. 2002. A minimized M13 coat protein defines the requirements for assembly into the bacteriophage particle. J. Mol. Biol. 322: 357-367.

Salcini, A.E., Confalonieri, S., Doria, M., Santolini, E., Tassi, E., Minenkova, O., Cesareni, G., Pelicci, P.G., and Di Fiore, P.P. 1997. Binding specificity and *in vivo* targets of the EH domain, a novel protein-protein interaction module. Gene Dev. 11: 2239-2249.

Scharn, D., Wenschuh, H., Reineke, U., Schneider-Mergener, J., and Germeroth, L. 2000. Spatially addressed synthesis of amino- and amino-oxy-substituted 1,3,5-triazine arrays on polymeric membranes. J. Comb. Chem. 2: 361-369.

Schild, L., Lu, Y., Gautschi, I., Schneeberger, E., Lifton, R.P., and Rossier, B.C. 1996. Identification of a PY motif in the epithelial Na channel subunits as a target sequence for mutations causing channel activation found in Liddle syndrome. EMBO J. 15: 2381-2387.

Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S.A., Baechi, T., Moelling, K., and Hovens, C.M. 1999. Mutagenesis and selection of PDZ domains that bind new protein targets. Nat. Biotechnol. 17: 170-175.

Shiraga, S., Ueda, M., Takahashi, S., and Tanaka, A. 2002. Construction of the combinatorial library of Rhizopus oryzae lipase mutated in the lid domain by displaying on yeast cell surface. J. Mol. Catal. B: Enzymatic. 17: 167-173.

Sidhu, S.S., Lowman, H.B., Cunningham, B.C., and Wells, J.A. 2000. Phage display for selection of novel binding peptides. Methods Enzymol. 328: 333-363.

Sidhu, S.S. and Weiss, G.A. 2002. Oligonucleotide-directed construction of phage display libraries. In: Phage Display: A practical approach. H.L. Lowman and T. Clackson, ed. Oxford University Press, In press.

Smith, G.P. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the viron surface. Science. 228: 1315-1317.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. 1997. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science. 275: 73-77.

Songyang, Z., Gish, G., Mbamalu, G., Pawson, T., and Cantley, L.C. 1995. A single point mutation switches the specificity of group III Src homology (SH) 2 domains to that of group I SH2 domains. J. Biol. Chem. 270: 26029-26032.

Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X.R., Barbacid, M., Sabe, H., and Hanafusa, H., and Yi, T. 1994. Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. Mol. Cel. Biol. 14: 2777-2785.

Songyang, Z., Shoelson, S.E. Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., Neel, B., Birge, R., Fajardo, J., Chou, M., Hanafusa, H., Schaffhausen, B., and Cantley, L.C. 1993. SH2 domains recognize specific phosphopeptide sequences. Cell. 72: 767-778.

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

Sparks, A.B., Rider, J.E., Hoffman, N.G., Fowlkes, D.M., Quillam, L.A., and Kay, B.K. 1996. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCγ, Crk, and Grb2. Proc. Natl. Acad. Sci. USA. 93: 1540-1544.

Stauffer, T.P., Martenson, C.H., Rider, J.E., Kay, B.K., and Meyer, T. 1997. Inhibition of Lyn function in mast cell activation by SH3 domain binding peptides. Biochemistry. 36: 9388-9394.

St. Hilaire, P.M., Alves, L.C., Sanderson, S.J., Mottram, J.C., Juliano, M.A., Juliano, L., Coombs, G.H., and Meldal, M. 2000. The substrate specificity of a recombinant cysteine protease from Leishmania mexicana: application of a combinatorial peptide library approach. ChemBioChem. 1: 115-122.

St. Hilaire, P.M., Alves, L.C., Herrera, F., Renil, M., Sanderson, S J., Mottram, J.C., Coombs, G.H., Juliano, M.A., Juliano, L., Arevalo, J., and Meldal, M. 2002. Solid-phase library synthesis, screening, and selection of tight-binding reduced peptide bond inhibitors of a recombinant *Leishmania mexicana* cysteine protease B. J. Med. Chem. 45: 1971-1982.

Stricker, N.L., Christopherson, K.S., Yi, B.A., Schatz, P.J., Raab, R.W., Dawes, G., Bassett, D.E.Jr., Bredt, D.S., and Li, M. 1997. PDZ domain of neuronal nitric acid oxide synthase recognizes novel C-terminal peptide sequences. Nat. Biotechnol. 15: 336-342.

Tao, J., Wendler, P., Connelly, G., Lim, A., Zhang, J., King, M., Li, T., Silverman, J.A., Schimmel, P.R., and Tally, F.P. 2000. Drug target validation: lethal injection blocked by inducible peptide. Proc. Natl. Acad. Sci. USA. 97: 783-786.

Toby, G.G. and Golemis, E.A. 2001. Using the yeast interaction trap and other two-hybrid-based approaches to study protein-protein interactions. Methods. 24: 201-217.

Tong, A., Drees, B., Nardelli, G., Bader, G.D., Brannetti, B., Castagnoli, L., Evangelista, M., Ferracuti, S., Nelson, B., Paoluzi, S., Quondam, M., Zucconi, A., Hogue, C.W., Fields, S., Boone, C., and Cesareni, G. 2002. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science. 295: 321-324.

Toshiyuki Obata, M. B. Y., German G. Leparc, Elizabeth T. Piro, Hiroshi Maegawa, Atsunori Kashiwagi, Ryuichi Kikkawa, and Lewis C. Cantley 2000. Peptide and Protein Library Screening Defines Optimal Substrate Motifs for AKT/PKB. The Journal of Biological Chemistry 275: 36108-36115.

Uetz, P. and Hughes, R.E. 2000. Systematic and large-scale two-hybrid screens. Curr. Opin. Microbiol. 3: 303-308.

Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Phillip, S., Helmer Citterich, M., Cesareni, G., and Dente, L. 2001. Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from Drosophila melanogaster. J. Biol. Chem. 276: 42122-42130.

Vajdos, F.F., Adams, C.W., Breece, T.N., Presta, L.G., de Vos, A.M., and Sidhu, S.S. 2002. Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. J. Mol. Biol. 320: 415-428.

van der Geer, P., and Pawson, T. 1995. The PTB domain: a new protein module implicated in signal transduction. TIBS. 20: 277-280.

Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J., and Salemme, F.R. 1989. Structural origins of high-affinity biotin binding to streptavidin. Science. 243: 85-88.

Weiss, G.A., Watanabe, C.K., Zhong, A., Goddard, A., and Sidhu, S.S. 2000. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. Proc. Natl. Acad. Sci. USA. 97: 8950-8954.

Wenschuh, H., Volkmer-Engert, R., Schmidt, M., Schulz, M., Schneider-Mergener, J., and Reineke, U. 2000. Coherent membrane supports for parallel microsynthesis and screening of bioactive peptides. Biopolymers (Peptide Science). 55: 188-206.

Wilson, D.S., Keefe, A.D., and Szostak, J.W. 2001. The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. USA. 98: 3750-3755.

Yang, J., Cron, P., Good, V.M., Thompson, V., Hemmings, B.A., and Barford, D. 2002. Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. Nat. Struct. Biol. 9: 940-944.

Yeung, Y.A. and Wittrup, K.D. 2002. Quantitative screening of yeast surface-displayed polypeptide libraries by magnetic bead capture. Biotechnol. Prog. 18: 212-220.

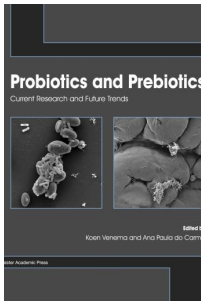
Zhou, J.M., Fujita, S., Warashina, M., Baba, T., and Taira, K. 2002. A novel strategy by the action of ricin that connects phenotype and genotype without loss of the diversity of libraries. J. Am. Chem. Soc. 124: 538-543.

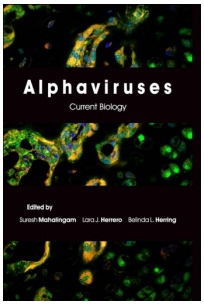
Zucconi, A., Dente, L., Santonico, E., Castagnoli, L., and Cesareni, G. 2001. Selection of ligands by panning of domain libraries displayed on phage lambda reveals new potential partners of synaptojanin 1. J. Mol. Biol. 307: 1329-1339.

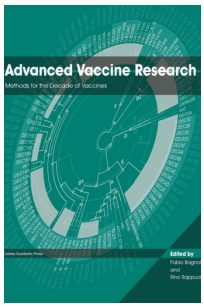
Further Reading

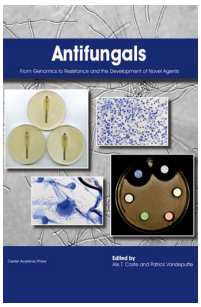
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile









- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order