Published in final edited form as:

Curr Protoc Chem Biol. 2012 December 1; 4(4): 331–355. doi:10.1002/9780470559277.ch120125.

Specificity Profiling of Protein-Binding Domains Using One-Bead-One-Compound Peptide Libraries

Andrew R. Kunys, Wenlong Lian, and Dehua Pei¹

Department of Chemistry and Biochemistry, The Ohio State University

Abstract

One-bead-one-compound (OBOC) libraries consist of structurally related compounds (e.g., peptides) covalently attached to a solid support, with each resin bead carrying a unique compound. OBOC libraries of high structural diversity can be rapidly synthesized and screened without the need of any special equipment and therefore can be employed in any chemical or biochemical laboratory. OBOC peptide libraries have been widely used to map the ligand specificity of proteins, to determine the substrate specificity of enzymes, and to develop inhibitors against macromolecular targets. They have proven particularly useful in profiling the binding specificity of protein modular domains (e.g., SH2 domains, BIR domains, and PDZ domains) and subsequently using the specificity information to predict the protein targets of these domains. The protocols outlined in this article describe the methodologies for synthesizing and screening OBOC peptide libraries against SH2 and PDZ domains and the related data analysis.

Keywords

protein-protein interaction; protein binding domains; sequence specificity; peptide libraries; one-bead-one-compound libraries; SH2 domains

INTRODUCTION

Protein-protein interaction is a fundamental component of life processes. A significant fraction of the interactions are mediated by modular domains that recognize specific short peptide motifs in their partner proteins (Seet et al., 2006). This article describes a general procedure for the synthesis of one-bead-one-compound (OBOC) peptide libraries and their applications in profiling the binding specificities of these protein domains. In this approach, a peptide library containing up to 10^7 random peptide sequences is chemically synthesized on a solid support by the split-and-pool method (Lam et al., 1991), so that each resin bead contains multiple copies of a unique peptide sequence. The resulting peptide library is subsequently screened against a protein domain of interest, which has been chemically or enzymatically labeled with biotin or a fluorescent tag, and positive beads are isolated and sequenced to produce individual binding sequences. Analysis of these sequences provides a consensus sequence(s) of the protein domain. For the purposes of this article the discussion will be limited to libraries of short, linear peptides and their binding to Src homology 2 (SH2) or PDZ domains, although the methodology is readily applicable to cyclic peptides, peptoids, and small molecules (Kodadek and Bachhawat-Sikder, 2006; Liu et al., 2011; Wang et al., 2004). The peptide libraries may be synthesized using either Fmoc- or Boc-based chemistry and several types of commercially available resins have been employed,

¹To whom correspondence should be addressed: Department of Chemistry and Biochemistry, The Ohio State University, 100 West 18th Avenue, Columbus, OH 43210. Phone: 614-688-4068; pei.3@osu.edu.

each with some advantages and disadvantages. The protein domains are produced recombinantly as a glutathione-S-transferase (GST) fusion or with a 12-amino acid ybbR tag (Yin et al., 2005). The GST or ybbR tag facilitates protein purification and/or chemically/enzymatically labeling of the protein with biotin or a fluorophore. Library screening is accomplished by incubating the peptide library with a small amount of the target protein, and positive beads are identified through an enzyme-linked assay or by their fluorescence signal. The hit beads are manually isolated with the aid of a micropipette and individually sequenced by partial Edman degradation-mass spectrometry (PED-MS) (Thakkar et al., 2006).

Although several other methods including phage display (Rickles et al., 1995; Tonikian et al., 2008), oriented peptide libraries (Songyang et al., 1993), and peptide arrays (Hoffmuller et al., 1999; Rodriguez et al., 2004) have also been used to profile protein domains, the OBOC library approach offers some distinct advantages. First, synthetic libraries permit the incorporation of posttranslationally modified amino acids; this is important because many protein domains require posttranslationally modified residues for binding [(e.g., SH2 domains recognize phosphotyrosine (pY) (Sadowski et al., 1986) whereas BRCT domains require phosphoserine (Yu et al., 2003)]. In contrast, biological methods such as phage display are generally incompatible with such modified residues. Second, the OBOC method gives individual binding peptide sequences. This is especially critical when a protein domain recognizes more than one class of peptide sequences, which appears to be quite common among the SH2 domains. For example, the N-terminal SH2 domain of the phosphatase SHP-2 can bind to four structurally distinct classes of pY peptides, as revealed by screening of an OBOC pY peptide library (Sweeney et al., 2005).

STRATEGIC PLANNING

Peptide Library Design

Designing a proper library for the target protein domain is one of the most crucial aspects in planning this experiment. Although in principle one can screen all protein domains against a generic library that contains a long stretch of random amino acids and all of the posttranslationally modified amino acids commonly found in proteins, in practice the synthesis and screening of such a large library would be prohibitively expensive and time consuming. Therefore, a more fruitful approach is to “custom” design a peptide library for each class of protein domains. One should gather as much information as possible on the binding properties of the protein domain such as the requirement of modified amino acids, the minimal peptide length, the position of modified amino acid(s) in the binding motif, and whether the interaction requires a free N- or C-terminus. For example, SH2 domains require a pY for binding and generally two to three residues immediately N- and C-terminal to pY, respectively (Songyang et al., 1993; Waksman et al., 1992), whereas PDZ domains require a free C-terminus (Manke et al., 2003; Rodriguez et al., 2003; Zhang et al., 1998). Thus, a proper library for the SH2 domains may take the form of XXpYXXXX, while that for a PDZ domain may be XXXXX-CO₂H (where X is any of the 20 proteinogenic amino acids).

A second consideration is the type of resin onto which the library is to be synthesized. Two different types of resins are commonly used in our laboratory, although other resins may also be used. Polyethyleneglycol-acrylamide (PEGA) resin is a co-polymer of acrylamide and polyethyleneglycol. The resin is hydrophilic, has large pore sizes, and is very flexible. The large pore sizes allow even large protein molecules (up to 70 kDa) to diffuse into the bead. These properties are crucial for on-bead screening of enzyme substrates (Ren et al., 2011). We have also found that PEGA resin is well suited for screening peptide libraries with fluorescently labeled proteins, because it has essentially no background fluorescence and its large pore size allows the labeled protein to penetrate into the bead interior thus

producing stronger fluorescent signals on a positive bead as compared to other resins (e.g., TentaGel). However, PEGA resin has low mechanical strength and is easy to break. It also sticks to many surfaces and thus extra care must be taken in handling the resin. PEGA is not compatible with the popular enzyme-linked assay involving alkaline phosphatase and 5-bromo-4-chloro-3-indolyl phosphate (BCIP), because the indigo dye formed does not stay on the hydrophilic bead. The second type of resin is TentaGel, a polystyrene polymer grafted with poly(ethyleneglycol). TentaGel resin is more rigid and has much greater mechanical strength than PEGA resin. It has excellent swelling properties in a wide range of organic and aqueous solutions, greatly facilitating both library synthesis in organic solvents and screening in aqueous buffers. TentaGel resin is ideally suited for enzyme-linked screening using the phosphatase/BCIP method, because the dye binds strongly to the polystyrene core. The mechanical strength of TentaGel resin also allows for spatial segregation of each resin bead into surface versus inner layers (Liu et al., 2002). This feature has been utilized to prepare peptide libraries of reduced ligand loading on the bead surface, thus increasing the stringency of the screening as well as reducing the amount of nonspecific binding caused by avidity effects (Chen et al., 2009; Wang et al., 2005). Joo and Pei (Joo and Pei, 2008) have also used this strategy to synthesize inverted peptides containing free C-terminus for screening against PDZ domains, while the normally-oriented peptides inside the beads served as encoding tags for peptide identification (See Alternative Protocol: **Synthesis of an Inverted Peptide Library**). However, TentaGel resin has small pore sizes and is essentially impermeable to proteins of any size. This property, together with its high autofluorescence signals, renders fluorescence-based screening on TentaGel resin somewhat challenging, although successful examples have been reported (Kritzer et al., 2005; Olivos et al., 2003). We and others have also found that TentaGel resin is generally unsuitable for on-bead screening of enzymatic reactions, probably because the enzymes can only access the substrates within a very thin layer on the bead surface and do not produce sufficient signal to render the bead positive.

A suitable linker between the random peptide sequence and the resin is also important for a successful screening experiment, peptide release, and subsequent sequence analysis by PED-MS. In our libraries, we have frequently used the hexapeptide LNBBRM as the linker (where B is β -alanine). The C-terminal methionine allows for the release of peptides from resin by treatment with cyanogen bromide (CNBr). The arginine residue increases the peptide solubility in aqueous solutions and provides a fixed positive charge to facilitate MS analysis. The two β -alanines add flexibility to the peptides and also distance between the solid phase and the random positions. The last two residues (LN) are added to increase the masses of the peptides to ~600 Da so that the peptide peaks do not overlap with those of matrix material during PED-MS analysis.

With the above considerations, a suitable library for screening against SH2 Domains is AXXpYXXXXLNBBRM-TentaGel, where X represents any of the 18 proteinogenic amino acids (except for cysteine and methionine) and norleucine (used as a replacement of methionine). Methionine is left out of the random sequence to avoid internal cleavage during peptide release from the resin, whereas cysteine is underrepresented among protein-protein interactions and may complicate peptide library synthesis and/or screening due to its ability to form disulfide bonds. We routinely synthesize the library on 5 g of 90- μ m TentaGel beads, each of which carries ~100 pmol of a unique peptide, resulting in a theoretical library diversity of 4.8×10^7 . Since 1 g of the above resin has $\sim 2.86 \times 10^6$ beads, 5 g of the library can cover $\sim 1.43 \times 10^7$ different sequences, or 30% of the theoretical sequence space. For a typical SH2 domain, we have found that screening against 100–200 mg of the library, which corresponds 0.6–1.2% of all possible sequences, is usually sufficient to generate an unambiguous consensus sequence(s). This is because an SH2 domain does not make specific interactions with all of the residues in the random region. For example, the SH2 domain of

Grb2 protein only requires the pY residue and an Asn at the pY+2 position (relative to pY, which is designated as position 0) for binding, although residues at the pY+1 and pY+3 positions also make minor contributions to the overall binding affinity (Kessels et al., 2002; Songyang et al., 1993).

For PDZ and other domains that require a free C-terminus for binding, peptide libraries must be synthesized on spatially segregated TentaGel beads (Figure 1) (Joo and Pei, 2008). The peptides are synthesized on the beads in the conventional C→N manner, with their C-termini attached to the support through an ester linkage on the bead surface but through an amide bond in the bead interior. The surface peptides are cyclized between their N-terminal amine and a carboxyl group installed at the C-terminal linker sequence, while the internal peptides are kept in the linear form. Base hydrolysis of the ester linkage in the cyclic peptides regenerates linear peptides that contain a free α-carboxyl group at their C-termini but remain covalently attached to the resin via the N-termini (“inverted” peptides). A suitable peptide library for PDZ domains may have the form of TentaGel-MLLBBEAAXXXXXX-CO₂H (Figure 1a), where B is β-alanine and X represents L-α-aminobutyrate (Abu, used as a replacement of cysteine), L-norleucine (Nle, as a replacement of methionine), or any of the 18 proteinogenic amino acids except for cysteine and methionine.

On-bead screening of OBOC libraries is sometimes plagued by false positives caused by nonspecific interactions, herein defined as any interaction other than the intended protein-peptide interaction. The TentaGel resin commonly used for OBOC libraries has a loading capacity of ~0.3 mmol/g, which corresponds to a ligand “concentration” of ~100 mM on the resin. This very high ligand density permits a protein domain to simultaneously interact with multiple ligands; as a result of the avidity effect, a bead carrying a weak binding ligand may be observed as a false positive (Chen et al., 2009). For example, screening against a protein of pI <7 may result in sequences dominated by Arg and Lys residues which, upon resynthesis and testing in solution, show little affinity to the protein domain. Presumably, the negatively charged protein engages in electrostatic interactions with several positively charged peptides on the bead surface. To overcome this problem, Chen et al. (2009) have synthesized spatially segregated OBOC libraries that contain a reduced ligand loading at the bead surface but a normal loading inside the bead (in order to have enough material for later sequencing analysis) (Figure 2). It was found that a 10-fold reduction in surface ligand density effectively eliminated most of the false positives and screening biases. We thus recommend that all protein domains be screened against reduced-density libraries. It should be noted that spatial segregation is most effective with resins of considerable mechanical strength and rigidity, such as the TentaGel. PEGA beads may be segregated, but a protein will have access to peptide ligands at both the bead surface and interior.

Protein Labeling

Screening of OBOC libraries against protein binding domains requires labeling of the protein with biotin, a fluorescent group, or some other tag for visualization. Two different labeling methods, one chemical and the other enzymatic, have been used in our laboratory. To chemically label a protein, one can simply mix the protein with 2–3 equivalents of the N-hydroxysuccinimide (NHS) ester of the tag of interest (e.g., NHS-biotin or NHS-Texas Red) in a NaHCO₃ buffer (pH 8). This results in selectively labeling of a Lys residue(s) on the protein. To minimize any interference with the binding property of the protein domain, it is advisable to express the protein as a GST fusion. The GST tag not only improves expression and solubility and facilitates purification, but also provides 20 additional lysine residues to which the dye molecule or biotin may be attached. If the protein of interest contains a Lys residue at the ligand binding site or for various reasons the experiment demands the expression of the protein as an isolated domain, enzymatic labeling is the method of choice.

The protein of interest is expressed with a short peptide sequence, DSLEFIASKLA (ybbR tag), at its N- or C-terminus. The ybbR tag is specifically recognized by the bacterial phosphopantetheinyl transferase Sfp (Yin et al., 2005). Treatment of the ybbR-tagged protein with Sfp in the presence of a coenzyme A-biotin (or coenzyme A-dye) conjugate results in selective labeling of the protein at the underlined serine residue.

BASIC PROTOCOL 1

Split-and-Pool Synthesis of One-Bead-One-Compound libraries

Each amino acid reacts at a slightly different rate during peptide coupling reactions. Although all of the naturally occurring amino acids and many unnatural ones can be coupled to completion, some modified amino acids such as pY cannot. Because the library against SH2 domains contains a pY, it is necessary to cap any unreacted N-termini after the addition of pY. The use of a more powerful coupling reagent (HATU instead of HBTU) is strongly advised for the coupling of pY and all subsequent residues. Library synthesis can be carried out in custom designed glass vessels or disposable plastic columns from Bio-Rad. This protocol below is designed for TentaGel resin, though it also works well with PEGA resin. Note that the PEGA resin should never be brought to complete dryness (air dry for more than 10 s or desiccation). Rather, it can be stored in DMF or other appropriate solvents at -20°C .

Materials

Chemglass* Solid Phase Peptide Synthesis Vessel with Inner Luer Joint > 25 mL; Disc O.D: 20 mm; GL size: 25 (CG186202, Chemglass No.:CG-1862-02)

Part A Materials

N'-N'-DimethylFormamide (Sigma Aldrich)
Dichloromethane (Sigma Aldrich)
Bio-Spin Disposable Chromatography Columns (BioRad cat # 732-6008)
TentaGel S NH₂ resin (90 μm , 0.3 mmol/g)
O-Benzotriazole-N,N',N'-tetramethyluronium hexafluorophosphate (HBTU) (Aapptech)
N-Hydroxybenzotriazole (HOBt) (Aapptech)
Diisopropylethylamine (DIPEA)
DCM/Ethyl Ether (55:45 v/v)
Fmoc-Met-OSu (Aapptech)
Fmoc-Met-OH (Aapptech)
Boc-Phe-OSu (Aapptech)
50% (v/v) trifluoroacetic acid (TFA)/DCM
2% acetaldehyde in DMF and 50 μL of 2% chloranil in DMF

Part B Materials:

All Fmoc-Amino Acids desired with acid-labile sidechain protecting groups (tBu, Boc etc) (Aapptech)
HBTU, HOBt, DIPEA (see above)

20% Piperidine in DMF

5 g ninhydrin in 100 mL of ethanol, 80 g phenol in 20 mL of ethanol, 1 mM KCN in 98% pyridine

Part C Materials:

All materials from previous section with the addition of

CD₃CO₂D (Sigma-Aldrich)

CH₃CD₂CO₂D (Sigma-Aldrich)

Part D Materials:

Fmoc-pY-OH (Aapptech)

HATU, DIPEA (Aapptech)

Acetic anhydride (Sigma Aldrich)

Part E Materials:

Modified reagent K: (6.5% phenol, 5% water, 5% thioanisole, 2.5% ethanedithiol, 1% anisole, and 1% Triisopropylsilane in TFA)

20% piperidine in DMF

A. Bead Segregation and Surface Density Reduction

This protocol decreases the peptide loading in the surface layer of each bead and greatly reduces the number of false positive beads due to nonspecific binding during library screening.

- 1 Soak the resin (5 g) in water overnight. In the morning, drain the resin and QUICKLY resuspend it in 30 mL of 55:45 (v/v) DCM/diethyl ether containing Fmoc-Met-OSu (0.10 equiv) and Boc-Phe-OSu (0.40 equiv). Let the coupling reaction proceed on a rotary shaker for 30 minutes. This will result in selective acylation of the surface amines, ~90% of which will be capped with Boc-Phe and the other ~10% will be coupled with Fmoc-methionine (the first residue).
- 2 Step 1 may fail at the first attempt. Therefore, it is important to test the quality of segregation/reduction by chloranil test before proceeding to the next step. Take 20–50 beads with the tip of a spatula and shake them into an Eppendorf tube. Into the tube, add 50 µL of 2% acetaldehyde in DMF and 50 µL of 2% chloranil in DMF. Mix the solutions and let the beads sit for 5 minutes at room temperature, examine the beads on a watchglass. If segregation was successful, then the inside of the beads should be green or blue and the outside should be clear (Figure 1b).
- 3 Wash the library resin with 55:45 DCM/diethyl ether (3 × 30 mL) and DMF (8 × 30 mL). Then treat the resin with 4 equiv of Fmoc-Met-OH and HBTU/HOBt/DIPEA in DMF (90 min). This will add a methionine to the remaining amines in the bead interior.
- 4 Treat the resin with 25 mL of 50% (v/v) trifluoroacetic acid (TFA)/DCM for 30 min to remove the Boc group from the Phe. Wash several times with DCM and DMF and then add 5% (v/v) DIPEA/DCM to neutralize the TFA. Finally, add excess acetic anhydride (2.5 g) in 25

mL of DMF and DIPEA. This will cap the N-terminus of the phenylalanine ensuring that there will be no further peptide coupling.

- 5 Remove the Fmoc group by treating twice with 30 mL of 20% piperidine in DMF (once for 5 min and again for 15 min), and wash with DMF (6×30 mL). At this point, the resin is ready for library synthesis.

B. Synthesis of the Linker Region

Synthesis of the linker sequence (LNBBRM) follows standard Fmoc chemistry and the basic protocol is described below. This protocol is for the addition of the arginine residue and must be repeated for the other residues in the linker.

- 6 Wash the resin with DMF 3 times and let it soak in 25 mL of DMF for 15 min (the soaking step may be skipped if the resin has already been resuspended in DMF after a coupling cycle).
- 7 Dissolve 4 equivalents of Fmoc-Arg(pbf)-OH (6.49 g), HBTU (3.79 g), and HOBt (1.35 g) in 25 mL of DMF. Add 8 equivalents of DIPEA (1.74 mL) and incubate the mixture on a rotary shaker for 2 min.
- 8 Add the mixture (which may have a slight color) to the resin and allow the reaction to proceed on a rotary shaker for 1.5 h. It is important that the beads do not clump together in the containers. As such, it is sometimes useful to manually mix the resin for a few minutes until the beads are evenly suspended.
- 9 Drain the vessel on a vacuum system and wash with DMF 5 times and then 5 times with DCM.
- 10 Perform a ninhydrin test on the resin. Take 20–50 beads with the tip of a spatula and shake them into a test tube. Into the tube, add 10 μ L of ninhydrin solution (5 g ninhydrin in 100 mL ethanol), 10 μ L of phenol solution (80 g phenol in 20 mL ethanol) and 10 μ L of potassium cyanide solution (1 mM KCN in 98% pyridine). Heat the solution for 5 min. The test should be negative (the solution should be yellow). If coupling was not complete (solution turns purple), repeat steps 2 – 5.
- 11 Remove the Fmoc group by treating twice with 25 mL of 20% piperidine in DMF (once for 5 min and again for 15 min), and wash 3 times with DMF.
- 12 Perform a ninhydrin test as above. This time the test should be positive (solution turns purple).. This step can usually be skipped, but is included here to ensure that the N-terminus is not capped and that the piperidine solution is good. If the test is negative, make fresh piperidine solution and redo step 5.
- 13 Repeat steps 1–6 for the rest of the amino acids in the linker sequence.
- 14* If you need to stop the experiment for an extended period of time (e.g., overnight at the end of the day), try to end after coupling an amino acid so the N-terminus is Fmoc protected. Wash the resin with 5 times with DCM and drain until dry and then for 1 additional minute. Place the vessel upright in a desiccator at room temperature for storage overnight (for long term storage, beads should be dried and stored at -20°C). We recommend taking extra care to clearly record the step at which the

experiment is stopped. Beads should be reswelled in DMF (step 1) before continuing.

C. Synthesis of the Random Positions:

Some amino acids (e.g., Leu, Ile, and Nle) have the same molecular mass (isobaric) and cannot be distinguished by MS. This problem has been resolved by employing a capping strategy (Thakkar et al., 2006). Briefly, small amounts of different acylating agents (5%) are added to the coupling reactions of isobaric amino acids during library synthesis, resulting in an extra peak for each isobaric residue during PED-MS analysis. We typically add 5% (mol/mol) $\text{CD}_3\text{CO}_2\text{D}$ to the coupling reaction vessels for leucine and lysine and 5% $\text{CH}_3\text{CD}_2\text{CO}_2\text{D}$ to that of norleucine. As a result, Ile, Leu, and Nle are distinguished by having a singlet peak, a doublet separated by 45 mass units, and a doublet separated by 58 mass units, respectively. Lys and Gln are differentiated in the same manner (addition of 5% $\text{CD}_3\text{CO}_2\text{D}$ to the Lys reaction). During the addition of the random residues, ninhydrin tests are generally not performed; instead, each coupling reaction is repeated (performed twice in succession) to ensure complete coupling.

- 15 For each random position, split the resin into 19 (or 20) equal portions and place them in 19 (or 20) different reaction vessels, one for each amino acid to be added at the next random position. Although one can dry the resin and split it by weight, a more convenient method is to split it volumetrically. Suspend the resin in ~25 mL of DMF in a large vessel and pipet 19 1-mL aliquots into separate reaction vessels. Drain all of the vessels including the original large one, and repeat the above procedure in the reverse direction (from vessel #19 to vessel #1). Repeat this one more time in the forward direction and repeat it if needed.
- 16 At this point, examine the 19 reaction vessels to ensure that they all have roughly equal amounts of resin. If not, balance them by transferring resin in between them (by pipetting).
- 17 Weigh out each amino acid to be coupled into clean Falcon tubes (not the reaction vessels) and use the same ratios for HBTU (or HATU) and HOBT (or HOAt) as described above (in step 2 of the linker heading). Prior to the addition of DIPEA, add 5% (mol/mol) $\text{CD}_3\text{CO}_2\text{D}$ to the vessels for leucine and lysine, and 5% (mol/mol) $\text{CH}_3\text{CD}_2\text{CO}_2\text{D}$ to the norleucine vessel. Add DIPEA and mix the contents for 1–2 min, and then add the mixture to the resin.
- 18 Allow the reaction to proceed for 1.5 h, wash the resin 5 times with DMF, and then repeat step 3.
- 19 Combine the resin from all 19 reaction vessels into the larger vessel. Transfer any resin remaining in the reaction vessels by suspension in DMF and pipetting. Remove the N-terminal Fmoc group by treatment twice with 25 mL of 20% piperidine in DMF (once for 5 min and again for 15 min). Wash the resin 3 times with DMF.
- 20 Repeat the above steps until all random positions are coupled.

D. Coupling of Phosphotyrosine

As mentioned above, coupling of phosphoamino acids such as pY is more difficult than the 20 proteinogenic amino acids. This problem is somewhat worsened because only 2.5 equivalents of Fmoc-pY are used in the reaction due to its high

costs. We recommend the use of HATU as the coupling agent, which is more powerful than HBTU. Obviously, if a library contains random residues N-terminal to pY, protocol 1C must be repeated following the addition of pY.

- 21 Dissolve 2.5 equivalents of Fmoc-pY-OH and 2.5 equivalents of HATU in 25 mL of DMF. Mix the solution with the resin and add 5 equivalents of DIPEA. Allow the reaction to proceed for 3 h. Drain the reaction solution and repeat the reaction.
- 22 Since the above reaction may not be complete, any remaining N-terminal amine needs to be capped. To do this, add 10 equivalents of acetic anhydride and 10 equivalents of DIPEA in 25 mL of DMF to the resin and incubate for 30 min.
- 23 Remove the N-terminal Fmoc group with 25 mL of 20% piperidine in DMF (once for 5 min and again for 15 min), and wash 3 times with DMF.
- 24 Continue the library synthesis as described above. We recommend the use of HATU/HOAt for all residues after the pY and 4–5 equivalents of Fmoc-amino acids.

E. Side-Chain Deprotection and Storage of Peptide Library

After the library synthesis is complete, we recommend storing it in the fully protected form, desiccated, and at -20°C . Peptide libraries stored under these conditions are stable for many years. As such, one should only deprotect a portion of the library on as-needed basis. Prior to screening, the library needs to be deprotected, as follows.

- 25 Treat the resin (1 g) with 20% piperidine in DMF (once for 5 min and again for 15 min) to remove the N-terminal Fmoc group. Wash 3 times with DMF and then 5 times with DCM.
- 26 Add 10 mL of a modified reagent K to the library resin and incubate for 2 h.
- 27 Wash the resin with TFA and DCM and dry under vacuum. Store resin dry at -20°C .

Alternative Protocol: Synthesis of Inverted Peptide Library

The protocols described in this unit can easily be modified to synthesize peptide libraries of more complex structures such as those containing free C-termini (inverted peptides) for screening against PDZ domains (Joo and Pei, 2008). As noted above, the key to this procedure is the spatial segregation of the resin beads during the coupling of Fmoc-Glu-O(allyl). Next, a hydroxymethylbenzoic acid (HMBA) linker is added followed by the random amino acids (Figure 1a). Note that the first random residue must be coupled to HMBA using DCC as the coupling reagent. After the addition of all random residues and dipeptide AA (as a linker), the C-terminal glutamic acid allyl ester is deprotected with a palladium catalyst and cyclized with the N-terminal amine with PyBOP. Finally, hydrolysis of the HMBA ester linkage with 1 M NaOH regenerates linear peptides that remain attached to the resin through their N-termini and display free C-termini.

Additional Materials:

Hydroxymethylbenzoic Acid (Sigma Aldrich)

N,N'-Dicyclohexylcarbodiimide (DCC) (Sigma Aldrich)

Fmoc-Glu(OAllyl) (Aaptech)

Palladium tetrakis(triphenyl) phosphine (Sigma Aldrich)

benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP)
(Aaptech)

Basic Protocol 2

Protein Labeling

For chemical labeling of proteins with NHS esters, it is imperative that the protein solution be free of buffer materials that contain primary amines (e.g., Tris and GSH, which is commonly used to elute GST fusion proteins). One should also be mindful that the protein may be unstable under the labeling conditions or after being labeled. Further, if a fluorescence based screen is planned, any unreacted dye should be removed (e.g., by passing the protein solution through a Sephadex G-25 column).

Reagents and Solutions

Bradford Assay (BioRad)

Amicon Ultracentrifugation unit (Millipore)

Bio-Spin Disposable Chromatography Columns (BioRad cat # 732-6008)

Protein Purification buffer: 50 mM HEPES, 100 mM NaCl. Adjust pH to 7.5 with HCl or NaOH

Bicarbonate buffer: 100 mM sodium bicarbonate, 100 mM NaCl

Stock NHS-Biotin solution: Dissolve NHS-biotin in DMSO for a final concentration of 10 mg/mL. Store the solution in small aliquots at -20 or -80 °C.

Chemical Labeling with N-Hydroxysuccinimide Ester

1. After a protein of interest is purified, determine its concentration using a Bradford or BCA assay. If the protein concentration is substantially below 1.5 mg/mL, concentrate the solution in an Amicon ultracentrifugation unit until the concentration reaches 1.5 mg/mL. High labeling efficiency is generally achieved when the protein concentration is ~ 2 mg/mL. We do not recommend attempting the labeling reaction at protein concentrations much higher than 2 mg/mL, as one runs significant risk of protein precipitation.
2. A typical screening experiment involves 100 mg of library resin and 10 mL of 250 nM protein solution, which requires ~ 2.5 nmol of protein. We thus typically label 50–100 nmol of protein, which is enough for multiple screens.
3. To the appropriate volume of protein solution, add about 0.1 volume of 100 mM sodium bicarbonate solution. Check the solution with pH paper to make sure that the pH is between 8 and 9. Add more sodium bicarbonate solution if needed. Add 2.5 equivalents of NHS-biotin (10 mg/mL in DMSO) and let the reaction proceed at room temperature for 30 min. If protein precipitation occurs, perform the reaction at 4 °C overnight. Be sure that the amount of DMSO does not exceed 3–5%.
4. After the reaction is complete, remove any free biotin (or dye) by three repetitions of concentration and dilution in an Amicon concentrator. Alternatively, the free biotin can be removed by passing the protein solution through a Sephadex G-25 column.

5. At this point, the protein solution is ready for use in the screening experiment and should be stored frozen in the presence of 30% glycerol at -80°C or 50% glycerol at -20°C .

Support Protocol

Enzymatic Labeling of Proteins

This protocol should be used if the chemical labeling method described above failed to give satisfactory screening results (e.g., chemical labeling results in the modification of a residue at or near the ligand binding site). The target protein must be expressed with a ybbR tag at the N- or C-terminus. This method allows specific labeling of the target protein at a single site (i.e., on the ybbR tag). The protein of interest does not need to be pure, as long as the impurities do not contain a sequence motif that is recognized as a substrate by the Sfp enzyme (Sfp has very stringent substrate specificity, and so far we have not observed any *E. coli* protein being labeled by the enzyme). Protein labeling with the Sfp enzyme may proceed at essentially any protein concentration, although we generally use between 50 and 500 μM protein. However, we have noted that Sfp labeling may cause some proteins to precipitate, the reason of which is yet unclear

1. To a protein solution, add 0.1 volume of 10 $\times$ Sfp buffer. Next add 5 equivalents of the coenzyme A-biotin (or dye) adduct (relative to the protein of interest) and 0.1 equivalents of the purified Sfp enzyme. The reaction should be complete in ~ 30 min at room temperature.
2. Remove any unreacted CoA-biotin (or CoA-dye) adduct on a Sephadex G-25 column. After this point, the protein is ready for screening and should be stocked in 30% glycerol at -80°C or 50% glycerol at -20°C .

Reagents and Solutions

10x Sfp Reaction Buffer: 500 mM HEPES, 1 M NaCl, and 100 mM MgCl_2 , pH 7.4

Coenzyme A-Biotin: Dissolve biotin-CoA in DMSO to a final concentration of 10 mg/mL and store it in small aliquots at -20 or -80°C .

Basic Protocol 3

On-Bead Screening of OBOC Libraries

For peptide libraries on TentaGel resin, the streptavidin-alkaline phosphatase (SAAP)/BCIP method is the method of choice for profiling the specificity of protein domains. Indeed, most of the SH2 domain specificity data in our laboratory were obtained using this method. Briefly, binding of a biotinylated protein domain to a bead recruits SAAP to the bead surface (Figure 2). Upon the addition of BCIP, the bead associated alkaline phosphatase dephosphorylates BCIP to produce an indole, which subsequently oxidizes in air and dimerizes to form indigo, a turquoise colored compound that binds tightly to the hydrophobic polystyrene core. As a result, the positive bead becomes turquoise colored and readily isolated with a micropipet under a dissecting microscope. The protocol below has been optimized for SH2 domains, but has worked well with other protein domains. The optimal protein concentration in the screening reaction is usually determined by trial and error.

Materials

Library resin (synthesized in step 1)

Micro-BioSpin columns (0.8 mL) are available from Bio-Rad (Hercules, CA).

DMF

DCM

SAAP is available from Prozyme (San Leandro, CA).

HBST buffer (30 mM HEPES, pH 7.4, 150 mM NaCl, 0.01% Tween 20, .01% gelatin, 10 mM sodium azide))

SAAP buffer (30 mM TRIS, pH 7.6, 1 M NaCl, and 20 mM K₂PO₄)

Staining buffer (30 mM TRIS, pH 8.5, 100 mM NaCl, 5 mM MgCl₂, and 20 mM ZnCl₂)

5 mg/mL BCIP in water (Sigma–Aldrich)

1 µg/µL SAAP (SAAP (Prozyme, San Leandro, CA)

Dissection microscope (10–40× magnification)

35-mm Petri dish (non-sterile, non-culture treated polystyrene dishes work fine)

1. Colorimetric Screening with Biotinylated Protein

Day 1:

It should be noted that multiple trials may have to be performed before the optimal conditions (protein concentration and washing conditions) can be found. These pilot experiments are typically carried out with ~10 mg of library resin. Once the proper conditions are found, one or two large scale experiments (50–100 mg resin) are then performed.

- 1 Weigh out 10 mg of library resin and place it in a *micro-BioSpin column (0.8 mL)*. Wash 3× with 1 mL of DMF, 3× with 1 mL of DCM, and ×x again with 1 mL of DMF.
- 2 Wash the resin with HBST blocking buffer, then suspend the resin in the same buffer and allow it to incubate for at least 1 h at 4 °C on a rotary shaker.
- 3 Drain the solution and suspend the library beads in 1 mL of fresh HBST blocking buffer containing varying concentrations of the biotinylated protein (typically 50–500 nM). Allow the protein to incubate with the resin at 4 °C for 4–16 h on a rotary shaker.

Day 2:

In addition to the protein concentration, the proper washing procedure is crucial for a successful screening experiment. While inadequate washing may result in false positives, excessive washing may result in false negatives. Thus, one must find a delicate balance. The number of positive beads depends on the affinity and specificity of the protein domain as well as the stringency of screening conditions (e.g., protein concentration, number of washings, and length of staining time). However, the screening experiments are generally very reproducible when multiple screenings are performed under the same conditions. In all screening experiments, a negative control with biotinylated GST should be included (it should yield no colored beads under identical conditions).

- 4 Drain the solution, and wash the resin very quickly with 1 mL of SAAP buffer. Resuspend the resin in 1 mL of SAAP buffer

and add 1 μ L of SAAP solution (1 mg/mL). Incubate the mixture for 10 min at 4 °C with mixing.

- 5 Drain the buffer solution, and quickly wash the resin with 1 mL of the blocking buffer (2 $\times$). Wash once with the reaction buffer and then resuspend the resin in the reaction buffer.

Note: If the experiment produces too many blue beads or has strong background color, more stringent washing with the blocking buffer should be performed at this stage in the next experiment.

- 6 Transfer the beads into a 35-mm petri dish. Add 0.1 volume of 5 mg/mL BCIP and place the dish on a slow shaker. Check the development of colored beads periodically under a microscope. For a typical SH2 domain, 10 strongly colored beads from 10 mg of resin (or $\sim$ 1/3000 beads) is indicative of a proper stringency level. Color development typically takes 30–45 min, but may take up to 2 h for some protein domains.

- 7 Quench the staining reaction by adding 1 mL of 1 M HCl. Manually remove all of the colored beads from the petri dish using a micropipet under a dissecting microscope and place them into another petri dish filled with water. Sort the beads into different categories according to their color intensities, placing all beads of the same color category into a microcentrifuge tube. Discard any broken beads or beads with color on only part of their surfaces. We typically separate positive beads into “intensely”, “medium”, and “lightly” colored categories.

Alternative Protocol: On-Bead Screening against Fluorescently Labeled Protein

Fluorescence-based screening requires that the protein of interest be labeled with a fluorophore and is best carried out with libraries synthesized on PEGA resin (the protocol below is designed for PEGA resin). The main advantage of this method is the simplicity of the screening process. One can simply add increasing concentrations of the labeled protein to the library and monitor the increase in fluorescence on positive beads. Since the color intensity on a bead is in equilibrium with the protein concentration in the screening solution, the intensity should directly reflect the binding affinity (avidity) of the ligand on the bead. This eliminates the need for washing steps and therefore the influence of binding kinetics on the color intensity of a bead. The disadvantage of this method is that the manual bead picking process under a fluorescence microscope is rather taxing, and it is more difficult to sort the positive beads into different categories based on the fluorescence intensity.

Materials:

Micro-BioSpin columns (0.8 mL) are available from Bio-Rad (Hercules, CA).

DMF, DCM

HBST buffer (see above)

1. Weigh out $\sim$ 10 mg of library resin. Because the library resin should be kept swollen in DMF all time, a convenient way to do this is to transfer the resin suspended in DMF into a Micro-spin column, drain the resin by gravity, and

weigh the amount of wet resin inside the column. Each mg of dry PEGA resin swells into ~14 mg of wet resin under these conditions.

2. Wash the resin several times with DMF, then several times with HBST buffer. Allow the resin to incubate in the blocking buffer for 1 h at 4 °C on a rotary shaker.
3. Add fluorescently labeled protein, incubate at 4 °C in the dark for 2 h on a rotary shaker, and examine the fluorescence intensity on beads under a microscope. It is a good practice to start with a low protein concentration (e.g., 10 nM) and gradually increase the concentration, until a desired number of positive beads (e.g., 1/3000 beads for SH2 domains) become clearly fluorescent.
4. Isolate the positive beads with a micropipette (they should have a fluorescent ring around them) and place them into a clean Eppendorf tube. Some beads may clump together. Be sure that only the positive beads are isolated.

Basic Protocol: 4

Partial Edman Degradation

PED-MS provides a high-throughput, reliable, and inexpensive method for sequencing support-bound peptides and was specifically developed for identification of hits derived from OBOC libraries (Thakkar et al., 2006). During a partial Edman degradation reaction, a bead is treated with a mixture of Edman's reagent, phenylisothiocyanate (PITC), and a small amount of fluorenylmethoxycarbonyl-N-hydroxysuccinimidyl ester (Fmoc-OSu). Both reagents will compete for reaction with the peptide N-terminal amine (Figure 4a). For 90–95% of the peptides that have reacted with PITC, subsequent treatment with TFA results in the cleavage of the N-terminal residue and a shortened peptide. For the remaining 5–10% of the peptides, which have reacted with Fmoc-OSu, their N-termini are blocked with an Fmoc group and TFA treatment has no effect and thus the retention of the N-terminal residue. Repetition of the partial Edman reaction generates a peptide “ladder”, which can be analyzed by MALDI-TOF MS (Figure 4b). By calculating the mass differences between adjacent peaks, one can unambiguously determine the sequence of the original full-length peptide.

Materials

Reaction vessel (12-mm diameter, 20 mm height, a 10–20 µm glass frit, and 1 mm Luer tip at the bottom)

PITC can be purchased in 1-mL sealed ampoules from Sigma–Aldrich.

4-Hydroxy- α -cyanocinnamic (Sigma–Aldrich.)

2:1 (v/v) pyridine/water solution containing 0.1% triethylamine

Trifluoroacetic Acid (Sigma Aldrich)

Fmoc-OSu solution (8.8–11.0 mM) in pyridine

Cyanogen bromide (40 mg/mL) in 70% TFA

Ammonium iodide (Sigma Aldrich)

Dimethyl sulfide (Sigma Aldrich)

Procedure

1. Place positive beads of each color category (intense, medium, or light etc.) into a custom designed glass reaction vessel (12-mm diameter, 20 mm height, a 10–20 μ m glass frit, and 1 mm Luer tip at the bottom). “Wash” in the following steps means filling the vessel completely with solution and then draining the vessel through its frit under vacuum.
2. Wash the beads several times with water, pyridine, and a 2:1 (v/v) pyridine/water solution containing 0.1% triethylamine. Dry the beads by vacuum suction of the reaction vessel through the Luer tip.
3. For each vessel, mix 148 μ L of the Fmoc-OSu solution with 12 μ L of PITC by pipetting. Add 160 μ L of the 2:1 (v/v) pyridine/water solution containing 0.1% triethylamine. Quickly mix the contents and transfer the solution into the reaction vessel (with beads).
4. Let the reaction proceed for 6 min and then wash once with pyridine and 3 times with DCM.
5. Wash once with anhydrous TFA, add in 500 μ L of TFA with a Pasteur pipette, and allow the reaction to proceed for 6 min. Drain the resin and repeat this step.
6. Wash 3 times with DCM, once with pyridine, and then with a 2:1 (v/v) pyridine/water solution containing 0.1% triethylamine. Drain the solution to dryness.
7. Repeat steps 3–6 until the desired number of reaction cycles are complete. One can stop the experiment after a complete cycle, but it is preferred that all PED cycles be finished in the same day.
8. After the final cycle of PED, treat the beads twice with 1 mL of 20% piperidine in DMF at room temperature (once for 5 min and again for 15 min).
9. Resuspend the beads in 1 mL of TFA and add 20 μ L (0.27 mmol) of dimethyl sulfide and 25 mg (0.18 mmol) of ammonium iodide. Allow the vessel to sit on ice for 30 min. This step reduces any oxidized methionine to facilitate CNBr cleavage.
10. Drain the vessel and wash exhaustively (at least 10 times) with water to remove all traces of ammonium iodide. Suspend the beads in 1 mL of water and transfer them by pipet into a petri dish. Use additional water if necessary to transfer all of the beads into a petri dish (use a dissecting microscope to ensure that all of the beads are transferred). Transfer the beads into individual Eppendorf tubes (1 bead/tube) with a pipet. Depending on the number of positive beads, this step may require from a few to a several hundred tubes.
11. Add 20 μ L of the CNBr solution to each Eppendorf tube and allow the reaction to proceed in the dark overnight (12–14 h is optimal).
12. Evaporate the TFA and CNBr under vacuum (e.g., in a Speedvac).
13. Store the dried samples (in the same Eppendorf tube) at 4 °C. **MALDI-TOF Analysis** This procedure should be preferably performed on the same day as when the samples are dried under vacuum. Although peptides are generally stable, we have found that the sooner this step is done, the better the MS results.
14. Add 5 μ L of 50:50:0.1 (v/v) acetonitrile/water/TFA into each tube.
15. Vortex the tube for 10 s and centrifuge for 10 s at 5000 RPM.
16. Repeat step 2 twice.

17. Wash the MALDI plate several times with methanol and water (The last washing step should be with methanol). Allow the plate to air dry. If the plate is dirty, it should be cleaned by sonication in 1:1 methanol/acetonitrile.
18. Mix 1 μ L of the peptide solution with 2 μ L of a matrix solution containing 10 mg/mL 4-hydroxy- α -cyanocinnamic acid in a 50:50:0.1 (v/v) acetonitrile/water/TFA.
19. QUICKLY mix the solution with a pipet and transfer 1 μ L of the mixture onto the MALDI plate. Hold the micropipette tip a few inches from the MALDI plate and slowly press down on the plunger to cause the solution to bead at the end of the pipette tip. Carefully dab the bead onto one of the sample wells on the MALDI plate. If the spots appear salty or crusty they may need to be re-spotted.
20. Once all of the samples are spotted, the plate is ready to be analyzed by MALDI-TOF. Consult with your local instrumentation facilities to determine the optimal protocol.

MS Spectral Data Analysis

Several programs have been designed to aid in reading mass spectrometry data. For example, Bruker provides its own FlexAnalysis. For PED-MS analysis, we have found a third party software called “MoverZ” to work well. The main piece of advice for assigning sequences is to look for landmarks. Our libraries contain an N-terminal alanine residue which can be used to determine where to start reading. Further, because the library is acylated at the pY, there is a distinctive +42 peak in the middle. Finally, the linker region adds up to 627, which can usually be seen.

Verification of Results

We typically resynthesize a few of the selected peptides and test them for binding to the cognate protein domain. Fluorescence anisotropy is the preferred method given that it requires a minimal amount of peptide sample and is high-throughput. For an overview of this technique please see CPCB 1: 1–15 2009.

COMMENTARY

Background Information

Comparison to Other Methods—At least two other methods have been widely employed to profile the specificity of protein domains. Phage display can yield similar information to OBOC libraries. The advantage of phage display libraries is that they can generate libraries of higher sequence diversity (10^9 for phage display vs 10^7 for OBOC libraries). A major drawback of phage display is that the peptides are generally limited to the 20 proteinogenic amino acids. As such, phage display is not well suited for protein domains that require modified amino acids (e.g., SH2 domains). Another method that has been employed to examine specificity is oriented peptide display libraries (OPAL), which can be carried out either in solution (Songyang et al. 1993) or as peptide arrays (Rodriguez et al., 2004). This method also relies on solid phase peptide synthesis and can easily incorporate non-natural amino acids such as pY. Further, the method is more high-throughput and can be completed faster than an OBOC library can be screened. However the OPAL method does not provide individual binding sequences. When a protein domain recognizes multiple ligand classes, this OPAL data represent a statistical average of the different classes and may not reveal much information at all about the individual consensus sequences.

Applicability to Other Systems—Although the focus of this unit has so far been limited to studies of the binding specificity of protein domains, OBOC libraries can be readily

applied to several other systems. A related area is the profiling of the substrate specificity of enzymes. We and other investigators have screened OBOC peptide libraries to determine the substrate specificities of protein kinases (Wu et al., 1994), protein phosphatases (Chen et al., 2010; Ren et al., 2011), aminopeptidases (Xiao et al., 2010), and endoproteases (Ekici et al., 2007), among others. OBOC libraries of cyclic peptides (Liu et al., 2011) and peptoids (Kodadek and Bachhawat-Sikder, 2006) have also been synthesized and screened to identify potent inhibitors against a wide variety of macromolecular targets.

Critical Parameters

Protein Stability and Functionality—Before any sort of screening should be considered, the stability and functionality of the protein in question must be confirmed. First, the protein needs to be functional under the screening conditions. All SH2 domains, if they can be successfully expressed and purified, have worked well under the conditions outlined in these protocols. However, many proteins are inherently unstable, including some of the SH2 domains, and must be used immediately after purification. Although it is beyond the scope of this review to discuss a comprehensive list of strategies to improve protein purification, we have found that longer induction times at lower temperatures and cell lysis in a French pressure cell (instead of sonication) can sometimes improve the quality of the proteins.

In our protocol, chemical labeling of the protein domain is a necessity. As such, this creates another concern in producing functional proteins as some proteins become unstable upon labeling. There have been a few SH2 domains that seem to be stable when purified but become less stable once they are labeled. The ybbR tag/Sfp labeling method has been especially problematic in this regard. Although we were able to circumvent this problem by performing the labeling reaction at 4 °C, it does not mean this strategy will work for all proteins. Therefore, it is important to test the function of the protein before and after labeling.

Chemical Stability—Some of the chemicals used in our experiments have special stability issues and due attention is required. First, PITC is very reactive to moisture in the air and should be purchased in sealed ampoules. Once an ampoule is opened, it should be used within a day. Second, DMF breaks down over time to formaldehyde and dimethylamine. During the coupling reaction, dimethylamine may remove the Fmoc group from the peptide chain, resulting in the addition of the same amino acid twice (or more). Therefore, it is a good practice to dissolve Fmoc-amino acids in DMF immediately before use. This problem can also be circumvented by using freshly distilled DMF or N-methylpyrrolidone (NMP) as the solvent. Third, BCIP is not stable indefinitely even when stored at –20 °C and attention should be paid to make sure that this reagent is in good condition. Fourth, the quality of the SAAP enzyme varies from vendor to vendor and when it is obtained from a new vendor, it should be tested to obtain the optimal concentration. Finally, CNBr is not only extremely toxic, but also unstable in air and toward moisture. The pure compound is a white crystal and if the reagent shows any color, it should be replaced or purified before use.

Peptide Stability—In general, peptides (either on resin or in solution) are quite stable if kept desiccated at –20 °C. Positive beads selected from libraries are stable for at least a month if stored in this manner. However, peptides released from positive beads by CNBr cleavage can undergo degradation and give poor MS spectra after long-term storage. Phosphotyrosine is known to undergo spontaneous dephosphorylation upon storage, especially under weakly acidic conditions (pH 1–6). Washing the OBOC pY peptide libraries with a 5% DIPEA/DCM solution before storage can slow down the dephosphorylation process.

Troubleshooting

Library Synthesis—Although peptide synthesis is fairly straightforward, problems can arise with some residues. As started earlier, phosphoamino acids are difficult to couple to completion, and capping is used to prevent further chain extension. Typically, this type of strategy should only be employed if coupling cannot be taken to completion after multiple rounds of coupling with HATU. It should be noted that the coupling reagents (HATU/HBTU/HCTU etc.) can cause guanidylation of the N-terminal amine resulting in peptide chain termination. Care should be taken to maintain the proper Fmoc-AA/HATU ratio (1:1) and the coupling reaction time should not exceed 2 h. Guanidylation can be prevented by performing the coupling reaction with PyBOP.

Library Deprotection—For a standard library, all amino acids are Fmoc-protected and have acid labile side chain protecting groups. Our Modified Reagent K is the deprotection agent of choice. If deprotection is incomplete, the reaction may be extended to 2.5 h, but should not exceed 3 h. If different protection schemes are used, then their deprotection conditions need to be carefully considered.

Screenings Yield Too Few or Too Many Beads or Beads Have Partial Color—Of all of the procedures outlined in this protocol, library screening requires the most practice and optimization. Several factors can lead to too many or too few positive beads. First, the concentration of the protein should be adjusted. For SH2 domains, we typically start with ~250 nM protein in solution. However, this will vary from protein to protein so different concentrations should be tested. If too many beads are positive, or if the background color is very high, one may also increase the stringency of the washing procedure. This will remove nonspecifically bound SAAP from false positive beads, but it may also reduce the amount of specifically bound protein. Thus, discretion needs to be exercised when adjusting this parameter. If no staining occurs under even the very mild washing conditions, it is likely that the protein is non-functional.

Sequences with Multiple Positively Charged Residues or Hydrophobic Residues—Many protein-protein interactions rely on hydrophobic residues for binding. However, if all of the selected peptides contain multiple aromatic residues in a random manner, then it may signal the presence of nonspecific binding. There are two potential solutions to the problem. One can use higher concentrations of blocking proteins (e.g., gelatin) during the blocking and screening steps. Alternatively, one can use a library with even lower ligand density at the bead surface to reduce the avidity effects. The random appearance of multiple positively charged residues in a sequence is also likely a result of nonspecific charge-charge interactions. One should again try to further reduce the ligand loading on the bead surface. We have had limited success in assuaging these difficulties by increasing the salt concentration in the buffers (to 300 or 500 mM). However, doing so will take the system out of physiological range. Finally, it is possible that results such as these can arise due to poor protein quality.

Poor Quality of MS Spectra—Poor MS spectral quality makes sequence assignment difficult and may be caused by poor library quality, improper partial Edman degradation, poor sample preparation/crystallization, or inappropriate setting of the MS instrument parameters. To check for library quality, one can randomly select a few beads and analyzed the released peptides by MALDI-TOF. Each bead should produce predominantly one peak, corresponding to the full-length peptide. The presence of many extra signals suggests that library synthesis was unsatisfactory.

During PED, finding a proper PITC/Fmoc-OSU ratio is critical in order to obtain a peptide ladder of even intensities from the MALDI MS spectrum. While a low ratio may result in too much chain termination and the absence of peaks in the low m/z region, a high ratio (too little Fmoc-OSU) may lead to weak signals for all peaks except for the one of the lowest m/z ratio. It is also important to vacuum dry the beads prior to the next cycle of PITC reaction, so that the pyridine/water solution can rapidly permeate through the entire bead and achieve uniform reaction.

Problems can also arise during preparing and analyzing the samples on the MALDI-TOF device. First, it is important to ensure that all of the peptide fragments derived from PED are dissolved. Repeated vortexing and centrifugation of the samples is advised. Hydrophobic peptides should be dissolved in 1:1 water and acetonitrile containing 0.1% TFA. Any solvents with higher boiling points (e.g., DMSO and DMF) should be avoided. Second, it is important to ensure that the samples are properly spotted onto the MALDI plate. If the solution starts to crystallize before spotting, it should be discarded. If any sample spots appear salty or crusty, it is likely that the matrix did not form crystals around the analyte correctly. These samples should be respotted until they look uniform. If any sample fails to give a complete peptide ladder (and therefore unambiguous sequence assignment) the first time, it should be re-spotted and analyzed.

Anticipated Results

Screening—Although no two protein domains are identical, one can anticipate the results of a screen after performing the first few. For example, screening a typical SH2 domain against 10 mg of the pY library typically yields between 10 and 30 positive beads. However, for a type of protein domain, the number of positive beads will depend on both the screening condition and the degree of sequence specificity of the domain.

PED-MS—For the SH2 and PDZ domain libraries described in this article, PED-MS sequencing should have 90% success rate (defined as the percentage of peptides with unambiguously assigned sequences). Lower success rates may result if the random sequence contains substantially more than six amino acids. Occasionally, a MS spectrum may be consistent with more than one peptide sequence (due to the presence of extra peaks). In such an event, one may be able to make the correct sequence assignment by taking advantage of the characteristic modifications of certain amino acids (Thakkar et al., 2006). For example, tryptophan is prone to oxidation and frequently gives an $M + 16$ (where M is the expected mass for the peptide). When glutamine is at the N-terminus of a peptide, it almost always undergoes partial conversion to a pyroglutamate and generates an $M - 17$ peak (loss of NH_3). Glutamic acid sometimes gives an $M - 18$ peak (loss of H_2O), though its intensity is usually quite weak. Lysine frequently gives an extra peak at $M + 119$ (due to addition of a phenylcarbamoyl group). Finally, α -cyano-4-hydroxycinnamic acid (the MALDI matrix) gives several peaks including two at m/z 650 and 825.

Data and Interpretation—To reliably arrive at a consensus sequence(s) for a protein domain, a statistically significant number of individual sequences are needed. We typically obtain ~100 individual binding sequences for each protein domain. If a domain has only a single tight consensus, this number can be substantially smaller (e.g., 30–50 sequences). If the domain recognizes multiple classes of sequences, the number should be considerably larger (e.g., 200 sequences). Another factor that can be used to determine the consensus is the color intensity of the beads. The color intensity of a bead is generally well correlated with the binding affinity of the peptide on the bead (Sweeney et al., 2005).

Time Consideration

Obviously, the time it takes to complete the entire project will vary depending on the nature of the protein domain and other factors. The estimates below assume that the researcher is already fairly experienced with each part of the protocol and there are no unforeseen problems.

Library Synthesis: 2 weeks. This assumes that the library contains six random positions. The addition of each random position takes about a day, while each invariable position takes a few hours. Bead segregation and library deprotection will take two days.

Protein Preparation: 2–3 days. This part varies greatly from one protein to the next and most of the failures are due to nonfunctional proteins or poor quality of the proteins. Protein labeling typically requires only a few hours or can be left overnight.

Screening: 4–n days. It is always a good idea to test multiple conditions in parallel to quickly arrive at the optimal condition. The screening procedure only takes about 2 days. If the proper condition (low background color, right number of beads) is found on the first attempt, then one large-scale screening will suffice to generate all of the necessary data. We recommend that one should sequence the positive beads from the small-scale experiment before going onto a large-scale screen, to make sure that the chosen condition will lead to meaningful results.

Partial Edman Degradation: 1 day. This is a robust technique and fails only as a result of human error.

MS Analysis: 0.5–1 day. Drying, dissolving, spotting 100 peptides onto a MALDI plate and analyzing the plate by MALDI-TOF MS take about half a day (3–5 h).

Acknowledgments

The work in this laboratory was supported by grants from NIH (GM062820 and CA132855).

Literature Cited

- Chen X, Tan PH, Zhang Y, Pei D. On-bead screening of combinatorial libraries: reduction of nonspecific binding by decreasing surface ligand density. *J Comb Chem.* 2009; 11:604–611. [PubMed: 19397369]
- Chen XW, Ren L, Kim S, Carpino N, Daniel JL, Kunapuli SP, Tsygankov AY, Pei DH. Determination of the Substrate Specificity of Protein-tyrosine Phosphatase TULA-2 and Identification of Syk as a TULA-2 Substrate. *Journal of Biological Chemistry.* 2010; 285:31268–31276. [PubMed: 20670933]
- Ekici OD, Karla A, Paetzel M, Lively MO, Pei DH, Dalbey RE. Altered-3 substrate specificity of *Escherichia coli* signal peptidase 1 mutants as revealed by screening a combinatorial peptide library. *Journal of Biological Chemistry.* 2007; 282:417–425. [PubMed: 17077081]
- Hoffmuller U, Russwurm M, Kleinjung F, Ashurst J, Oschkinat H, Volkmer-Engert R, Koesling D, Schneider-Mergener J. Interaction of a PDZ protein domain with a synthetic library of all human protein C termini. *Angew Chem Int Edit.* 1999; 38:2000–2004.
- Joo SH, Pei D. Synthesis and screening of support-bound combinatorial peptide libraries with free C-termini: determination of the sequence specificity of PDZ domains. *Biochemistry.* 2008; 47:3061–3072. [PubMed: 18232644]
- Kessels HW, Ward AC, Schumacher TN. Specificity and affinity motifs for Grb2 SH2-ligand interactions. *Proc Natl Acad Sci U S A.* 2002; 99:8524–8529. [PubMed: 12084912]
- Kodadek T, Bachhawat-Sikder K. Optimized protocols for the isolation of specific protein-binding peptides or peptoids from combinatorial libraries displayed on beads. *Mol Biosyst.* 2006; 2:25–35. [PubMed: 16880920]

- Kritzer JA, Luedtke NW, Harker EA, Schepartz A. A rapid library screen for tailoring beta-peptide structure and function. *Journal of the American Chemical Society*. 2005; 127:14584–14585. [PubMed: 16231906]
- Lam KS, Salmon SE, Hersh EM, Hraby VJ, Kazmierski WM, Knapp RJ. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature*. 1991; 354:82–84. [PubMed: 1944576]
- Liu R, Marik J, Lam KS. A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries. *J Am Chem Soc*. 2002; 124:7678–7680. [PubMed: 12083920]
- Liu T, Qian Z, Xiao Q, Pei D. High-throughput screening of one-bead-one-compound libraries: identification of cyclic peptidyl inhibitors against calcineurin/NFAT interaction. *ACS Comb Sci*. 2011; 13:537–546. [PubMed: 21848276]
- Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*. 2003; 302:636–639. [PubMed: 14576432]
- Olivos HI, Bachhawat-Sikder K, Kodadek T. Quantum dots as a visual aid for screening bead-bound combinatorial libraries. *Chembiochem*. 2003; 4:1242–1245. [PubMed: 14613120]
- Ren L, Chen X, Luechapanichkul R, Selner NG, Meyer TM, Wavreille AS, Chan R, Iorio C, Zhou X, Neel BG, Pei D. Substrate specificity of protein tyrosine phosphatases 1B, RPTPalph, SHP-1, and SHP-2. *Biochemistry*. 2011; 50:2339–2356. [PubMed: 21291263]
- Rickles RJ, Botfield MC, Zhou XM, Henry PA, Brugge JS, Zoller MJ. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A*. 1995; 92:10909–10913. [PubMed: 7479908]
- Rodriguez M, Li SS, Harper JW, Songyang Z. An oriented peptide array library (OPAL) strategy to study protein-protein interactions. *J Biol Chem*. 2004; 279:8802–8807. [PubMed: 14679191]
- Rodriguez M, Yu X, Chen J, Songyang Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. *J Biol Chem*. 2003; 278:52914–52918. [PubMed: 14578343]
- Sadowski I, Stone JC, Pawson T. A Nucleocatalytic Domain Conserved among Cytoplasmic Protein-Tyrosine Kinases Modifies the Kinase Function and Transforming Activity of Fujinami Sarcoma-Virus P130gag-Fps. *Molecular and Cellular Biology*. 1986; 6:4396–4408. [PubMed: 3025655]
- Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol*. 2006; 7:473–483. [PubMed: 16829979]
- Songyang Z, Shoelson SE, Chaudhuri M, Gish G, Pawson T, Haser WG, King F, Roberts T, Ratnofsky S, Lechleider RJ, et al. SH2 domains recognize specific phosphopeptide sequences. *Cell*. 1993; 72:767–778. [PubMed: 7680959]
- Sweeney MC, Wavreille AS, Park J, Butchar JP, Tridandapani S, Pei D. Decoding protein-protein interactions through combinatorial chemistry: sequence specificity of SHP-1, SHP-2, and SHIP SH2 domains. *Biochemistry*. 2005; 44:14932–14947. [PubMed: 16274240]
- Thakkar A, Wavreille AS, Pei DH. Traceless capping agent for peptide sequencing by partial Edman degradation and mass spectrometry. *Anal Chem*. 2006; 78:5935–5939. [PubMed: 16906744]
- Tonikian R, Zhang YN, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, Xin XF, Chan AC, Seshagiri S, Lasky LA, Sander C, Boone C, Bader GD, Sidhu SS. A specificity map for the PDZ domain family. *Plos Biol*. 2008; 6:2043–2059.
- Waksman G, Kominos D, Robertson SC, Pant N, Baltimore D, Birge RB, Cowburn D, Hanafusa H, Mayer BJ, Overduin M, Resh MD, Rios CB, Silverman L, Kuriyan J. Crystal-Structure of the Phosphotyrosine Recognition Domain Sh2 of V-Src Complexed with Tyrosine-Phosphorylated Peptides. *Nature*. 1992; 358:646–653. [PubMed: 1379696]
- Wang X, Peng L, Liu R, Xu B, Lam KS. Applications of topologically segregated bilayer beads in 'one-bead one-compound' combinatorial libraries. *J Pept Res*. 2005; 65:130–138. [PubMed: 15686543]
- Wang XB, Zhang JH, Song AM, Lebrilla CB, Lam KS. Encoding method for OBOC small molecule libraries using a biphasic approach for ladder-synthesis of coding tags. *Journal of the American Chemical Society*. 2004; 126:5740–5749. [PubMed: 15125667]
- Wu J, Ma QN, Lam KS. Identifying substrate motifs of protein kinases by a random library approach. *Biochemistry*. 1994; 33:14825–14833. [PubMed: 7993909]

- Xiao Q, Zhang F, Nacev BA, Liu JO, Pei D. Protein N-terminal processing: substrate specificity of *Escherichia coli* and human methionine aminopeptidases. *Biochemistry*. 2010; 49:5588–5599. [PubMed: 20521764]
- Yin J, Straight PD, McLoughlin SM, Zhou Z, Lin AJ, Golan DE, Kelleher NL, Kolter R, Walsh CT. Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc Natl Acad Sci U S A*. 2005; 102:15815–15820. [PubMed: 16236721]
- Yu XC, Chini CCS, He M, Mer G, Chen JJ. The BRCT domain is a phospho-protein binding domain. *Science*. 2003; 302:639–642. [PubMed: 14576433]
- Zhang X, Morera S, Bates PA, Whitehead PC, Coffey AI, Hainbuecher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. *EMBO J*. 1998; 17:6404–6411. [PubMed: 9799248]

(a) Scheme showing the main steps in the spatial segregation of TentaGel beads and synthesis of an inverted peptide library containing free C-termini. (b) Picture of spatially segregated beads from (a) after staining of the N-terminal amines of peptides with chloranil (under a microscope and stained prior to side chain deprotection).

Figure 2.

Scheme showing bead segregation and the reduction of surface ligand density. Reagents and conditions: (a) soak in water overnight; (b) 0.4 equiv Boc-Phe-OSu, 0.1 equiv Fmoc-Met-OSu, and 0.5 equiv DIPEA in 55:45 $\text{Et}_2\text{O}/\text{DCM}$, 30 min; (c) 4 equiv Fmoc-Met-OH, HBTU; (d) 55% TFA in DCM; (e) Ac_2O ; (f) split-and-pool synthesis by Fmoc/HBTU chemistry; (g) TFA.

Figure 3.
Scheme showing the reactions involved in the colorimetric library screening of a pY peptide library against biotinylated GST-SH2 domains.

Figure 4.

(a) Reactions involved in partial Edman degradation. Reagents and conditions: (a) 10–30:1 PITC:Fmoc-OSu; (b) TFA; (c) Repeat steps a and b 7 times; (d) 20% piperidine in DMF; and (e) CNBr in 70% TFA. (b) A representative MALDI-TOF spectrum showing the peptide ladder derived from a single bead that displays the peptide sequence AHIpYDFIYLNBBRM-resin.