Current Topics in Peptide & Protein Research
Vol. 15, 2014

Review

Synthesis and screening of peptide libraries with free C-termini

Yen-Chih Wang and Mark D. Distefano*

Departments of Chemistry and Medicinal Chemistry, University of Minnesota, Minneapolis, MN, 55455, USA.

ABSTRACT

Peptide libraries are useful tools to investigate the relationship between structure and function of proteins. The creation of peptide libraries with free C-termini presents unique synthetic challenges. In this review, methods for creating peptide libraries using either solid-phase peptide synthesis or phage display are described. Methods for screening such libraries and their application in studying several important biological problems are also reported.

KEYWORDS: combinatorial libraries, C-terminal peptide libraries, peptide libraries, phage display, protein libraries, screening, solid-phase peptide synthesis

1. Introduction

Combinatorial synthesis has been a rapidly developing field in both academic and industrial settings in the past two decades and is now seen as an essential tool in both the discovery and development of new bioactive molecules and drugs. In simplest terms, combinatorial synthesis is a means of producing and analyzing the properties of a large number of compounds in a short period of time. Prior to the development of combinatorial methods, protein-protein and protein/enzyme – ligand/substrate interactions were typically studied using individual molecules in experiments conducted one at a time. In contrast, the introduction of combinatorial methods has allowed investigators to examine such interactions with a large number of probes simultaneously resulting in an acceleration of the research process. In this article, the application of combinatorial methods for studying interactions with the C-termini of proteins is reviewed.

*Corresponding author: diste001@umn.edu

2. Introduction to peptide library synthesis

Chemical methods: solid-phase peptide synthesis

Solid phase synthesis was pioneered for the synthesis of peptides, and most of the early work carried out on combinatorial synthesis was performed on peptides as well because the building blocks (amino acids) have regular functionalities (amino and carboxylic acid groups). There are already 20 amino acids available from Nature meaning that there are already many building blocks to generate mixtures, which will be referred to as "combinatorial libraries" in this review. For example, if a library of tetrapeptides is to be synthesized with 20 different amino acid building blocks, a mixture of $20^4 = 1.6\times10^5$ different peptides will be obtained at the end. If unnatural amino acids are also incorporated for library construction, the overall library size can be even greater.

The general approach for the chemical synthesis of peptides is that the first amino acid is attached to a solid support through its carboxyl group and then each N-protected amino acid is added in turn. During the coupling, the carboxyl group of the incoming amino acid must be activated. The commonly used activating reagents include

carbodiimides (for example, DIC), uronium (for example, HCTU) or phosphonium salts (for example, BOP) (Figure 1). After each addition, the N-protecting group must be removed before the next amino acid is added.

Currently, the most commonly used N-protecting group is the Fmoc group which can be removed under basic conditions (typically 20% piperidine); hence this process is also referred to as Fmoc-solid phase peptide synthesis (Fmoc-SPPS). The growing peptide chain is attached to the solid support so that all waste products, removed Fmoc protecting groups, excess reagents and inorganic by-products can be removed by simple filtration after each operation.

Combinatorial synthesis can be carried out such that a single product is obtained in each different reaction flask, a process known as parallel synthesis. Alternatively, the procedure can be designed such that mixtures of compounds are produced in each reaction vessel. Parallel synthesis can be performed on resins which are typically cross-linked insoluble polymeric beads or on pieces of chemically functionalized paper which will be referred to as membranes in this manuscript. An example of an early manual approach to parallel synthesis is called the teabag procedure in which the polymeric resin support is sealed in labeled meshed containers (known as teabags) [1]. The teabags are then placed in bottles which act as the reaction vessels (Figure 2). In the case of a peptide synthesis, the first amino acid is added to the resin (different amino acids for different bottles). All teabags in one specific bottle now have the same amino acid linked to the resin. The teabags from every bottle are now combined in one vessel for deprotection and washing. The teabags can then be redistributed among the bottles for the addition of a second amino acid, recombined for deprotection and washing, redistributed for addition of the next amino acid, and so on.

Another approach to parallel synthesis is called SPOT synthesis in which solutions of activated amino acids are positionally addressed and delivered in small drops to distinct points on a solid surface such as a functionalized cellulose membrane or glass slide forming a pattern of small spots. Peptide arrays are synthesized in a stepwise manner following standard Fmoc-based peptide chemistry (Figure 3) [2]. Using automated SPOT synthesis with commercial instruments, it is possible to routinely synthesize and screen 600 peptides on a 150 cm^2 membrane.

One disadvantage of parallel synthesis is that the number of peptides in the libraries is limited by the physical size of the membrane used for synthesis; typically libraries contain 1000 members per membrane. For generating larger libraries,

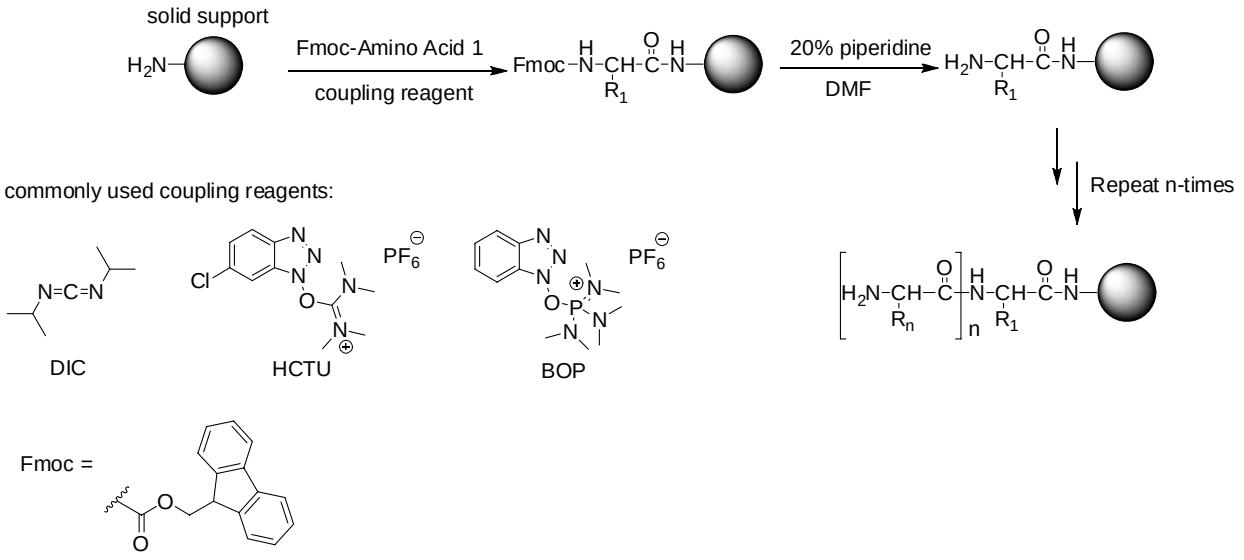

Figure 1. A schematic view of Fmoc-solid phase peptide synthesis and commonly used reagents.

the split and mix method based on peptide synthesis resin beads can be used. This strategy consists of three steps [3]: (1) Each amino acid is coupled to resin beads. (2) The beads are mixed together and separated into equal portions. The number of portions depends on how many residues are going to be coupled. Then each portion is reacted with a different amino acid. (3) Repeat the process of (1) and (2). All beads are mixed together and split into portions. This is represented in Figure 4 where the synthesis of a library of tripeptides is shown using three amino acids (Gly, Ala and Val) as building blocks.

After the split and mix synthesis, all 27 possible tripeptides will be produced. An important feature in this approach is that an individual bead contains a single peptide sequence (not a mixture). For this reason, this approach is also referred to as a one-bead-one-compound (OBOC) library. If 20 different amino acids were used to prepare a pentapeptide library, a library size of $20^5 = 3.2\times10^6$ members will be generated. The size is much greater than what can be made by parallel synthesis. However, it is important to note that in this case, the sequence of a peptide present on an individual bead is not known in advance. After screening the bead library, the peptide sequences from the selected beads must be determined via Edman degradation [4], partial Edman degradation-mass spectrometry (PED/MS) [5] or tandem mass spectrometry (MS/MS) [6].

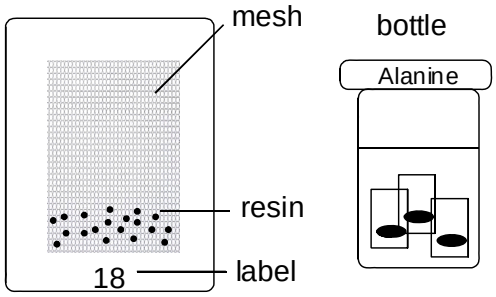

Figure 2. The teabag procedure for library synthesis. In this example several “teabags” are incubated in a bottle containing activated alanine.

Biological methods: phage display

In contrast to solid-phase peptide synthesis, in biological systems a peptide or protein is synthesized from its N-terminal end to its C-terminus. In cells, the genetic message is transcribed from DNA into RNA and then translated into protein. A sequence of three nucleotide bases, called a codon, specifies a particular amino acid that is to be incorporated into a protein. Each amino acid is specified by a three-base sequence and this relationship is known as the genetic code. Since in Nature the protein sequence is determined by DNA sequence, the problem of how to make peptide libraries is actually related to the problem of how to make DNA libraries and there are several ways of accomplishing this. First, degenerate DNA primers can be synthesized by solid-phase DNA synthesis. For example, if a random nonapeptide is to be attached to the C-terminus of a protein, a primer sequence containing $(NNK)_9$ can be made and incorporated at the C-terminus of the protein (Figure 5) where

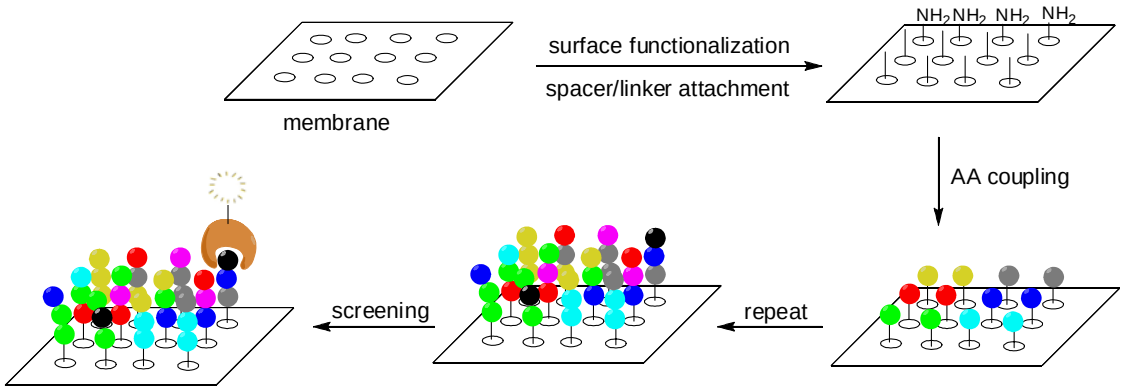

Figure 3. A schematic view of the SPOT synthesis and its application in subsequent binding assays.

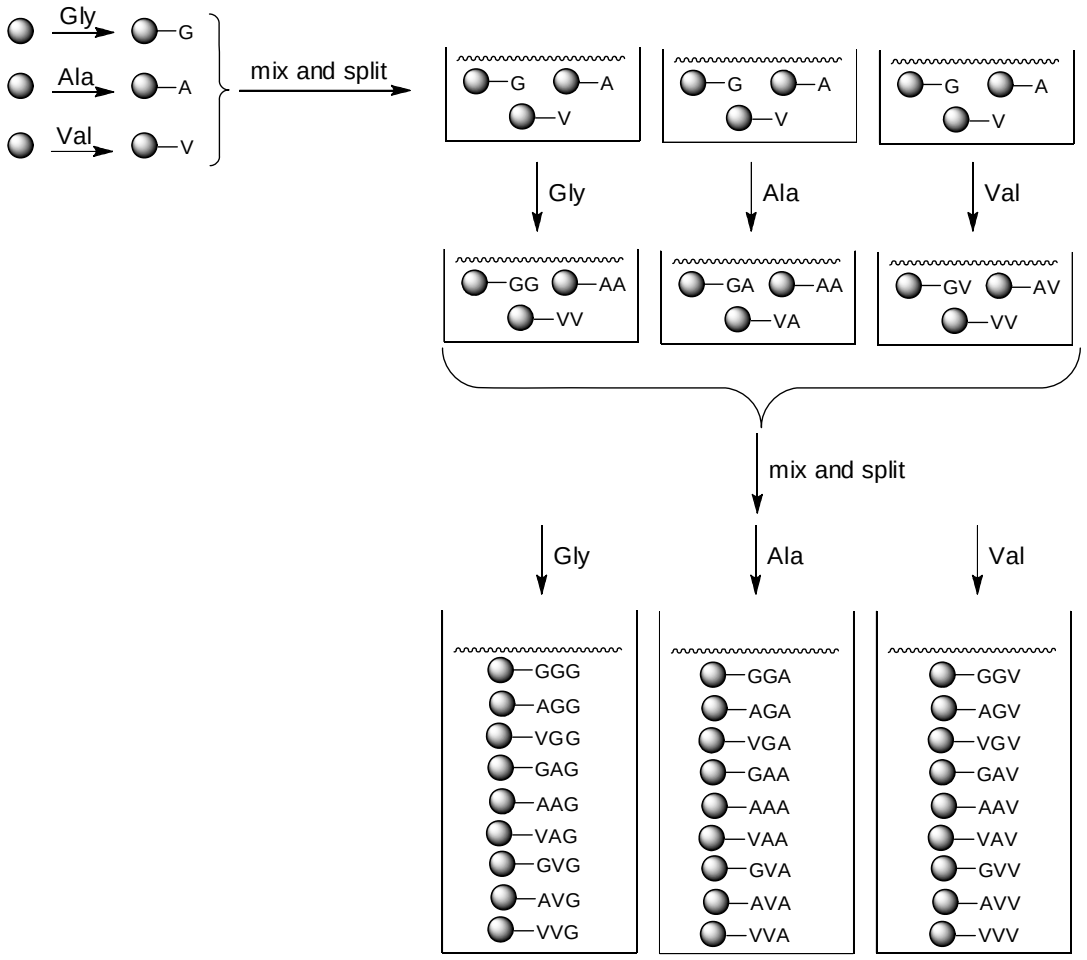

Figure 4. A schematic view of the split and mix method (or OBOC method).

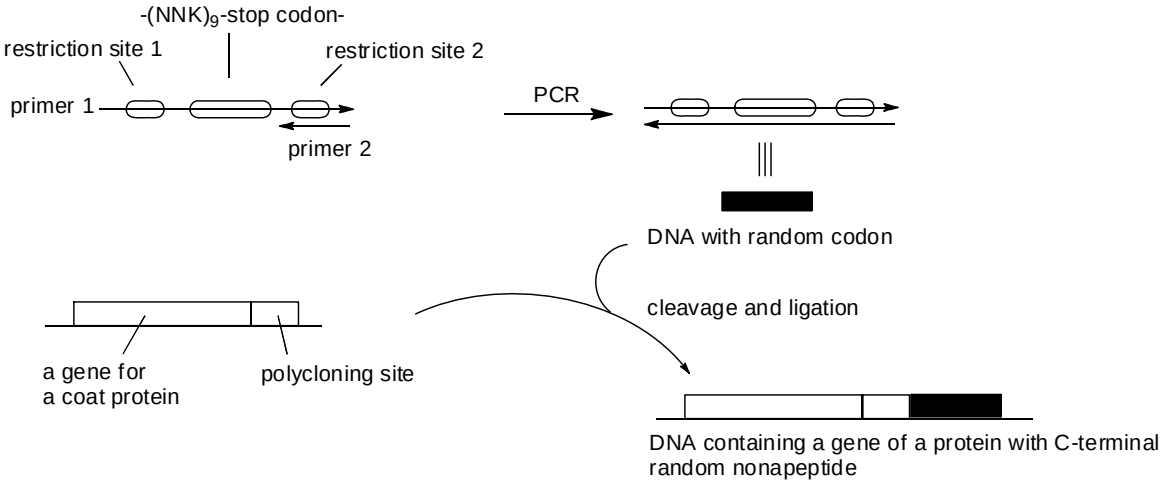

Figure 5. Illustration of preparing a random DNA library using a degenerate primer (primer 1).

N represents any of the 4 nucleotides (A, T, C, G) and K indicates G or T. NNK sequences for random codons are frequently used instead of NNN to minimize the introduction of unwanted stop codons. All 20 canonical amino acids are still retained using a NNK codon scheme.

If a library containing only certain specific sequences, for example DNA sequences present in the human genome, is desired, a completely random DNA library is not appropriate. In this case, a customized DNA microarray can be prepared/purchased and the sequences can be amplified by PCR (Polymerase Chain Reaction) and then cloned into a suitable vector such as a phagemid (Figure 6).

A DNA sequence can also be randomized by error-prone PCR. In this method, during PCR dNTPs along with mutagenic nucleotides such as 8-oxo-2'-deoxyguanosine (8-oxo-dGTP) and 4-dihydro-8H-pyrimido[4,5-c][1,2]oxazine-7-one (dPTP) are added. During DNA replication, the mutagenic nucleotides mis-pair causing mutagenesis whose rate can be controlled by the number of PCR cycles and the concentration of the mutagenic nucleotide(s).

Once suitable DNA libraries are prepared, a technique that allows a specific DNA sequence to be physically linked with its cognate translated peptide/protein such as phage display. Phage display capitalizes on the fact that the phage are made up of DNA coated with proteins encoded by that DNA. If a particular peptide sequence is translationally fused to the coding sequence for one of the coat proteins of the phage, all the progeny of that phage will display the particular peptide sequence on their surface. To use this technology, a randomized peptide-coding sequence is fused to the gene encoding a coat protein so that different phage display different versions of the peptide on their surface.

The phages that display a peptide which binds to a target of interest are "panned for" and isolated. If only a few phage particles from the mixture can be separated from most of other phages that express different, unwanted, peptides, the phages can be propagated to make hundreds of millions of identical copies. The cloned DNA can then be sequenced to determine the particular peptide sequence which binds to the target (Figure 7). Many different types of phages have been adapted for phage display. In some cases, different coat proteins from a phage can be chosen to display the peptides. Which phage and which coat protein to use depend on the specific application.

3. Some biological interactions involving the C-termini of proteins

Protein-protein interactions and protein post-translational modifications participate in a wide range of cellular process such as the control of protein localization, protein turn-over, enzyme activity and regulation of signal transduction. Misregulation of protein-protein interactions and post-translational modifications are often involved in many diseases such as cancer highlighting their importance in cell physiology. Among these interactions, some proteins bind to their partners at the C-terminus. The PDZ domain (acronym of the synapse-associated protein PSD-95/SAP90, the septate junction protein Discs-large and the tight junction protein ZO-1) is one of the most common protein modules in eukaryotic cells. These domains bind C-terminal motifs of target proteins. It is well documented that the last four amino acids in the C-terminal targets are of crucial importance in determining PDZ domain affinity. The residue selectivity in these positions is commonly used to distinguish peptide ligands into classes [7, 8]. For example, S/T-X-Ψ denotes class I, Φ-X-Ψ indicates class II, and E/D-X-Ψ describes class III where Ψ and Φ represent aliphatic and hydrophobic amino acids, respectively. Other classes of peptides recognized by different PDZ domains have also been reported [9-11]. Some proteins harbor multiple PDZ domains that are often arranged in closely linked groups. This renders the proteins capable of binding to several different transmembrane or intracellular partners at the same time, thus organizing multimeric complexes.

Neurophilin-1 (NRP-1) is another example of a protein that recognizes the C-termini of its partners. NRP-1 is a cell-surface receptor that plays an important role in angiogenesis and cardiovascular development [12, 13]. NPR-1 mediates vascular permeability induced by VEGF and certain semaphorins [14, 15]. NPR-1 interacts with VEGF

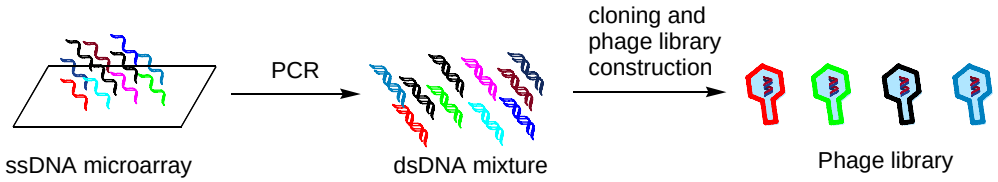

Figure 6. A general procedure for generating a customized DNA library from a microarray.

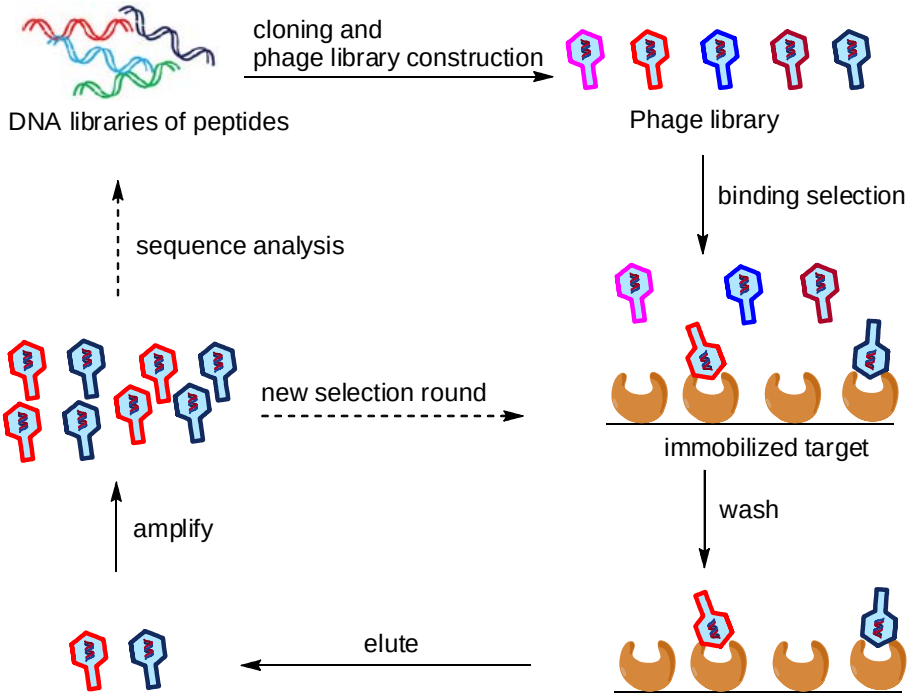

Figure 7. A schematic procedure for performing phage display.

and semaphoring 3A through C-terminal basic residues. Several peptides homologous to the VEGF C-terminus are known to compete with VEGF-NPR-1 interaction [16, 17].

After a protein is synthesized, it can be covalently modified by other enzymes. For example, it can be cleaved by a protease, phosphorylated by a kinase or lipidated by a prenyltransferase through a process called post-translational modification whose purpose is to switch on (or switch off) an enzyme's activity, to localize the protein to a specific place in the cell or to change its binding affinity to other proteins. Some enzymes that perform such reactions recognize the C-termini of their protein substrates. For example, prenylation of a variety of proteins with isoprenoids is essential for normal cellular functions and has important roles in numerous diseases. Protein prenylation involves the addition of farnesyl (C15) or geranylgeranyl (C20) isoprenoids near the C-termini of proteins and is catalyzed by protein farnesyltransferase (PFTase) and geranylgeranyltransferases (GGTase), respectively [18]. Some of those enzymes recognize a C-terminal tetrapeptide sequence called the Ca_1a_2X motif where C is the cysteine to which the isoprenoid is attached, a_1 and a_2 are usually aliphatic amino acids and X is the major determinant for modification by either PFTase or GGTase I (Figure 8) [19, 20].

The linking of proteins to ubiquitin (Ub) is another modification associated with a variety of cellular functions, including regulation of protein

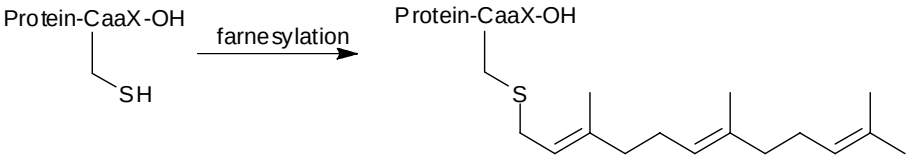

Figure 8. C-terminal farnesylation of proteins catalyzed by protein prenyltransferases.

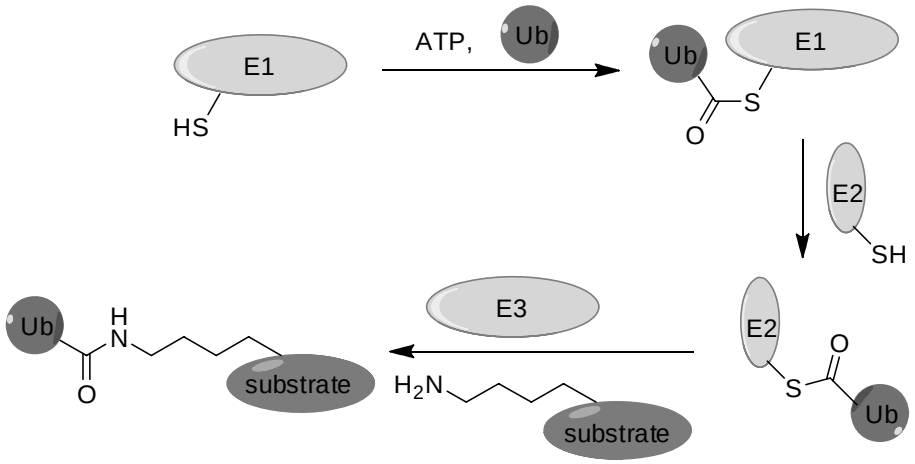

Figure 9. The enzymatic cascade of E1-E2-E3 that catalyzes Ub transfer to substrate proteins.

degradation, control of protein subcellular localization, induction of protein-protein interactions and regulation of enzymatic activities. Ubiquitin is a 76-residue protein that can be linked to other cellular proteins through an isopeptide bond between the C-terminal carboxylate of Ub and the ε-amino group of a Lys residue on the modified protein [21]. The Ub modification pathway involves three enzymes, E1, E2 and E3. The E1 enzyme catalyzes the attachment of Ub molecule to a Cys residue of E1 itself to form a Ub-E1 thioester bond. Next, Ub is transferred to E2 to form a Ub-E2 conjugate via a trans-thioesterification reaction. Finally an E3 enzyme catalyzes the transfer of Ub from E2 to the substrate protein (Figure 9) [22]. So far, two E1s, fifty E2s and more than 1000 E3s have been identified in the human genome [23, 24]. These enzymes assemble overlapping and intersecting networks for Ub transfer to the cellular target proteins.

There are nearly 20 structurally Ub-related proteins in eukaryotes including Nedd8 and SUMO. These proteins are called ubiquitin-like proteins (UBLs). Although structurally related, UBLs regulate a strikingly diverse set of cellular processes, including nuclear transport, proteolysis, translation, autophagy, and antiviral pathways [25]. The UBL modification involves a similar E1-E2-E3 cascade. Several proteins involved in these ubiquitin-related processes need to specifically recognize Ub or UBL to perform their function, including deubiquitylation enzymes (DUBs), certain trafficking proteins and ubiquitin conjugating enzymes. To recognize Ub or UBL, many proteins utilize a variety of ubiquitin-binding domains [26]. Some domains, like BUZ domain, have been found to interact with ubiquitin by binding to its free C-terminus [27, 28].

Not only are there examples of proteins recognizing C-terminal peptides, but also some peptide analogues that can bind to the C-termini of their targets. Vancomycin is a narrow-spectrum antibiotic produced by a bacterium *Streptomyces orientalis*. This complex natural product is derived biologically from a linear heptapeptide containing 5 aromatic residues. Oxidative coupling, chlorination, hydroxylation and final addition of two sugars complete the structure (Figure 10). Vancomycin

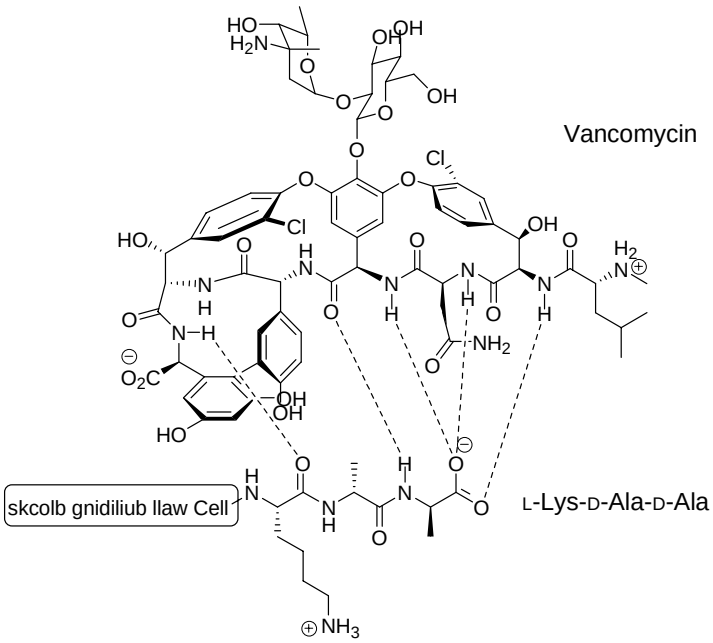

Figure 10. Vancomycin and binding interactions to the L-Lys-D-Ala-D-Ala moiety involved in cell wall biosynthesis.

forms a three-dimensional pocket into which the tail of the bacterial cell wall building block's pentapeptide moiety can fit. It is held there by the formation of five hydrogen bonds between the cell wall building block and vancomycin (Figure 10). Because vancomycin is a large molecule, it covers the tail and acts as a steric shield, blocking access to the transglycosidase and transpeptidase enzymes for cell wall construction [29].

4. Synthesis and screening of C-terminal peptide libraries using chemical methods

The well-established strategy for solid-phase peptide synthesis proceeds from C-terminus to N-terminus. The C-terminus of the peptide is tethered to the solid support (Figure 1). This creates a technical problem since it prevents the analysis of targets that require free C-termini for recognition. One way to solve this problem is to prepare peptides via solid phase synthesis but then release them and perform the screening assays in solution. Spaller and coworkers created two chemical libraries of the hexapeptide YKQTXV template in a parallel synthesis platform with 92 variations of X in library I and 186 variations of X in library II [30]. Resins were arrayed into 96-well filter plates. After the peptides were cleaved from the beads, an ELISA-type assay was used to screen the binding of peptides to the third PDZ domain (PDZ3) of PSD-95 protein (Figure 11).

Cantley and coworkers used a similar approach to investigate the peptide-binding specificities of nine PDZ domains from various proteins [7]. A primary library, KNXXXXXXXX-COOH, and a secondary library, KNXXXXXX(S,T,Y)XX-COOH, were created by the split and mix method. The peptides were released into solutions as a mixture. Some peptides with affinity to PDZ domains were enriched and then sequenced by Edman degradation. Since a peptide pool was sequenced at the same time, preference for certain amino acids at a given position was obtained, not the individual binding sequences. The above approaches used an off-bead screening platform meaning that the peptides were not attached to the solid phase when they were tested for biological activity. However, there are two disadvantages for off-bead screening. First, the off-bead screening is usually slower than on-bead screening. Using an on-bead screening platform, 10^8 beads can be readily screened which is faster than the 96-well plate format in the first example. Second, there are cases where the peptides released prove to be insoluble in the test assay and give a false-negative result which could be avoided if the peptides were attached to the bead during the assay.

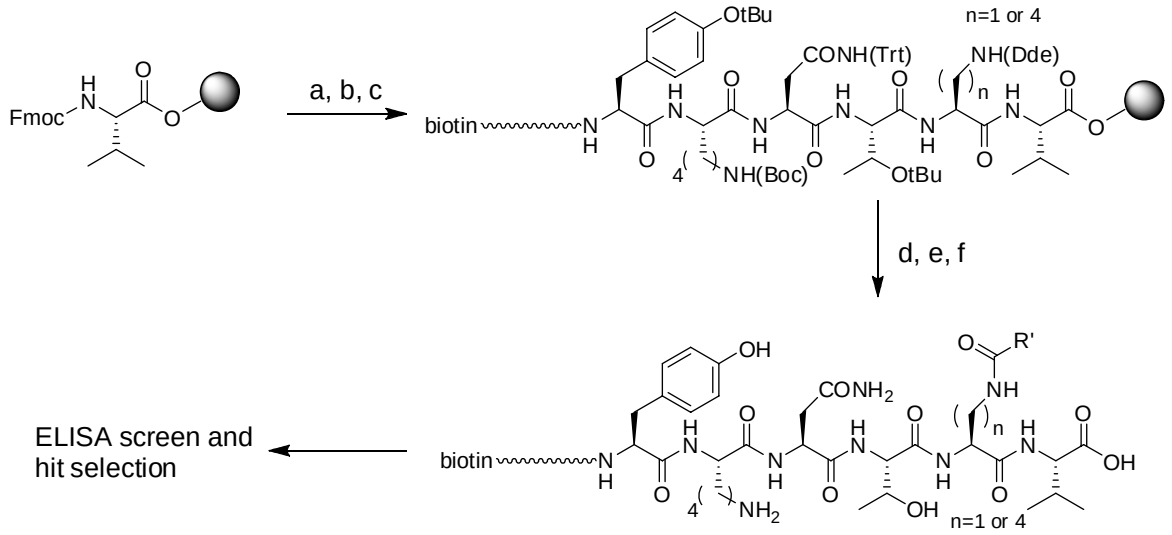

Figure 11. Synthesis and screening of C-terminal-displayed organic acid-modified libraries for PSD-95 PDZ3. Reagents and conditions: (a) Solid-phase peptide synthesis; (b) linker coupling; (c) biotinylation; (d) Dde removal (2% hydrazine); (e) array resin into 96-well filter plates and couple to a variety of organic acids ($R'CO_2H$); (f) TFA treatment.

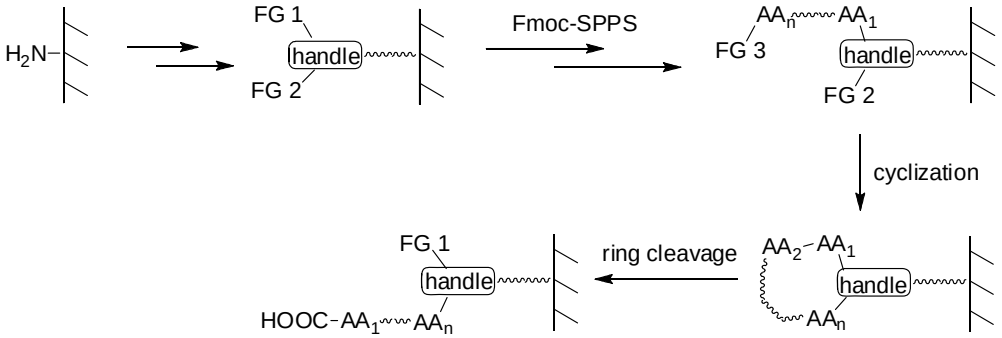

Figure 12. Illustration of peptide inversion strategy, a method for synthesis of C-terminal libraries.

If on-bead screening is to be performed, some ingenuity must be employed to present a free carboxyl terminus for target recognition because in a standard solid-phase peptide library, the carboxyl terminus of the peptide is tethered to the solid support. Several investigators have prepared peptide libraries with free C-termini using a method called “peptide inversion” (Figure 12). In this approach, a peptide is first synthesized in the conventional C to N manner. However, during this synthesis, a special handle is installed. This moiety contains two functional groups (FG 1 and FG 2). Peptide synthesis continues on FG 1 in the traditional C to N manner. Next, the complete peptide (AA_n—AA_1) is cyclized between its N-terminus and the internal FG 2 to give a cyclic peptide. Note that the N-terminus of the complete peptide can be an exposed amino group or another functional group installed (FG 3 in Figure 12). Finally, cleavage of the cyclic peptide from C-terminus of AA_1 releases the free C-terminus for screening.

Marlowe and coworkers used this strategy to generate resin-bound peptides with free C-termini (Figure 13) [31]. Note that in the structure **13-1**, the HMPA-Lys-Suc moiety present in that compound is the special handle noted in Figure 12. The HMPA is FG 1 and Suc is FG 2 which is

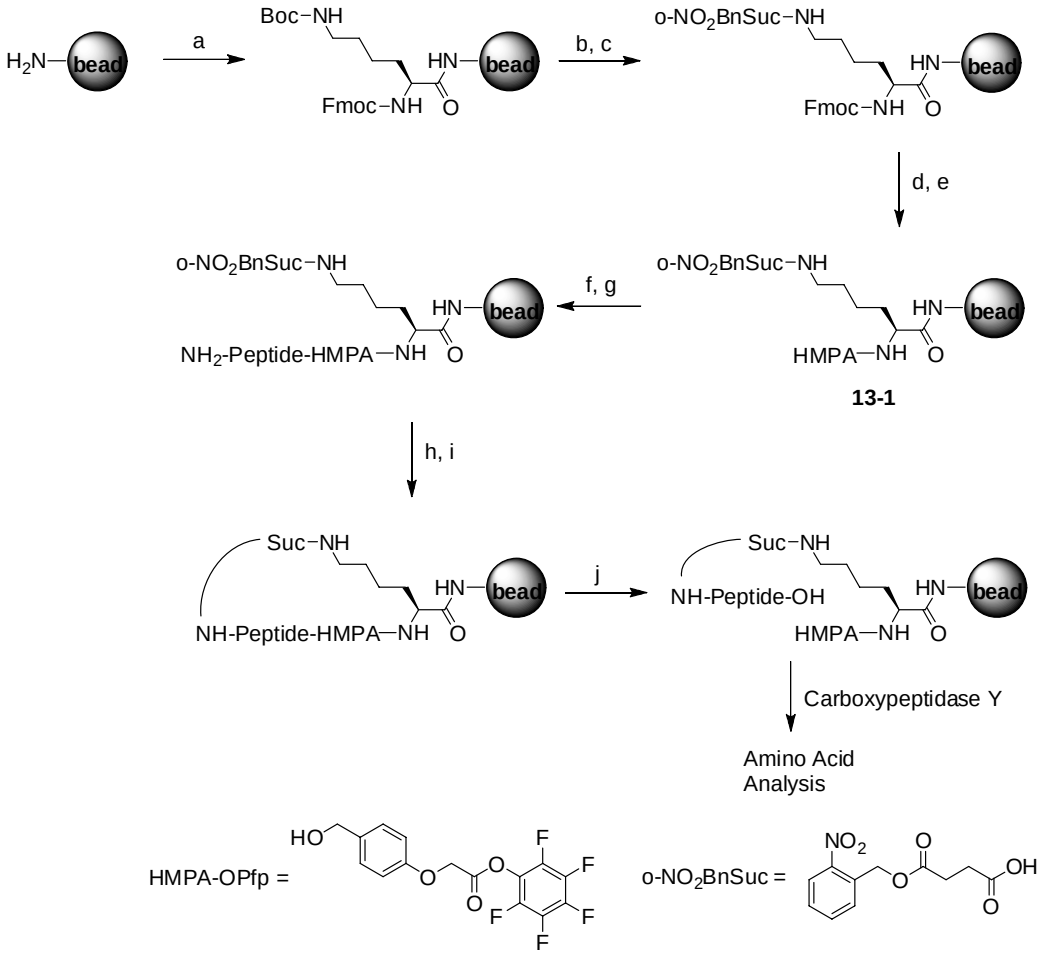

Figure 13. Synthesis of resin-bound peptides with free C-termini and enzymatic analysis. Reagents and conditions: (a) Fmoc-Lys(Boc)-OH, DIC, HOBt, CH_2Cl_2; (b) 50% TFA/CH_2Cl_2; (c) o-NO_2BnSuc, DIC, HOBt, DMF; (d) 20% piperidine/DMF; (e) HMPA-OPfp, DIEA, DMF; (f) Fmoc-Amino acid, DIC, HOBt, DMAP, DMF; (g) Solid-phase peptide synthesis; (h) hv at 360 nm, 5% NH_4OH; (i) BOP, DIEA, DMF; (j) 5% anisole, 5% ethanedithiol, 90% TFA.

protected using an o-nitrobenzyl group. Peptide synthesis continues from HMPA with the bond between the first amino acid and HMPA being an ester. After peptide synthesis is complete, the o-nitrobenzyl group is removed with UV light to expose a free carboxylic acid (FG 2). Intramolecular amide bond formation generates a cyclic peptide. The ester bond between the peptide and HMPA is labile under acidic conditions because of the electron-donating oxygen atom directly attached to the phenyl ring. Using 95% TFA, the ester bond undergoes cleavage to produce a peptide with a free C-terminus and N-terminus attached to the surface. Later, those peptides can be processed by carboxypeptidase Y to release the C-terminal 4 amino acids into solution for subsequent analysis.

If peptide libraries are created using the above strategy, one important problem is created. How can peptide sequences be determined from mixtures of beads? Edman degradation is a common way to characterize peptide sequences. However, Edman degradation requires peptides with free N-termini, a situation that does not exist for peptides linked via their N-termini. This problem is avoided if parallel synthesis is adopted. For example, in SPOT synthesis, the peptide sequences are known based on their locations on the membrane. Volkmer and coworkers used SPOT synthesis to analyze PDZ

domain regulation mechanisms (Figure 14) [32]. In their synthetic method, the Cys-βAla-HMPA in structure **14-1** was the special handle noted in Figure 12. The HMPA is FG 1 and Cys is FG 2 which is protected by a trityl group. Peptide synthesis continues from HMPA and the bond between the first amino acid to the HMPA is an ester bond.

Typical esterification methods are not compatible with the SPOT synthesis format due to the long reaction time and volatility of the solvent (typically CH_2Cl_2). In previous work, they reported a new ester bond formation method utilizing CDI (1,1'-carbonyldiimidazole) as the activating reagent for amino acid coupling and the reaction conditions were optimized for SPOT synthesis [33]. CDI was replaced by CDT (1,1'-carbonyl-di-(1,2,4-triazole)) as an activator when the amino acids are Gln, Tyr or Pro to avoid a precipitation problem (step b in Figure 14). After the peptide is complete, the N-terminus is reacted with bromoacetic acid to provide a reactive electrophile (FG 3 in Figure 12) that can react with the deprotected Cys under mild basic conditions to form a cyclic peptide. Then, the peptides with free C-termini can be generated by TFA treatment because of the lability of the ester bond between the HMPA and the amino acid 1. The use of Cys as the nucleophilic FG 2 means that Cys cannot be incorporated in the peptide library synthesis but the use of orthogonally protected Cys [34, 35] could solve this problem.

Volkmer and coworkers determined the specificity of PDZ domains by incubating the membrane in a solution containing PDZ domains with GST fusions. The high affinity peptides were detected using an anti-GST antibody fused with HRP [32, 36, 37]. Later they used a similar strategy to find candidate substrates of a periplasmic protease containing PDZ domains (Tsp) from *E. coli* [38]. A total of 648 proteins were predicted to be potential substrates based on bioinformatic analysis. Their C-terminal sequences were generated by SPOT synthesis (Figure 14) and the resulting membrane-bound peptides were incubated with a His-tagged Tsp solution. The protease Tsp was rendered inactive by mutating Ser at 430 to Ala to prevent proteolytic degradation of the immobilized peptides. Following incubation, bound Tsp was identified through the use of an anti-His antibody. This screen identified 24 candidate proteins that potentially interact with Tsp [38].

Distefano and coworkers adopted a similar approach to make peptide libraries with free C-termini by SPOT synthesis (Figure 15) [39, 40]. In their synthetic scheme, the HMPA-Glu-ODmab in structure **15-1** corresponds to the special handle shown in Figure 12. The HMPA is FG 1 and the carboxylic acid side chain of Glu is FG 2 which is protected by a Dmab group. Peptide synthesis

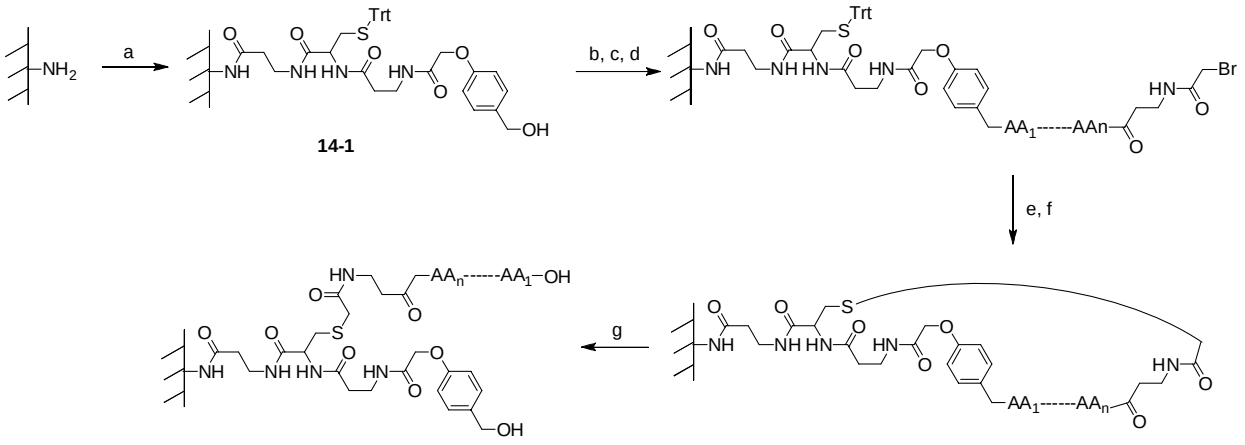

Figure 14. Synthesis of peptides with free C-termini by SPOT synthesis for PDZ binding assay. Reagents and conditions: (a) Standard Fmoc-SPPS, β-Ala, Cys(Trt), β-Ala, HMPA; (b) Fmoc-Amino acid, CDI, DMF (the CDI is replaced by CDT when the amino acids are Gln, Tyr or Pro); (c) Standard Fmoc-SPPS; (d) Bromoacetic acid 2,4-dinitrophenyl ester, NMP; (e) 7% TFA, CH_2Cl_2; (f) Cs_2CO_3/H_2O/DMF; (g) 60% TFA followed by 90% TFA.

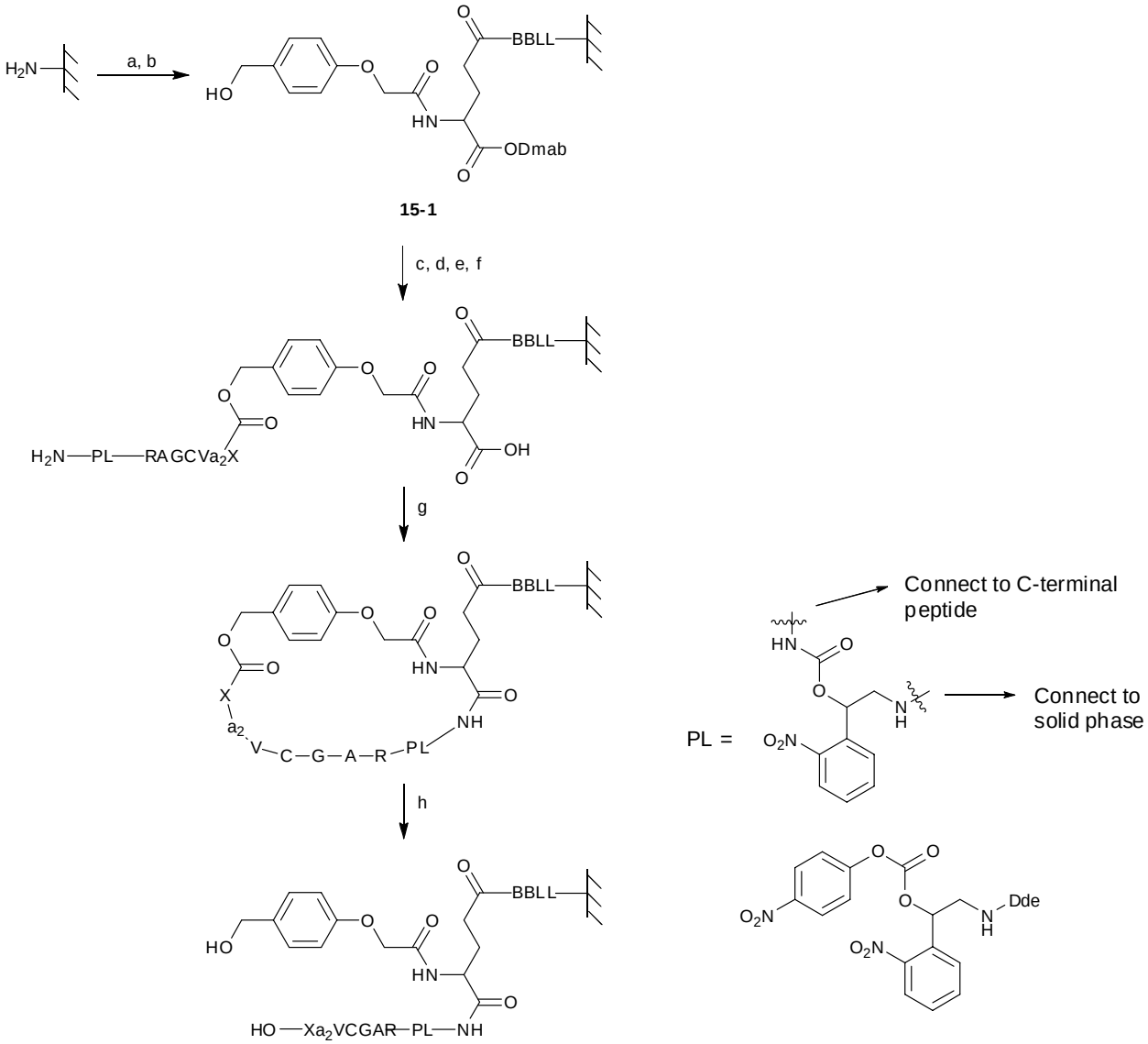

Figure 15. Synthesis of inverted peptides for screening assay. Reagents and conditions: (a) standard DIC coupling, then, capping, then 20% piperidine; (b) standard DIC coupling; (c) 0.4 M Fmoc-AA and 1.2 M CDI in DMF, then, capping, then 20% piperidine; (d) standard DIC coupling, then, capping, then 20% piperidine; (e) 0.5 M photocleavable linker, 0.5 M Et_3N in DMF; (f) 2% NH_2NH_2; (g) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF; (h) modified reagent K.

proceeded from the HMPA moiety. The N-terminal amino group of the peptide library was protected by a Dde group which can be removed along with the Dmab group at the same time using hydrazine thereby exposing the N-terminal amino group and the internal carboxylic acid group. Intramolecular amide bond formation between these two resulted in a cyclic peptide. The desired peptide libraries with free C-termini were then generated by TFA treatment because of the lability of the ester bond between the HMPA and the amino acid at the X position. It is also noteworthy that a photocleavable linker was incorporated to allow for peptide release via photolysis for subsequent analysis by MALDI to confirm product formation. They constructed 20x19 RAGCVa_2X and RAGCCa_2X libraries with the X being one of the 20 proteogenic amino acids

except Pro and a_2 being any one of the 20 proteogenic amino acids.

Wang *et al*. screened those C-terminal libraries to determine the peptide specificity of protein farnesyltransferase (PFTase). PFTase catalyzes the transfer of a farnesyl (C15) group of farnesyl diphosphate to the C-terminal Ca_1a_2X motif of a peptide or protein (Figure 8). To determine the peptide substrate specificity, each membrane was subjected to PFTase-catalyzed prenylation with an alkyne-containing FPP analogue followed by derivatization with biotin-azide via copper-catalyzed azide-alkyne cycloaddition (CuAAC). Those peptides that were prenylated by PFTase were conjugated to biotin at this step. The membrane was then subjected to an enzyme-linked assay involving streptavidin-alkaline phosphatase (SA-AP) and the chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Spots containing prenylated peptides appear turquoise colored, whereas spots where the prenylation reaction was inefficient remain colorless (Figure 16). Initially, they screened the C-terminal libraries to study the specificity of *R. norvegicus* PFTase (rPFTase). They also investigated the interplay between peptides and isoprenoid substrates of varying length (C5, C10 and C15) and the specificity of PFTases from different organisms (*R. norvegicus*, *S. cerevisiae* and *C. albicans*) [39, 40].

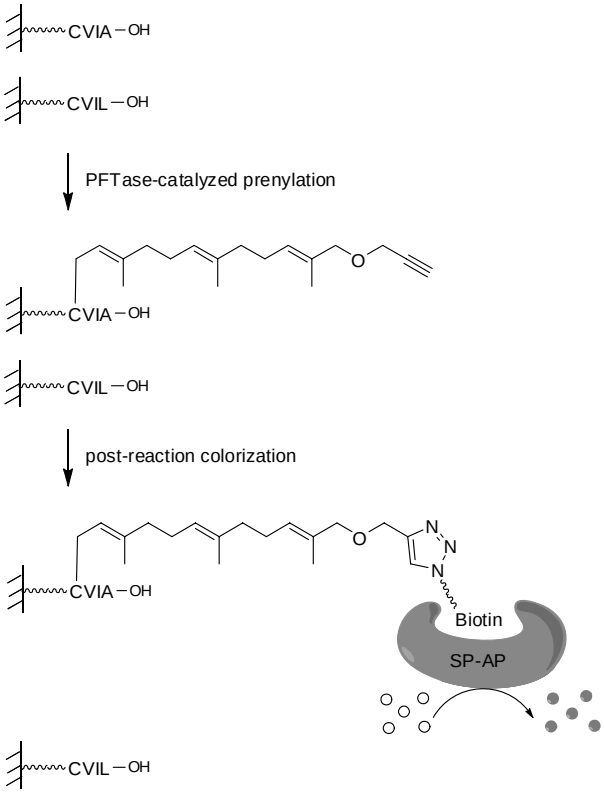

Figure 16. Screening and imaging strategy using CVIA (a substrate) and CVIL (non-substrate) as examples. Post-reaction colorization was accomplished by click reaction with biotin-azide followed by incubation with SP-AP. Colorization occurred upon the addition of BCIP.

As noted above, one limitation of parallel synthesis strategies such as SPOT synthesis is the small library size, usually < 10^3 members per 150 cm^2 membrane. If the OBOC approach is adopted, it is routine to generate and screen a library of 10^5-10^7 members. However, since all beads are present together during synthesis and screening, to determine the sequence of a positive bead (decoding), a tagging procedure must be used during library synthesis. For example, Bradley and coworkers prepared an inverted tripeptide library to be screened using a guanidinium-based tweezer receptor (Figure 17) [41]. In their synthetic scheme, 80% of the functional groups on each bead were used to prepare an inverted peptide library and the remaining 20% were employed to incorporate a coding tag. After the C-terminal tripeptide library was generated, it was screened by binding to a solution of dansyl-labeled peptide tweezer receptor or dansyl-labeled peptoid tweezer receptor (Figure 18). Highly fluorescent beads were picked up and treated with 100% TFA to expose N-terminal peptides whose sequences are the same as the C-terminal peptides of the same beads. The peptide sequences on those beads were then sequenced by Edman degradation.

Saving a portion of the loading of a bead for tagging is a good way for later structural determination. However, if both the inverted peptides and the encoding peptides are randomly attached on the beads as in the above example, the encoding peptides could potentially interfere with library screening. To solve this problem, Lam and coworkers developed a simple method to spatially segregate TenaGel beads into inner and outer layers [42, 43]. TentaGel resins are grafted coplymers consisting of polystyrene matrix onto

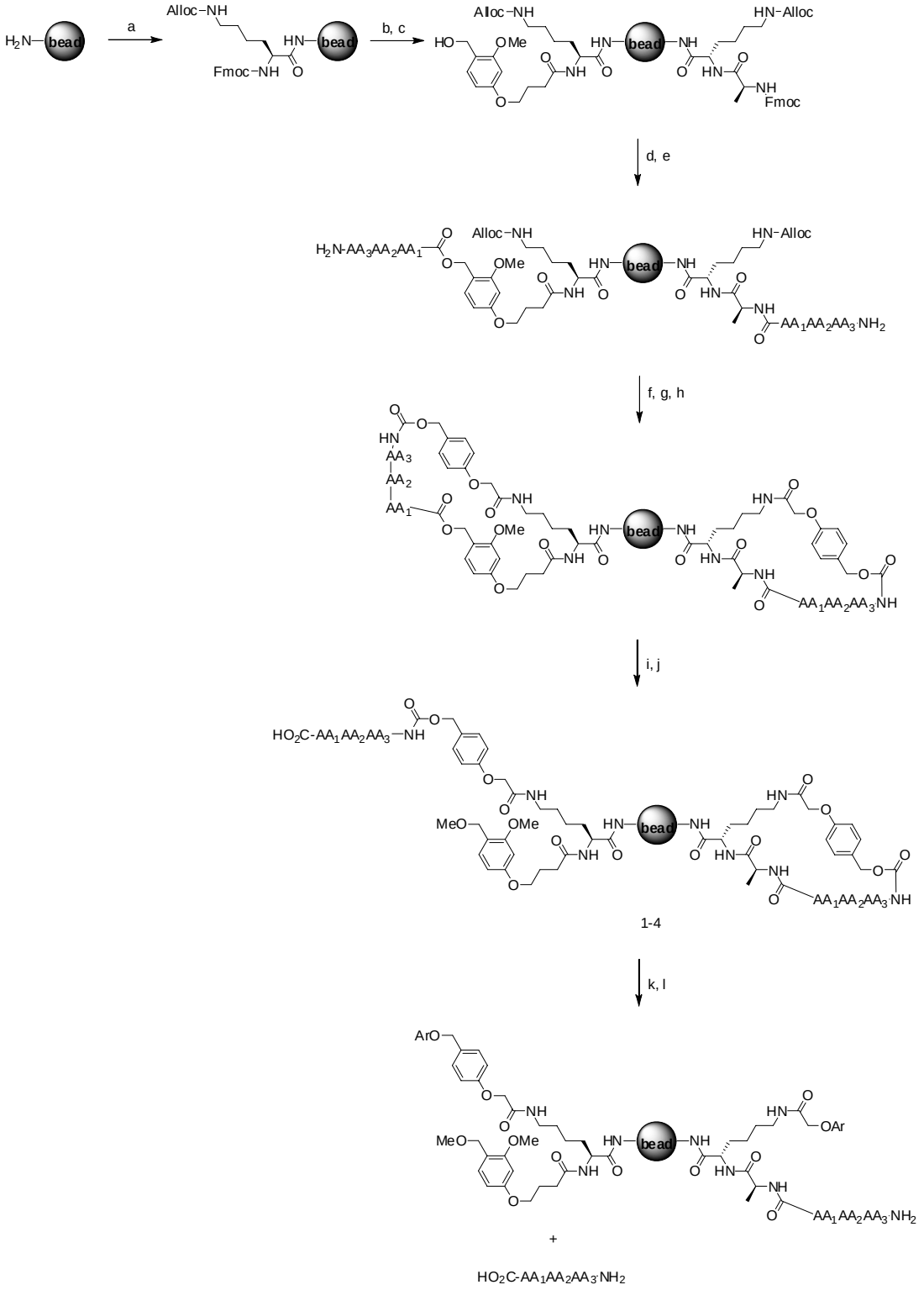

Figure 17. Synthesis of resin-bound peptides with free C-termini and post-screening treatment. Reagents and conditions: (a) standard SPPS; (b) HMPB, DIC, DMAP (80% of amine sites); (c) Fmoc-Ala, DIC, HOBt; (d) 20% piperidine; (e) split and mix Fmoc-SPPS; (f) p-$NO_2C_6H_4OCO_2CH_2C_6H_4OCH_2CO_2$Allyl; (g) $Pd(PPh_3)_4$, dimedone; (h) PyBOP, DIEA, DMAP; (i) 1% TFA/DCM; (j) MeOH; (k) screen library; (l) treat positive beads with 100% TFA.

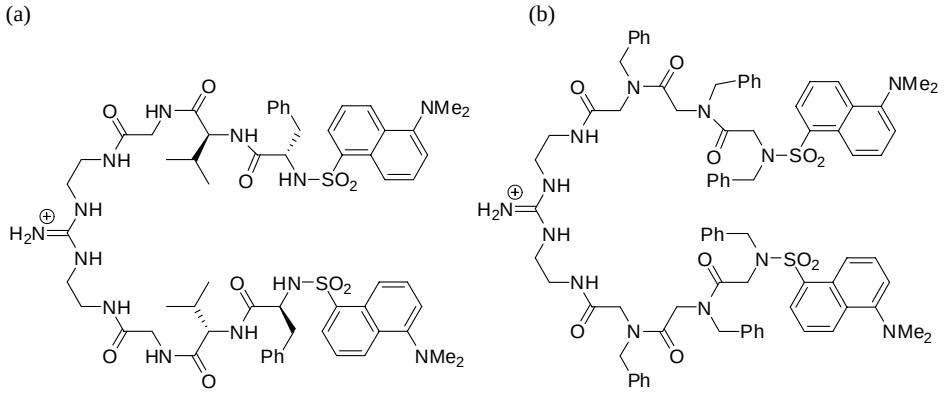

Figure 18. Structure of (a) dansyl-labeled peptide tweezer receptor. (b) dansyl-labeled peptoid tweezer receptor.

which polyethylene glycol is grafted. TentaGel beads are first soaked in water, drained and then quickly suspended in a mixture of dichloromethane and diethyl ether containing the derivatizing reagent. Because the inner phase of the bead retains water which is not miscible with the organic solvent, only peptides on the bead surface are exposed to and react with the reagent. The beads are then washed with DMF, and the free N-terminal amino groups remaining within the bead interior can react with a different reagent.

Pei and coworkers adopted this method to create peptide libraries with free C-termini on the outer layer and N-termini in the inner layer (Figure 19) [11, 44]. About 33% of the sites on a bead were used for C-terminal peptide library synthesis whereas the majority of the loading was used for coding. After forming the cyclic structure, the ester bond can be selectively hydrolyzed in NaOH to generate a peptide with a free C-terminus. An inverted peptide library containing five randomized residues was screened for binding to biotinylated PDZ domains. PDZ-bound resins were identified by subsequent treatment with streptavidin conjugated to horseradish peroxidase (SP-AP). The peptide sequences on those beads were sequenced by partial Edman degradation-mass spectrometry (PED/MS) [5, 45] since PED/MS sequencing is less time consuming than traditional Edman degradation sequencing. Pei and coworkers determined the binding specificity of the four PDZ domains of the sodium-hydrogen exchanger regulatory factor-1 (NHERF1) and channel-interacting PDZ domain protein (CIPP), [44] and later PDZ domains of the T-cell lymphoma invasion and metastasis 1 (Tiam1) and Tiam2 proteins [11]. They also used a similar strategy to evaluate the peptide specificity of BUZ domains of histone deacetylase 6 (HDAC6) and mutant ubiquitin processing protease (Ubp-M) to investigate whether these BUZ domains can bind to other peptide sequences other than Ub [46]. In that work, they used HMPA instead of HMBA as the linker to form the ester bond. Because ester bond between the HMPA and the connected amino acid can be cleaved under acidic conditions, ring cleavage and side-chain protecting group removal can be performed at the same time by TFA treatment.

Lam and coworkers used a similar bilayer/peptide inversion strategy to discover specific peptide ligands that bind to vancomycin [47]. Vancomycin was labeled with biotin to facilitate the identification of bound beads. XXXK(Ac)aa tripeptide libraries were screened first (a indicates D-Ala) (Figure 20). Later, more focused secondary libraries (YELK(XXX)aa, and XXXXaa) were designed and screened. A ligand with 50-fold higher binding affinity to vancomycin than the native ligand, Kaa, was identified.

5. Preparation and screening of C-terminal peptide libraries using biological methods

As noted above, many types of phages have been adapted for phage display. If phage display is used to screen C-terminal peptide libraries, it is important to choose a phage and a coat protein where the latter has its C-terminus exposed so that

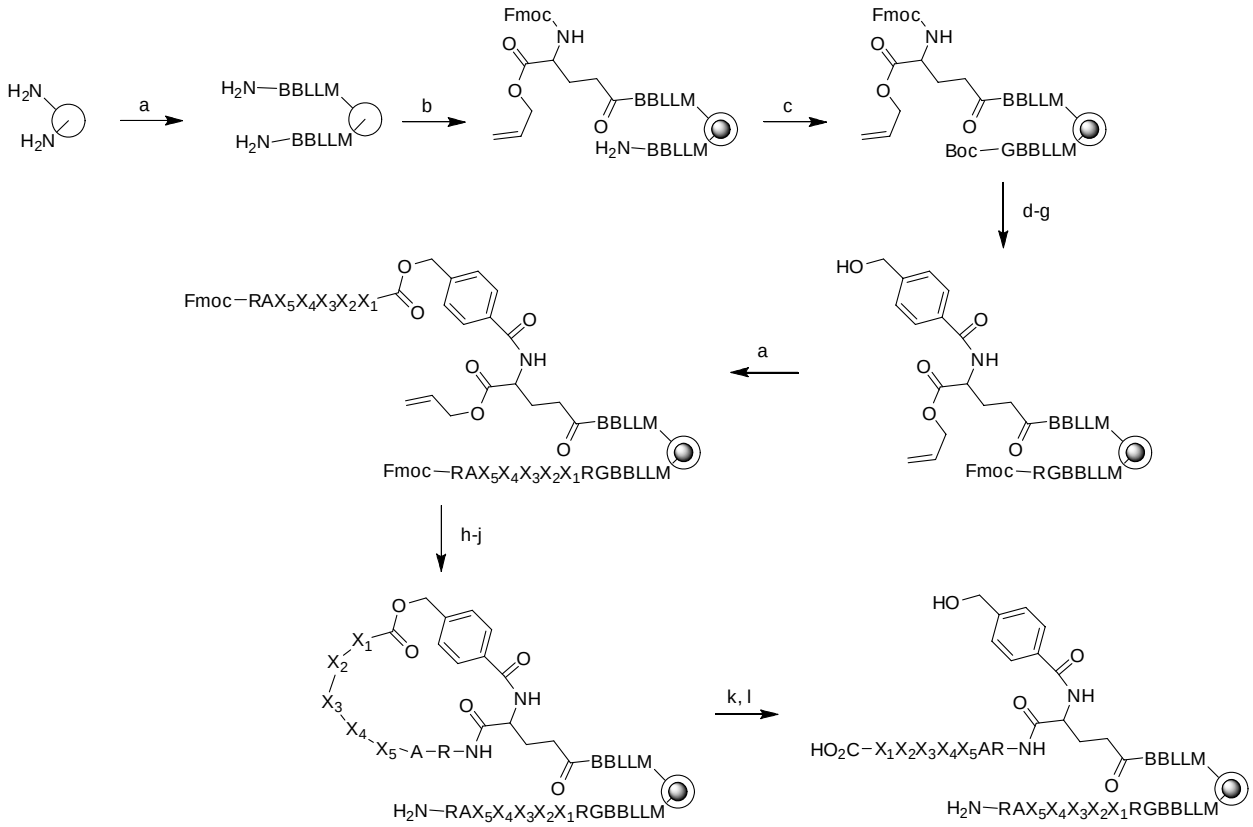

Figure 19. Synthesis of a spatially segregated and inverted peptide library. Reagents and conditions: (a) standard Fmoc-SPPS; (b) soak in water and then 0.33 equiv. of Fmoc-Glu(δ-NHS)-OAll in CH_2Cl_2/Et_2O; (c) Boc-Gly-OH, HBTU; (d) 20% piperidine; (e) HMBA, HBTU; (f) TFA; (g) Fmoc-Arg(Pbf)-OH, HBTU; (h) $Pd(PPh_3)_4$; (i) 20% piperidine; (j) PyBOP, HOBt; (k) $NaOH_{(aq)}$; (l) TFA.

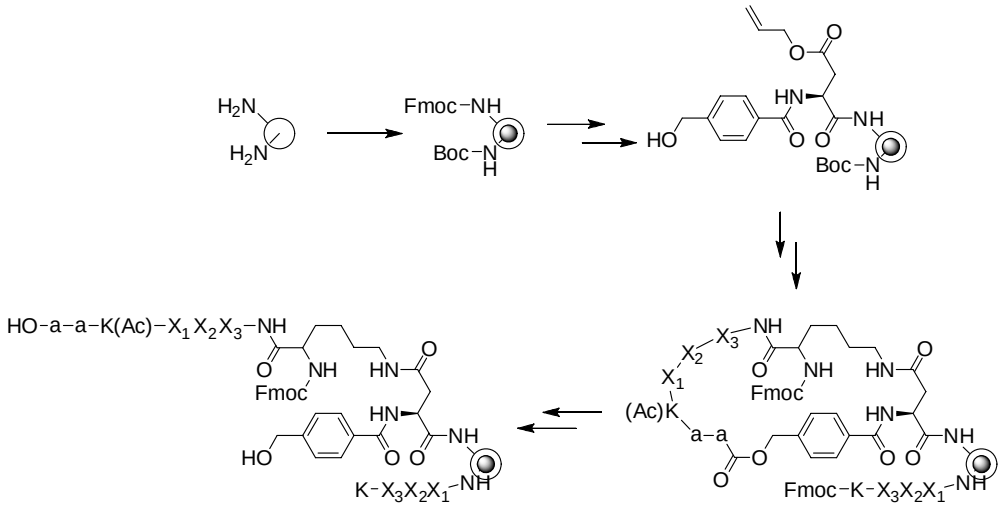

Figure 20. Synthesis of a spatially segregated and inverted peptide library.

a peptide library can be appended to its C-terminus. Phage λ is one good choice (Figure 21) [48]. The carboxyl terminus of the D-capsid protein is not involved in head formation [49]; so random peptides can be displayed on the surface of λ phage by fusion to the carboxyl terminus of D proteins. After a DNA library of phage λ sequences is prepared, it can be inserted into a phage display vector for subsequent incorporation into phage particles via *in vitro* packaging. Those λ phages can be propagated by infecting host bacteria cells to produce progeny that display the corresponding protein sequence on their surface.

Dente and coworkers used λ phage to display nonapeptides of random sequences with free C-termini [9]. In their system, about 95% of the total D protein was chimeric and the remaining 5% was synthesized by a second wild-type gene in the phage vector. Seven PDZ domains of the human INADL protein were expressed, immobilized on beads and screened to determine their binding specificities. Different consensus binding sequences were defined for each hINADL PDZ domain. Later they used a similar λ phage strategy to determine the five PDZ domains in the protein tyrosine phosphatase PTP-BL [50]. In that work they not only screened and characterized the specificity of each separate PDZ domain but also investigated the binding preferences of several combinations of PDZ domains. Interestingly while a protein segment spanning the last two PDZ domains of PTP-BL (PDZ4+5) did not reveal an altered binding preference relative to PDZ4, the combination of the first two PDZ domains (PDZ1+2) displayed a more stringent selectivity than PDZ2 alone. This demonstrates that there is an allosteric interaction between PDZ1 and PDZ2, and that the ligand selectivity of PDZ2 can be modulated by PDZ1.

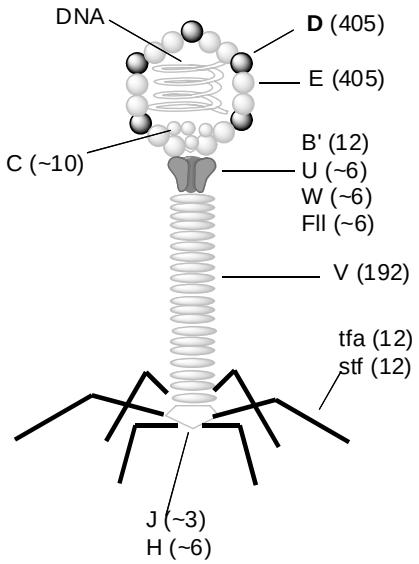

Figure 21. The structure of phage λ. Numbers indicate the number of protein copies in the particle (Adapted from Ref. 48).

T7 phage is also a good choice for C-terminal peptide display. The procedure of T7 phage display is similar to λ phage display except that the peptide-coding sequence to be displayed is cloned into the phage DNA to make a translational fusion to the phage head protein, gp10 (Figure 22) [51]. Since there are 415 copies of the head protein per phage, a large number of copies of a peptide can also be expressed on each phage, enhancing binding. Bredesen and coworkers used T7 phage display to determine the specificity of six PDZ domains from the synaptic proteins PSD-95 and SAP97 [52]. Later Spaller and coworkers also used the T7 display method to determine the specificity of the 10th PDZ domain of the multi-PDZ domain protein 1 (MUPP1) and the 3rd PDZ domain of PSD-95 [53].

Ruoslahti and coworkers used T7 phage display to identify peptides that can bind to human prostate cancer (PPC-1) cells [54]. The cancer cells were incubated with the phage libraries, washed, and then lysed to release the phages. The selected

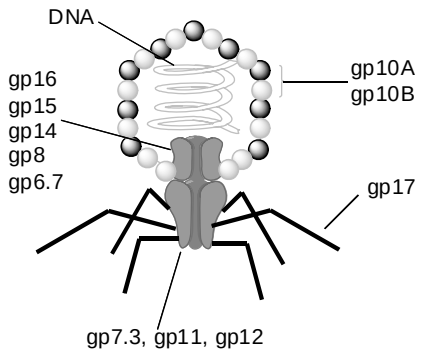

Figure 22. The structure of bacteriophage T7 (Adapted from [51]).

phages were determined to display a consensus sequence R/K-X-X-R/K. The strict requirement for a basic C-terminal amino acid is termed the C-end rule. It was also found that the bound phage could also be internalized into the cells at 37 ^{o}C. Affinity chromatography and immunostaining identified the cell receptor NRP-1 as the target responsible for the C-end rule. NRP-1 is a known mediator of vascular permeability [14, 15]. Intravenously injected positive phages penetrated into tissue parenchyma, but not negative control, indicates that the C-end rule peptides can cause vascular leakage and tissue penetration.

M13 phage is another common phage used for phage display. In contrast to some phages that lyse the cell to release themselves such as λ, T4 and T7 phages, M13 phage is secreted through the membrane of bacteria. Of the five coat proteins of the M13 phage, pIII and pVIII are the most commonly used coat proteins to construct translational fusions for phage display (Figure 23). In the M13 phage display technique, after a DNA library is cloned into an M13 display vector, the DNA is used to transform *E. coli* cells.

However, the M13 vector alone does not encode all the necessary proteins to produce progeny phage so the cells are subsequently infected with a helper phage which supplies normal head proteins. Thus in the assembled phage, the fusion proteins are usually mixed with the wild-type protein in the head.

Unfortunately there is a problem for the display of C-terminal peptides: in the natural assembly of M13 phage, it is the N-termini of pIII and pVIII that are solvent-exposed whereas the carboxyl terminus is buried within the particle [55, 56]. Hence, in M13 phage display, the random peptide sequence is usually inserted between the signal sequence required for secretion of the head protein and the N-terminus of the mature protein where it will be displayed on the surface of the phage. Sidhu and coworkers found that although the pVIII C-terminus is buried in the particle core, C-terminal fusions can still be displayed. They screened a heptapeptide library to investigate the binding specificities of PDZ domains 2 and 3 of MAGI 3, a membrane-associated guanylate kinase [57]. In their M13 phage system, there is an intervening linker of 13 residues between the C-terminus of pVIII and the random peptide to make sure the C-terminal peptides are accessible for interaction with the target protein. They also used this methodology to identify ligands for the Erbin PDZ domain [58]. Later they improved their system by screening libraries of mutated pVIII protein (plus a ten residue linker) for better C-terminal peptide display. Interestingly, the selected pVIII mutants contained mutations ranging from near the N-terminus to near the C-terminus. Their optimal pVIII mutants showed C-terminal display more than 100-fold relative to the wild-type [59]. They used this optimized methodology to determine the peptide specificity of about half of the PDZ domains known to exist in the human and *C. elegans* proteomes [60]. They also screened the M13 phage libraries of all human and viral C-terminal peptides against nine PDZ domains of human Densin-180, Erbin, Scribble, and Disks large homolog1 [61].

Yin and coworkers used a different strategy to display protein libraries connected to pIII of M13 phage. pIII, like pVIII, has its N-terminus exposed to solvent. One clever way to use pIII to display a

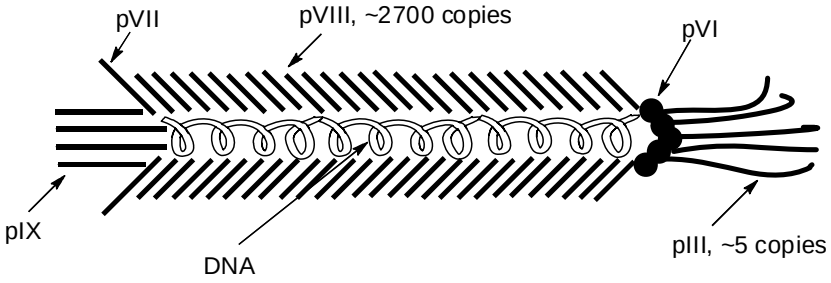

Figure 23. Schematic representation of the filamentous bacteriophage M13.

C-terminal library is to link the peptide library to pIII in the periplasm of the cell after it has been secreted instead of translationally fusing the library to pIII. In their system, pIII and the protein to be displayed were expressed from different genes on the M13 display vector and both have a signal sequence. A Jun peptide sequence was placed before the N-terminus of pIII, and a Fos peptide sequence was positioned before the N-terminus of the protein library. Thus, the Jun-pIII and the Fos-protein were secreted separately into the periplasm but formed a heterodimeric complex due to the formation of a Jun-Fos coiled coil and two disulfide bonds between Jun and Fos. In this way, phage assembly anchored the Jun-Fos dimers on the phage surface and the displayed protein was connected to the dimer with an exposed C-terminus (Figure 24). Because the cross-linking of the two proteins occurs before the phage leaves the cell, the vector DNA that encodes both of them is packaged in the phage head.

As noted above, in the ubiquitin pathway, Ub and UBLs are activated by specific E1 enzymes (see Figure 10). Based on crystallographic experiments, it had been reported that the C-terminal portion of Ub plays a key role in recognition of E1 [62]. A Ub library with random sequences from residue 71 to 75 (Ub has 76 residues) was screened to determine the specificity of two E1 enzymes (Ube1 and Uba6) [63]. The E1 protein was immobilized on a solid support. Selection of a phage-displayed Ub library was based on the formation of a Ub-E1 conjugate. The enriched phage particles were eluted with DTT treatment which resulted in cleavage of the thioester-linked conjugate (Figure 25) [63]. Later they used a similar strategy to screen a C-terminal library of Nedd8, a ubiquitin-like protein, to profile the specificity of Nedd8 activating E1 enzyme (NAE) [64]. They also made a mutant E1 (termed A1) that could not activate wild-type Ub and screened a Ub library against the A1 using the above screening strategy to find Ub mutants which could be recognized by A1. Subsequent optimization resulted in an orthogonal Ub-E1 system in which the engineered Ub can be transferred to A1 but there is no cross-activities with wild-type Ub and E1 [65].

6. Conclusion

A variety of methods to synthesize and screen peptides have been reviewed here with the focus on peptide libraries with free C-termini. Each method has its own advantages and disadvantages summarized in Table 1. The principle of SPOT synthesis is simple. Different peptides are displayed on different locations of a membrane. The positions of the spots on the membrane reveals the corresponding peptide sequences. This makes sequence determination trivial. In contrast, in the OBOC approach, the beads (and hence sequences) are mixed making a tagging procedure necessary. However, the library size obtained from SPOT synthesis is significantly smaller compared with the OBOC method or phage display making the sampling of all chemical space difficult. Phage particles can display proteins fused to their coat proteins. Hence, whole protein libraries can be screened by phage display whereas in solid-phase peptide synthesis, the length of the sequences is generally limited to less than 30 residues due to synthetic considerations. Phage libraries can also be injected into animals for *in vivo* screening.

As the solid support of SPOT synthesis and OBOC method is inert, chemical reactions can be

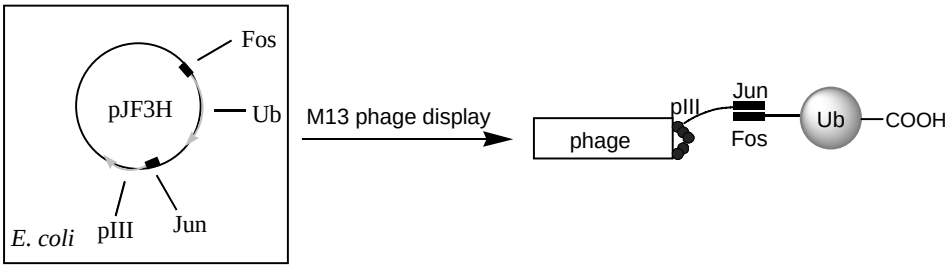

Figure 24. Phage display of Ub with free C-terminus (Adapted from [62]).

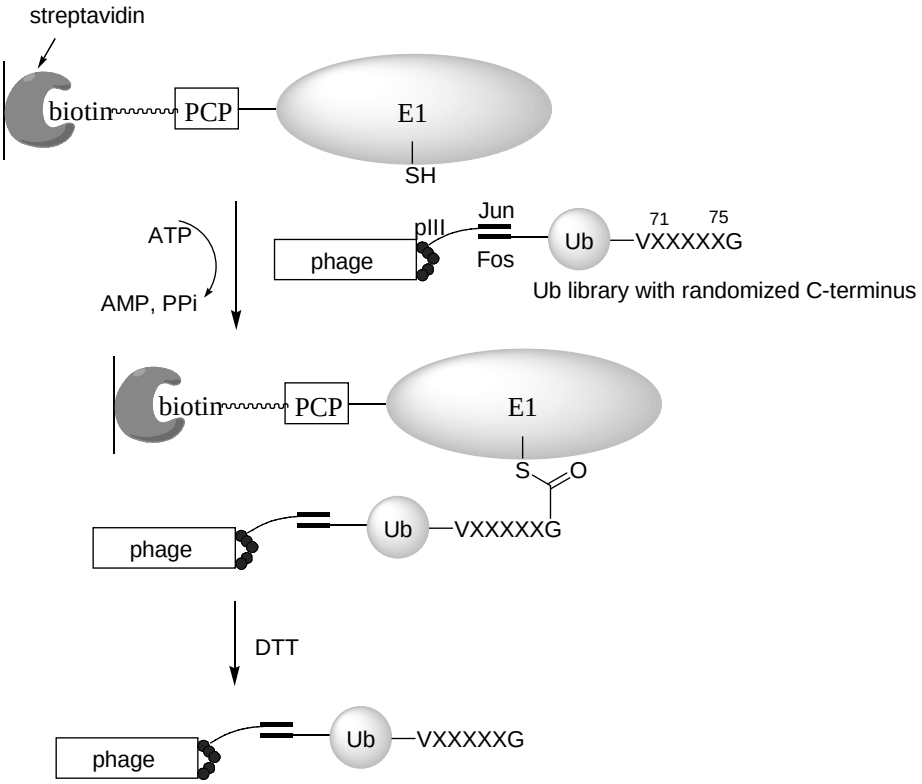

Figure 25. Phage selection of the Ub library (Adapted from [63]).

Table 1. A comparison of SPOT synthesis, one-bead-one-compound (OBOC) and phage display methods for the synthesis and screening of peptide libraries.

Category	Advantage	Disadvantage
SPOT synthesis	Easy to decode sequences Easy to incorporate UAAs Easy to do chemical derivatization	Small library size (~10^3)
OBOC	Large library size (~10^6) Easy to incorporate UAAs Easy to do chemical derivatization	Tagging is needed to decode sequences Cannot make protein libraries
Phage display	Large library size (~10^9) Can display peptide or protein libraries *In vivo* screening is possible	Difficult to incorporate UAAs Unstable in organic solvent

performed easily on it whereas proteins tend to denature in organic solvent making phages lose activity. Unnatural amino acids (UAAs) can be easily incorporated into peptides by SPPS whereas phage display is generally limited to the 20 naturally-occuring amino acids. Although enzymatic transformation or genetic code expansion technique are able to diversify peptides made by cellular machinery, the diversity of the resulting peptides still cannot compete with chemical synthesis.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health (NIH GM 084152), the National Science

Foundation (NSF CHE-1308655) and the University of Minnesota.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ABBREVIATIONS

BCIP : 5-bromo-4-chloro-3-indolyl phosphate
BOP : (Benzotriazol-1-yloxy)tris(dimethylamino) phosphonium hexafluorophosphate
CDI : 1,1'-carbonyldiimidazole
CDT : 1,1'-carbonyl-di-(1,2,4-triazole)
CuAAC : Copper-catalyzed azide-alkyne cycloaddition
Dde : N-(1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl)
DIC : N,N'-diisopropylcarbodimide
Dmab : 2-{1-[4-(hydroxymethyl)phenylamino]-3-methylbutylidene}-5,5-dimethyl-1,3-cyclohexanedione
DMF : N,N-dimethylformamide
DUB : Deubiquitylation enzyme
Fmoc : Fluorenylmethyloxycarbonyl
FPP : Farnesyl diphosphate
GGTase : Geranylgeranyltransferase
GST : Glutathione S-transferase
HCTU : O-(6-chlorobenzotriazol-1-yl)-N,N,N′,N′-tetramethyluronium hexafluorophosphate
HDAC6 : Histone deacetylase 6
HMBA : 4-Hydroxymethylbenzoic acid
HMPA : 4-(Hydroxymethyl)phenoxyacetic acid
HRP : Horseradish peroxidase
MALDI : Matrix-assisted laser desorption ionization
NAE : Nedd8 activating E1 enzyme
NRP-1 : Neurophilin-1
OBOC : One-bead-one-compound
PCR : Polymerase chain reaction
PED/MS: Partial Edman degradation-mass spectrometry
PFTase : Farnesyltransferase
SP-AP : Streptavidin-alkaline phosphatase
SPPS : Solid-phase peptide synthesis
TFA : Trifluoroacetic acid
Tiam1 : T-cell lymphoma invasion and metastasis 1
Tsp : Tail specific protease
UAA : Unnatural amino acid
Ub : Ubiquitin
UBL : Ubiquitin-like protein
VEGF : Vascular endothelial growth factor

REFERENCES

1. Houghten, R. A. 1985, Proc. Natl. Acad. Sci. USA., 82, 5131.
2. Hilpert, K., Winkler, D. F. H. and Hancock, R. E. W. 2007, Nat. Protoc., 2, 1333.
3. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M. and Knapp, R. J. 1991, Nature, 354, 82.
4. Edman, P. 1950, Acta Chem. Scand., 4, 283.
5. Thakkar, A., Wavreille, A.-S. and Pei, D. 2006, Anal. Chem., 78, 5935.
6. Biederman, K. J., Lee, H., Haney, C. A., Kaczmarek, M. and Buettner, J. A. 1999, J. Pept. Res., 53, 234.
7. 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, Science, 275, 73.
8. Vaccaro, P. and Dente, L. 2002, FEBS Lett., 512, 345.
9. Vaccaro, P., Brannetti, B., Montecchi-Palazzi, L., Philipp, S., Citterich, M. H., Cesareni, G. and Dente, L. 2001, J. Biol. Chem., 276, 42122.
10. Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A. and MacBeath, G. 2007, Science, 317, 364.
11. Shepherd, T. R., Hard, R. L., Murray, A. M., Pei, D. and Fuentes, E. J. 2011, Biochemistry, 50, 1296.
12. Takashima, S., Kitakaze, M., Asakura, M., Asanuma, H., Sanada, S., Tashiro, F., Niwa, H., Miyazaki, J.-i., Hirota, S., Kitamura, Y., Kitsukawa, T., Fujisawa, H., Klagsbrun, M. and Hori, M. 2002, Proc. Natl. Acad. Sci. USA., 99, 3657.
13. Staton, C. A., Kumar, I., Reed, M. W. R. and Brown, N. J. 2007, J. Pathol., 212, 237.
14. Becker, P. M., Waltenberger, J., Yachechko, R., Mirzapoiazova, T., Sham, J. S. K., Lee, C. G., Elias, J. A. and Verin, A. D. 2005, Circ. Res., 96, 1257.
15. Acevedo, L. M., Barillas, S., Weis, S. M., Göthert, J. R. and Cheresh, D. A. 2008, 111, 2674.
16. von Wronski, M. A., Raju, N., Pillai, R., Bogdan, N. J., Marinelli, E. R., Nanjappan, P., Ramalingam, K., Arunachalam, T.,

Eaton, S., Linder, K. E., Yan, F., Pochon, S., Tweedle, M. F. and Nunn, A. D. 2006, J. Biol. Chem., 281, 5702.
17. Starzec, A., Ladam, P., Vassy, R., Badache, S., Bouchemal, N., Navaza, A., du Penhoat, C. H. and Perret, G. Y. 2007, Peptides, 28, 2397.
18. Berndt, N., Hamilton, A. D. and Sebti, S. M. 2011, Nat. Rev. Cancer, 11, 775.
19. Moores, S. L., Schaber, M. D., Mosser, S. D., Rands, E., O'Hara, M. B., Garsky, V. M., Marshall, M. S., Pompliano, D. L. and Gibbs, J. B. 1991, J. Biol. Chem., 266, 14603.
20. Fu, H. W., and Casey, P. J. 1999, Recent Prog. Horm. Res., 54, 315.
21. Hershko, A. and Ciechanover, A. 1998, Annu. Rev. Biochem., 67, 425.
22. Pickart, C. M. 2001, Annu. Rev. Biochem., 70, 503.
23. Deshaies, R. J. and Joazeiro, C. A. P. 2009, Annu. Rev. Biochem., 78, 399.
24. Wenzel, D. M., Stoll, K. E. and Klevit, R. E. 2010, Biochem. J., 433, 31.
25. van der Veen, A. G. and Ploegh, H. L. 2012, Annu. Rev. Biochem., 81, 323.
26. Hicke, L., Schubert, H. L. and Hill, C. P. 2005, Nat. Rev. Mol. Cell. Biol., 6, 610.
27. Pai, M.-T., Tzeng, S.-R., Kovacs, J. J., Keaton, M. A., Li, S. S. C., Yao, T.-P. and Zhou, P. 2007, J. Mol. Biol., 370, 290.
28. Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., Cheng, X. and Wilkinson, K. D. 2006, Cell, 124, 1197.
29. Williams, D. H. and Bardsley, B. 1999, Angew. Chem. Int. Ed., 38, 1172.
30. Udugamasooriya, D. G., Sharma, S. C. and Spaller, M. R. 2008, ChemBioChem, 9, 1587.
31. Kania, R. S., Zuckermann, R. N. and Marlowe, C. K. 1994, J. Am. Chem. Soc., 116, 8835.
32. Boisguerin, P., Ay, B., Radziwill, G., Fritz, R. D., Moelling, K. and Volkmer, R. 2007, ChemBioChem, 8, 2302.
33. Ay, B., Volkmer, R. and Boisguerin, P. 2007, Tetrahedron Lett., 48, 361.
34. Postma, T. M., Giraud, M. and Albericio, F. 2012, Org. Lett., 14, 5468.
35. Dekan, Z., Mobli, M., Pennington, M. W., Fung, E., Nemeth, E. and Alewood, P. F. 2014, Angew. Chem. Int. Ed., 53, 2931.
36. Boisguerin, P., Leben, R., Ay, B., Radziwill, G., Moelling, K., Dong, L. and Volkmer-Engert, R. 2004, Chem. Biol., 11, 449.
37. Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R. and Oschkinat, H. 2004, J. Mol. Biol., 343, 703.
38. Weski, J., Meltzer, M., Spaan, L., Mönig, T., Oeljeklaus, J., Hauske, P., Vouilleme, L., Volkmer, R., Boisguerin, P., Boyd, D., Huber, R., Kaiser, M. and Ehrmann, M. 2012, ChemBioChem, 13, 402.
39. Wang, Y.-C. and Distefano, M. D. 2012, Chem. Commun., 48, 8228.
40. Wang, Y.-C., Dozier, J. K., Beese, L. S. and Distefano, M. D. 2014, ACS Chem. Biol., 9, 1726.
41. Davies, M., Bonnat, M., Guillier, F., Kilburn, J. D. and Bradley, M. 1998, J. Org. Chem., 63, 8696.
42. Liu, R., Marik, J. and Lam, K. S. 2002, J. Am. Chem. Soc., 124, 7678.
43. Song, A., Zhang, J., Lebrilla, C. B. and Lam, K. S. 2003, J. Am. Chem. Soc., 125, 6180.
44. Joo, S. H. and Pei, D. 2008, Biochemistry, 47, 3061.
45. Sweeney, M. C. and Pei, D. 2003, J. Comb. Chem., 5, 218.
46. Hard, R. L., Liu, J., Shen, J., Zhou, P. and Pei, D. 2010, Biochemistry, 49, 10737.
47. Yao, N., Wu, C.-Y., Xiao, W. and Lam, K. S. 2008, Biopolymers, 90, 421.
48. Rajagopala, S., Casjens, S. and Uetz, P. 2011, BMC Microbiol., 11, 213.
49. Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E., Steven, A. C., Pluckthun, A. and Wlodawer, A. 2000, Nat. Struct. Mol. Biol., 7, 230.
50. van den Berk, L. C. J., Landi, E., Walma, T., Vuister, G. W., Dente, L. and Hendriks, W. J. A. J. 2007, Biochemistry, 46, 13629.
51. Serwer, P., Wright, E., Hakala, K. and Weintraub, S. 2008, BMC Res. Notes, 1, 36.
52. Kurakin, A., Swistowski, A., Wu, S. C. and Bredesen, D. E. 2007, PLoS One, 2, e953.
53. Sharma, S. C., Memic, A., Rupasinghe, C. N., Duc, A.-C. E. and Spaller, M. R. 2009, Biopolymers, 92, 183.

54. Teesalu, T., Sugahara, K. N., Kotamraju, V. R. and Ruoslahti, E. 2009, Proc. Natl. Acad. Sci. USA, 106, 16157.
55. Stricker, N. L., Christopherson, K. S., Yi, B. A., Schatz, P. J., Raab, R. W., Dawes, G., Bassett, D. E., Bredt, D. S. and Li, M. 1997, Nat. Biotech., 15, 336.
56. Palzkill, T., Huang, W. and Weinstock, G. M. 1998, Gene, 221, 79.
57. Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A. and Sidhu, S. S. 2000, J. Biol. Chem., 275, 21486.
58. Laura, R. P., Witt, A. S., Held, H. A., Gerstner, R., Deshayes, K., Koehler, M. F. T., Kosik, K. S., Sidhu, S. S. and Lasky, L. A. 2002, J. Biol. Chem., 277, 12906.
59. Held, H. A. and Sidhu, S. S. 2004, J. Mol. Biol., 340, 587.
60. Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J.-H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D. and Sidhu, S. S. 2008, PLoS Biol., 6, e239.
61. Ivarsson, Y., Arnold, R., McLaughlin, M., Nim, S., Joshi, R., Ray, D., Liu, B., Teyra, J., Pawson, T., Moffat, J., Li, S. S.-C., Sidhu, S. S. and Kim, P. M. 2014, Proc. Natl. Acad. Sci. USA., 111, 2542.
62. Lee, I. and Schindelin, H. 2008, Cell, 134, 268.
63. Zhao, B., Bhuripanyo, K., Schneider, J., Zhang, K., Schindelin, H., Boone, D. and Yin, J. 2012, ACS Chem. Biol., 7, 2027.
64. Zhao, B., Zhang, K., Villhauer, E. B., Bhuripanyo, K., Kiyokawa, H., Schindelin, H. and Yin, J. 2013, ChemBioChem, 14, 1323.
65. Zhao, B., Bhuripanyo, K., Zhang, K., Kiyokawa, H., Schindelin, H. and Yin, J. 2012, Chem. Biol., 19, 1265.