Synthesis and Screening of Support-Bound Combinatorial Peptide Libraries with Free C-Termini: Determination of the Sequence Specificity of PDZ Domains[†]

Sang Hoon Joo and Dehua Pei*

Department of Chemistry, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210

Received November 30, 2007

ABSTRACT: Preparation of support-bound combinatorial peptide libraries with free C-termini has been challenging in the past because solid-phase peptide synthesis usually starts from the C-terminus, which must be covalently attached to the solid support. In this work, we have developed a general methodology to synthesize and screen one-bead-one-compound peptide libraries containing free C-termini. TentaGel microbeads (90 μm) were spatially segregated into outer and inner layers, and peptides were 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 were cyclized between their N-terminal amine and a carboxyl group installed at a C-terminal linker sequence, while the internal peptides were kept in the linear form. Base hydrolysis of the ester linkage in the cyclic peptides regenerated linear peptides that contained a free α-carboxyl group at their C-termini but remained covalently attached to the resin via the N-termini ("inverted" peptides). An inverted peptide library containing five random residues (theoretical diversity of 3.2×10^6) was synthesized and screened for binding to four postsynaptic density-95/discs large/zona occluden-1 (PDZ) domains of sodium-hydrogen exchanger regulatory factor-1 (NHERF1) and channel-interacting PDZ domain protein (CIPP). The identity of the binding peptides was determined by sequencing the linear encoding peptides inside the bead by partial Edman degradation/mass spectrometry. Consensus recognition motifs were identified for the PDZ domains, and representative peptides were resynthesized and confirmed for binding to their cognate PDZ domains. This method should be generally applicable to all PDZ domains as well as other protein domains and enzymes that recognize the C-terminus of their target proteins.

Combinatorial peptide libraries have been widely used to identify specific binding ligands against receptors (*1–7*), define the substrate specificity of enzymes, (*8–14*) and develop new catalysts (*15–20*). However, proteins/enzymes that recognize the C-terminus of another peptide/protein (especially those that require posttranslationally modified C-termini) remain challenging targets for peptide library screening. Although ribosomally synthesized peptide libraries [e.g., phage display (*21, 22*) and *lac*I repressor (*23, 24*) or GFP fusion (*25*)] have been generated with free C-termini, they are generally limited to 20 proteinogenic amino acids. On the other hand, while modified amino acids or unnatural building blocks can be readily incorporated into chemically synthesized libraries, it has been technically difficult to prepare and screen a support-bound combinatorial peptide library containing a free C-terminus. The well-established solid-phase peptide synthesis methodologies all start from the C-terminus, which must be covalently attached to the support (C → N synthesis). Attempts to synthesize peptides starting from the N-terminus (N → C synthesis) are generally met with low yields and racemization problems (*26*). Another obstacle is the lack of a reliable method to routinely sequence peptides from the C-terminus after they are positively identified from a library. To avoid these technical difficulties, Songyang et al. (*27*) synthesized C-terminal peptide libraries on the solid phase, released the peptides from the resin, and performed screening assays in the solution phase. The enriched peptide pool was sequenced by Edman degradation to reveal the relative preference for certain amino acids at a given position. This method, however, does not give individual binding sequences and therefore is unable to reveal any sequence covariance (e.g., the existence of multiple consensus sequences). Other investigators have prepared peptide libraries with free C-termini by "peptide inversion" (*28–32*). In this approach, a peptide is first synthesized in the conventional C → N manner and attached to a solid support via an ester linkage. Next, the completed peptide is cyclized between its N-terminus and a C-terminal carboxyl group installed on a bifunctional linker. Finally, cleavage of the ester linkage releases the free C-terminus, but the peptide remains attached to the support at its N-terminus. To facilitate hit identification, inverted peptide libraries were synthesized

[†] This work was supported by the National Institutes of Health (GM62820).

* To whom correspondence should be addressed. Phone: 614-688-4068. Fax: 614-292-1532. E-mail: pei.3@osu.edu.

[1] Abbreviations: BCIP, 5-bromo-4-chloro-3-indolyl phosphate; Fmoc-OSU, *N*-(9-fluorenylmethoxycarbonyloxy) succinimide; HBTU, *O*-benzotriazole-*N*,*N*,*N*′,*N*′-tetramethyl-uronium-hexafluoro-phosphate; HMBA, *p*-hydroxymethylbenzoic acid; HOBt, *N*-hydroxybenzotriazole; MALDI-TOF, matrix assisted laser desorption ionization-time-of-flight; NMM, *N*-methylmorpholine; OBOC, one-bead-one-compound; PED/MS, partial Edman degradation/mass spectrometry; PITC, phenyl isothiocyanate; RU, response unit; SA-AP, streptavidin-alkaline phosphatase; SPR, surface plasmon resonance; TFA, trifluoroacetic acid.

10.1021/bi7023628 CCC: $40.75
Published on Web 01/31/2008

on cellulose membranes in a spatially addressable manner (SPOT synthesis) (*29, 30*). Unfortunately, SPOT synthesis limits the practical library size at 10^3-10^4 members and requires robotic systems. Alternatively, Davies and Bradley (*31, 32*) have synthesized a 1000-member one-bead-one-compound (OBOC)[1] library using the split synthesis method (*33–35*) by inverting only 80% of the resin-bound peptides. The remaining 20% peptides were not inverted and served as encoding tags, which were sequenced by Edman degradation. A drawback of this method is that the encoding tags may interfere with library screening by binding to the screening target. Edman degradation is also expensive and time-consuming.

Postsynaptic density-95/discs large/zona occluden-1 (PDZ) domains are a large family of protein modular domains found in organisms from bacteria to man (*36, 37*). They typically comprise ~90 amino acids and occur more than 450 times in ~250 human proteins. PDZ domains mediate protein–protein interactions by binding to the C-termini of their target proteins. Previous peptide library (*21–25, 27, 29, 30*) and structural studies (*38, 39*) indicate that most of the recognition determinants are contained within the four C-terminal residues and that a free C-terminus is required for high-affinity binding. Each PDZ domain recognizes a specific subset of peptide sequences, but different PDZ domains may have substantial overlap in their binding specificities. Since the definition of their sequence specificity is key to identifying their physiological partner proteins and cellular functions, there have been extensive studies, both theoretical (*40, 41*) and experimental (*21–25, 27, 29, 30, 42, 43*), directed toward defining the sequence specificity of PDZ domains on a genome-wide scale.

In this work, we have developed a general methodology for the synthesis and screening of OBOC peptide libraries containing free C-termini and applied it to determine the recognition motifs of four PDZ domains. This methodology is readily applicable to all PDZ domains and other proteins and enzymes that recognize the C-termini of their partner proteins/substrates.

MATERIALS AND METHODS

Materials. Fmoc-protected L-amino acids and other reagents for peptide synthesis were purchased from Advanced ChemTech (Louisville, KY), Peptides International (Louisville, KY), or NovaBiochem (La Jolla, CA). Phenyl isothiocyanate (PITC) was purchased in 1-mL sealed ampules from Sigma (St. Louis, MO), and a freshly opened ampule was used in each experiment. Streptavidin-alkaline phosphatase (SA-AP) conjugate (~1 mg/mL) was purchased from Prozyme (San Leandro, CA). 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) was from Sigma. N_α-Fmoc-Glu(δ-*N*-hydroxysuccinimidyl)-O-$CH_2CH{=}CH_2$ was prepared as previously described (*44*). DNA plasmids containing sodium-hydrogen exchanger regulatory factor-1 (NHERF1) PDZ1 domain (pGEX-NHERF1-PDZ1) and channel-interacting PDZ domain protein (CIPP) protein (pcDNA-CIPP) were generous gifts from Dr. Mike Zhu of The Ohio State University. All oligonucleotides were custom synthesized at Integrated DNA Technologies (Coralville, IA).

Synthesis of Inverted Peptide Library. The inverted peptide library was synthesized on 2.0 g of TentaGel S NH_2 resin (90 μm, 0.29 mmol/g). All of the manipulations were performed at room temperature unless otherwise noted. The linker sequence (BBLLM) was synthesized with 2 equiv of Fmoc-amino acids, using HBTU/HOBt/*N*-methylmorpholine (NMM) as the coupling reagents. The coupling reaction was typically allowed to proceed for 1 h, and the beads were washed with DMF (3×) and DCM (3×). The Fmoc group was removed by treatment twice with 20% piperidine in DMF (5 + 15 min), and the beads were exhaustively washed with DCM, DMF, and again with DCM (3× each). To spatially segregate the beads into outer and inner layers, the resin was treated with 20% piperidine in DMF (5 + 15 min), washed with DMF and water, and soaked in water overnight. The resin was drained and suspended in a solution of N_α-Fmoc-Glu(δ-*N*-hydroxysuccinimidyl)-O-$CH_2CH{=}CH_2$ (0.19 mmol, 0.33 equiv) in 30 mL of 55:45 (v/v) DCM/diethyl ether. The mixture was vigorously shaken briefly and incubated on a rotary shaker for 30 min at room temperature. The beads were washed with 55:45 DCM/diethyl ether (3×) and DMF (8×) to remove water from the beads and then treated with 2 equiv of Boc-Gly OH plus HBTU/HOBt/NMM in DMF (45 min). Next, *p*-hydroxymethylbenzoic acid (HMBA) (2 equiv) was coupled to the surface layer using standard Fmoc/HBTU chemistry. The Boc group on the encoding peptide was then removed with the treatment of trifluoroacetic acid (TFA) for 1 h, and an Arg was added to the N-terminus by treatment with Fmoc-Arg(Pbf)-OH/HBTU/HOBt/NMM. To synthesize the random region, the resin was split into 20 equal portions and each portion (100 mg) was coupled twice with 4 equiv of a different Fmoc-amino acid. The first random residue was coupled with diisopropylcarbodiimide/4,4-diaminopyridine in DCM for 6 h, while the other positions were coupled with HBTU/HOBt/NMM for 2 h. To differentiate isobaric amino acids during MS sequencing, 5% (mol/mol) CD_3CO_2D was added to the coupling reactions of Lys and Leu, and 5% $CH_3CD_2CO_2D$ was added to the reaction of Nle (*45, 46*). After the five random positions, a dipeptide Arg-Ala (or Ala-Ala) was added to the N-terminus of all peptides. Next, the allyl group on the C-terminal glutamate was removed by overnight treatment with a solution containing tetrakis(triphenylphosphine)palladium (1 equiv), triphenylphosphine (3 equiv), formic acid (10 equiv), and diethylamine (10 equiv) in anhydrous THF. The beads were washed sequentially with 0.5% diisopropylethylamine in DMF, 0.5% sodium dimethyldithiocarbamate hydrate in DMF, DMF (3×), DCM (3×), and DMF (3×). The N-terminal Fmoc group was then removed in 20% piperidine, and the beads were washed with DMF (6×), 1 M HOBt in DMF (3×), DMF (3×), and DCM (3×). For peptide cyclization, a solution of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP)/HOBt/NMM (5, 5, 10 equiv, respectively) in DMF was mixed with the resin, and the mixture was incubated on a rotary shaker for 3 h. The resin was washed with DMF (3×) and DCM (3×) and dried under vacuum for >1 h. The resin was treated with 1 M NaOH for 1 h, followed by side-chain deprotection with a modified reagent K (6.5% phenol, 5% water, 5% thioanisole, 2.5% ethanedithiol, 1% anisole, and 1% triisopropylsilane in TFA) for 1.5 h. The resin was washed with TFA and DCM, and dried under vacuum before storage at −20 °C.

Peptide Sequencing by Partial Edman Degradation/Mass Spectrometry (PED/MS). Selected resin beads (typically

1–100 beads) were pooled and subjected to partial Edman degradation in a single reaction vessel as described previously (*45, 46*). The beads were suspended in 160 μL of pyridine/water (v/v 2:1) containing 0.1% triethylamine and mixed with an equal volume of the pyridine containing Fmoc-OSU (2.5 μmol) and PITC (100 μmol). The reaction was allowed to proceed for 6 min with mixing, and the beads were washed with DCM and TFA. The beads were treated twice with ~300 μL of TFA for 6 min each. After washing the resin with DCM and pyridine, the cycle was repeated 6 times. After the final cycle, the Fmoc group was removed by treatment with 20% piperidine in DMF twice (5 + 15 min). For MALDI-TOF analysis, the degraded beads were treated with ~1 mL of TFA containing ammonium iodide (10 mg) and dimethylsulfide (20 μL) on ice for 30 min to reduce any oxidized Met. After the beads were washed with water, they were transferred into microcentrifuge tubes (1 bead/tube), and each was treated with 20 μL of 70% TFA containing CNBr (40 mg/mL) overnight in the dark. The solvents were evaporated under vacuum to dryness, and the peptides released from the bead were dissolved in 5 μL of 0.1% TFA in water. One microliter of the peptide solution was mixed with 2 μL of saturated 4-hydroxy-α-cyanocinnamic acid in acetonitrile/0.1% TFA (1:1), and 1 μL of the mixture was spotted onto a MALDI sample plate. Mass spectrometry was performed on a Bruker III MALDI-TOF instrument in an automated manner at Campus Chemical Instrument Center of The Ohio State University. The data obtained were analyzed by either Moverz software (Proteometrics LLC, Winnipeg, Canada) or Bruker Daltonics flexAnalysis 2.4 (Bruker Daltonic GmbH, Germany).

Recombinant DNA Constructs. The DNA sequence encoding each of the four CIPP PDZ domains was amplified from plasmid pcDNA-CIPP by the polymerase chain reaction (PCR). The PCR primers for subcloning were 5′-C GAA TTC CAT ATG GGA GAA CTA CAC ATT ATC GAA CTG G-3′ and 5′-G GGA TCC GTC GAC CTG ACT GAC TGC ATC TTC GTT TC-3′ for PDZ1 domain; 5′- C GAA TTC CAT ATG GGC CAG GAA ATG ATC ATA GAA ATA TCC-3′ and 5′-G GGA TCC GTC GAC CTG TGC TTC ATC TCT GTA TAC CAC C-3′ for PDZ 2 domain; 5′-C GAA TTC CAT ATG GAC GAG GAG AAC TTG GAG GTG-3′ and 5′-C ATC GTG GTC GAC GGA ACC GGC GCG GAG TCT-3′for PDZ 3 domain; 5′-C GAA TTC CAT ATG ACA GAG GAG GAA CCA AGG ACT-3′ and 5′-C ATC GTG GTC GAC CTG GGT CGC TAT GGC ACT TAT-3′ for PDZ 4 domain. The PCR products were digested with restriction enzymes *Nde*I and *Sal*I and ligated into an engineered pET-22b(+) vector that contains a ybbR tag for the specific labeling with phosphopantetheinyl transferase Sfp (*47, 48*). All DNA constructs were confirmed by dideoxy sequencing at Genewiz (South Plainfield, NJ) or Plant Microbe Genomics Facility of The Ohio State University.

Purification and Biotinylation of PDZ domains. Escherichia coli BL21(DE3) cells harboring the proper expression vector were grown in Luria–Bertani media to the midlog phase and induced by the addition of 100–150 μM isopropyl β-D-thiogalactoside for 6 h at 30 °C. The cells were harvested by centrifugation and lysed in a lysis buffer (50 mM sodium phosphate, pH 7.4, 5 mM imidazole, 300 mM NaCl for CIPP PDZ domains; or 100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol for GST-PDZ1) by sonication in the presence of protease inhibitors (35 mg/L phenylmethylsulfonyl fluoride, 20 mg/L trypsin inhibitor, 1 mg/L pepstatin). The GST-PDZ1 fusion protein (for NHERF1) was purified from the crude cell lysate on a glutathione-agarose column, and the eluted protein was dialyzed against phosphate-buffered saline (pH 7.4). To label the protein with biotin, 2 mg of protein (0.7 mL) was incubated with 2 equiv of (+)-biotin *N*-hydroxysuccinimide (NHS) ester (Sigma) in 0.1 M $NaHCO_3$ (pH 8.4) for 30 min. The unreacted biotin ester was quenched by the addition of 50 μL of 1 M Tris buffer (pH 8.5). The CIPP PDZ domains, which contained C-terminal ybbR and six-histidine tags, were purified by metal-affinity chromatography and enzymatically labeled with biotin by using phosphopantetheinyl transferase Sfp and a biotin-CoA adduct (*47, 48*). The labeled proteins were passed through a G-25 column (eluted with phosphate-buffered saline) to remove any unlabeled biotin. Protein concentration was determined by the Bradford method using bovine serum albumin as standard.

Peptide Library Screening. In a MicroBio-Spin column (0.8 mL, Bio-Rad), 50 mg of the peptide library was swollen in DCM and extensively washed with DMF and then water. The resin was incubated in the HBST-gelatin buffer (30 mM Hepes, pH 7.4, 150 mM NaCl, 0.05% Tween 20, and 0.1% gelatin) for 2 h to overnight with gentle mixing at 4 °C. The resin was drained, resuspended in 800 μL of the same buffer containing 0.5–1 μM biotinylated PDZ domain protein, and incubated overnight at 4 °C. The resin was again drained, resuspended in SA-AP buffer (30 mM Tris-HCl, pH 7.4, 250 mM NaCl, 10 mM $MgCl_2$, and 70 μM $ZnCl_2$) containing SA-AP (1 μg/mL), and incubated for 15 min at room temperature with gentle mixing on a rotary shaker. The resin was washed with 800 μL of SA-AP buffer (2×), 800 μL of HBST-gelatin buffer (3×), and 800 μL of SA-AP reaction buffer (30 mM Tris-HCl, pH 8.5, 100 mM NaCl, 5 mM $MgCl_2$, 20 μM $ZnCl_2$) (3×). The resin was transferred into a single well on a 12-well plate (BD Falcon) by rinsing with 3 × 300 μL of the SA-AP reaction buffer. Upon the addition of 100 μL of 5 mg/mL BCIP in the SA-AP reaction buffer, intense turquoise color developed on positive beads in ~30 min, when the staining was terminated by the addition of 100 μL of 1 M HCl. The resin was washed extensively with water and transferred into a 35-mm Petri dish, from which the positive beads were manually picked with a pipet under a dissecting microscope.

Synthesis of Individual PDZ-Binding Peptides. Each peptide was synthesized on 100 mg of Wang resin (0.8 mmol/g) with standard Fmoc chemistry. A tripeptide (Ac-YAA or Ac-YSS) was added to their N-terminus to facilitate concentration determination. After cleavage and deprotection, the identity of the peptides was confirmed by MALDI-TOF mass spectrometry. Analytical HPLC analysis (C_{18} column) indicated that most of the peptides had >80% purity except peptide VFHRV-OH (70% pure). The crude peptides (after repeated trituration with diethyl ether) were directly used in binding assays. Peptide concentration was determined by measuring UV absorbance at 280 nm.

Determination of Dissociation Constants by Surface Plasmon Resonance (SPR). The binding affinity of the individual peptides to CIPP PDZ domains was determined by SPR analysis on a BIAcore 3000 instrument. Assays were performed in HBS-EP buffer (10 mM HEPES, pH 7.4, 150

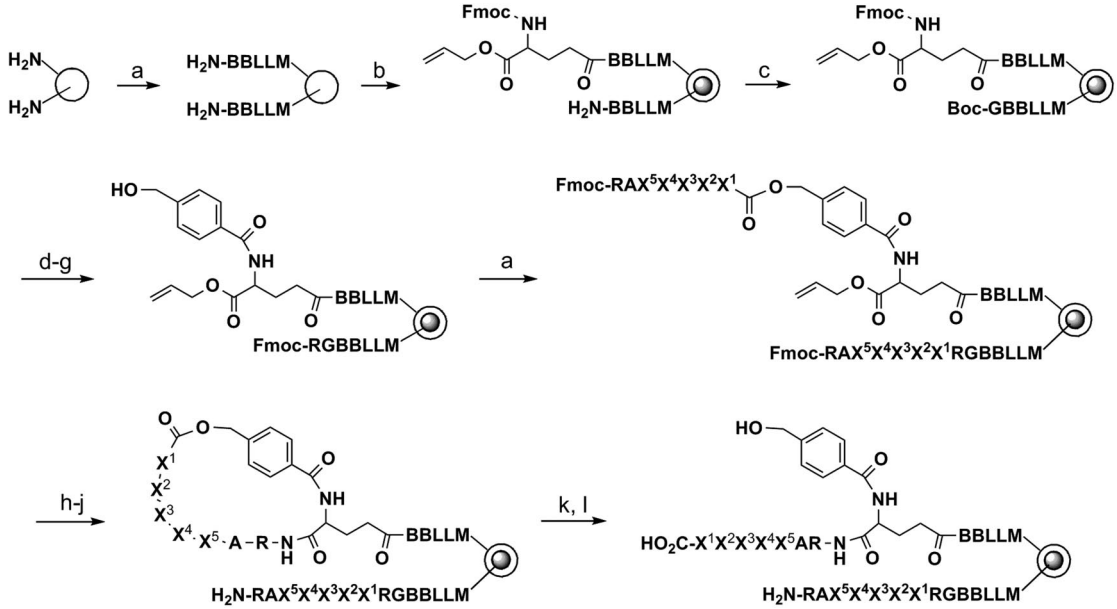

FIGURE 1: Synthesis of spatially segregated and inverted peptide library. Reagents: (a) standard Fmoc/HBTU chemistry; (b) soak in water and then 0.33 equiv of N^{α}-Fmoc-Glu(δ-NHS)-O-$CH_2CH{=}CH_2$ in Et_2O/CH_2Cl_2; (c) excess Boc-Gly-OH, HBTU; (d) piperidine; (e) HMBA, HBTU; (f) TFA; (g) Fmoc-Arg(Pbf)-OH, HBTU; (h) $Pd(PPh_3)_4$; (i) piperidine; (j) PyBOP, HOBt; (k) NaOH, H_2O; and (l) TFA.

mM NaCl, 3 mM EDTA, 0.005% polysorbate 20) at 25 °C. Biotinylated PDZ domains were immobilized to a sensorchip coated with streptavidin. The chip was first conditioned with an activation buffer containing 1 M NaCl and 50 mM NaOH following the manufacturer's instructions. Injection of the biotinylated PDZ protein (20 μg/mL) was continued until there was no further increase in the signal (typically 3500–6500 RU increase over background). For K_D measurements, varying concentrations of the peptides (0−1500 μM) dissolved in HBS-EP buffer were passed over the chip for 2 min at a flow rate of 5 μL/min. Flow cell 1 on the sensorchip, which was not loaded with any PDZ protein, was the blank. In between runs, regeneration of the surface was achieved by flowing HBS-EP buffer in most cases; when necessary, the sensorchip was regenerated by injecting a strip buffer (0.001% SDS in HBS-EP buffer) for 3−6 s at a flow rate of 50−100 μL/min. The equilibrium response unit (RU_{eq}) at a given peptide concentration was obtained by subtracting the RU of the blank flow cell from that of the PDZ domain-containing flow cell. The dissociation constant (K_D) was obtained by nonlinear regression fitting of the data to the equation: $RU_{eq} = RU_{max}[peptide]/(K_D + [peptide])$, where RU_{max} is the maximum response unit.

RESULTS

Design Strategy and Synthesis of Free C-Terminal Peptide Library. We adopted the strategy of Davies and Bradley (*31, 32*) to invert only a fraction of the peptides on each resin bead (~33%), so that the noninverted peptides (which have the same sequence but with a free N-terminus) serve as an encoding sequence to facilitate later sequence determination. To prevent the encoding peptides from interfering with library screening, we topologically segregated each resin bead into two different layers; the bead surface would display the inverted peptide containing a free C-terminus, whereas the inner core would carry the corresponding encoding peptide with a free N-terminus (Figure 1). During library screening against a macromolecular target (e.g., a PDZ domain), which is too large to diffuse into the bead interior, only the inverted peptides on the bead surface would have access to the target. After a positive bead is identified from the library, the identity of the binding peptide would be determined by sequencing the encoding peptide on the bead by PED/MS (*45, 46*), a high-throughput peptide sequencing technique previously developed in our laboratory.

We tested this strategy on PDZ domains. Since PDZ domains generally make contacts with the four to five C-terminal residues of their partner proteins, we designed a peptide library containing five random residues at the C-terminus, resin-MLLBBERAX^{5}X^{4}X^{3}X^{2}X^1-COOH (library I), where B is β-alanine and X^1-X^5 represent L-α-aminobutyrate (Abu, 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 (Figure 1). Library I had a theoretical diversity of 20^5 (or 3 200 000) and was synthesized on TentaGel S resin (90 μm, ~2 860 000 beads/g, ~100 pmol peptide/bead). The terminal methionine permits peptide release by CNBr prior to MS analysis (vide infra). The two β-alanines were intended to provide some flexibility to the peptides to facilitate their binding to the PDZ domain. The two leucines were included to ensure that the smallest peptides would have an m/z ratio of >600 in MALDI mass spectra to avoid signal overlap with the MALDI matrix. Synthesis of library I started with the addition of the BBLLM linker to the TentaGel resin using standard Fmoc chemistry. Next, the beads were spatially segregated into outer and inner layers using the method of Lam (*49*), with the concomitant incorporation of Glu as a bifunctional linker. Briefly, TentaGel beads bearing the BBLLM linker were soaked in water, drained, and quickly suspended in 55:45 (v/v) dichloromethane/diethyl ether containing 0.33 equiv of a side chain *N*-hydroxysuccinimidyl (NHS) ester of L-glutamic acid, N_{α}-Fmoc-Glu(δ-NHS)-O-$CH_2CH{=}CH_2$. Because the organic solvent is immiscible with water, only peptides on the bead surface were exposed

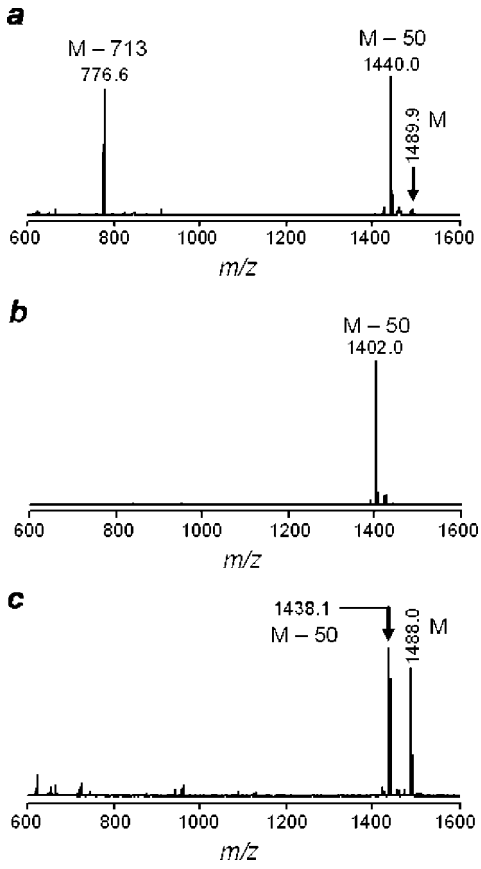

FIGURE 2: MALDI-TOF mass spectra of library I peptides at different stages of synthesis. (a) Prior to cyclization and no NaOH treatment; (b) NaOH treated uncyclized library; and (c) NaOH treated cyclized library.

to and reacted with the activated ester. The beads were washed with DMF, and the remaining free N-terminal amines in the inner core (0.67 equiv) were acylated with Boc-Gly OH. After the Fmoc group was removed from surface peptides, a *p*-hydroxymethylbenzoic acid linker was coupled to the N-terminal amine. The Boc protecting group was then removed from the internal peptides, and an Arg was added; the Arg provides a fixed positive charge to the encoding sequences and greatly facilitates MS analysis. The random region was synthesized by the split-and-pool method (*33–35*) to generate an OBOC library. A dipeptide Arg-Ala was added to the N-terminus; the Arg provides a positive charge to the surface peptides to facilitate MS analysis, whereas the Ala serves as a spacer to minimize any potential bias the positively charged Arg may exert on library screening. The α-allyl group on the C-terminal glutamate and the N-terminal Fmoc group were removed by treatment with $Pd(PPh_3)_4$ and piperidine, respectively. Subsequent treatment with PyBOP cyclized the surface peptides, while the encoding peptides in the bead interior were kept in the linear form. Finally, hydrolysis of the ester linkage with NaOH exposed the free C-termini of the surface peptides, which remained attached to the support via their N-termini (Figure 1). Note that any uncyclized surface peptides would be cleaved off the resin by the NaOH treatment, and therefore the bead surface should contain only the inverted peptides. Deprotection of side chains were carried out with TFA as usual.

Characterization of the Inverted Peptide Library. To assess the quality of the peptide library, resin beads from different stages of the library synthesis were subjected to MS analysis.

Table 1: MS Analysis of the Internal/Surface Peptide Ratio on 24 Randomly Selected Beads[a]

bead	M − 50 (encoding)	area	M (inverted)	area	ratio (inverted/encoding)
C1	1540.2	7022	1590.1	4623	0.66
C2	1523.4	2393	1573.3	566	0.24
C3	1360.4	5083	1410.4	4394	0.86
C4	1438.5	2436	1488.4	2060	0.85
C5	1365.5	5993	1415.4	877	0.15
C6	1579.4	1676	1629.3	692	0.41
C7	1294.5	8831	1344.5	693	0.08
C8	1493.1	3729	1543.1	2	0.00
C9	1510.5	3489	1560.4	131	0.04
C10	1435.6	5057	ND	0	0.00
C11	1435.6	4328	1485.5	3102	0.72
C12	1471.6	3243	1521.5	4484	1.38
C13	1558.5	10668	1608.4	6521	0.61
C14	1449.6	2160	1499.5	2535	1.17
C15	1416.6	5669	1466.5	172	0.03
C16	1614.5	5329	1664.4	153	0.03
C17	1441.5	8414	1491.5	2373	0.28
C18	1492.6	3681	1542.5	5483	1.49
C19	1525.5	6416	1575.4	1358	0.21
C20	1584.4	9229	1634.3	4670	0.51
C21	1529.3	2721	1579.3	2492	0.92
C22	1446.3	2479	1496.2	2413	0.97
C23	1531.2	4064	1581.2	309	0.08
C24	1634.2	2130	1684.1	1876	0.88
average					0.52

[a] ND, not detected by MALDI-TOF (threshold = signal/noise >3).

First, a small aliquot of resin (~2 mg) was withdrawn from the library immediately before the peptide cyclization step and treated with reagent K to remove side-chain protecting groups. Twenty-four beads were randomly selected, and the peptide from each bead was released by CNBr cleavage and analyzed by MALDI-TOF MS. Out of the 24 beads, 23 showed a pair of peaks separated by 50 amu (Figure 2a), indicating that both surface (*m*/*z* M, the pseudomolecular ion of the inverted peptide) and internal peptides (*m*/*z* M − 50) were successfully synthesized. For some of the beads, the *m*/*z* M peak had much lower abundance relative to the *m*/*z* (M – 50) peak. This was due to partial hydrolysis of the ester linkage of the surface peptides during overnight treatment with CNBr, which was dissolved in 70% TFA in water. The hydrolysis product, $RAX^5X^4X^3X^2X^1$, gave an intense peak at *m*/*z* (M – 713) (Figure 2a). Second, in a parallel experiment, 24 randomly selected beads were subjected to the same treatment as described above, except that the beads were treated with 1 N NaOH solution for 1 h prior to CNBr cleavage and MS analysis. The resulting MS spectrum of each bead showed a single peak in the expected *m*/*z* region for the encoding sequences (Figure 2b), consistent with complete hydrolysis (and therefore release) of the surface peptides. Third, 24 beads were randomly selected from the final library (after NaOH hydrolysis and deprotection with TFA), treated with CNBr, and analyzed by MS. The MS spectra of 23 beads showed a pair of peaks at *m*/*z* M and *m*/*z* (M – 50) (Figure 2c). Since any uncyclized surface peptides would have been cleaved off the resin by the NaOH treatment (Figure 2b), the *m*/*z* M peak in Figure 2c must be derived from the inverted peptides on the bead surface (Figure 1). The molar ratio of the inverted/encoding peptides on each bead was estimated from the relative abundance of the *m*/*z* (M – 50) and *m*/*z* M peaks, by assuming that the two peptides have equal ionization efficiency in the MS. The molar ratios of the 24 beads ranged

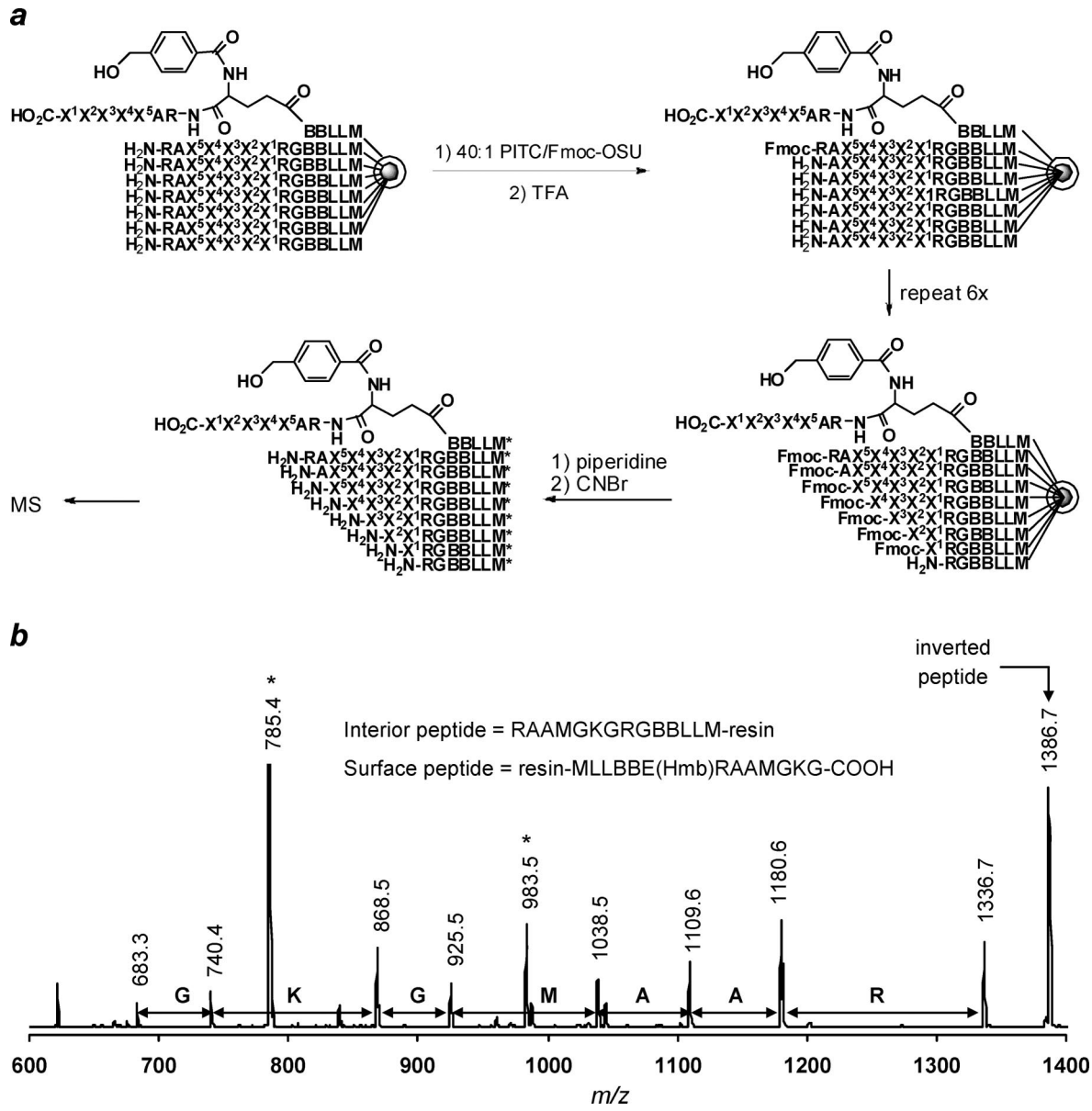

FIGURE 3: (a) Partial Edman degradation of resin-bound peptide. M*, homoserine lactone. (b) MALDI-TOF mass spectrum of the peptide and its degradation products from a randomly selected bead with the sequence RAAMGKG-CO_2H. *, Peptide fragments N-terminally acylated with CD_3CO- (Leu and Lys) or CH_3CD_2CO- group (Nle); they were derived from library synthesis to differentiate isobaric amino acid (*46*). Hmb, *p*-hydroxymethylbenzoyl.

from 0.00 to 1.49, but had an average value of 0.52 (the theoretical value is 0.50) (Table 1). Although for an individual bead, the estimated molar ratio may not be very accurate (due to potentially different ionization efficiency of the inverted vs encoding peptides), collectively the above data indicate that peptide cyclization and NaOH hydrolysis occurred at high yields and that substantial amounts of inverted peptides were present on each bead surface. We have previously shown that on-bead peptide N-to-C cyclization reaction is highly efficient (nearly quantitative) (*44*).

Peptide Sequencing by PED/MS. Fifty beads were randomly selected from library I and subjected to seven cycles of PED, (*45, 46*) which converted the encoding peptide on each bead into a series of progressively shorter peptides (Figure 3a). The inverted peptide on the surface was not affected by the PED reaction. Fourteen of the beads were randomly chosen and placed into individual microcentrifuge tubes, and the peptides were released by cleavage with CNBr and analyzed by MALDI-TOF MS. Out of the 14 beads, 13 (93%) produced spectra of sufficient quality to allow unambiguous sequence assignment, while one bead failed to show any signals, possibly due to poor sample preparation. Figure 3b shows a representative MS spectrum derived from one of the 14 beads; the other 13 spectra are provided in Supporting Information (Figure S1). The inverted peptide produced an intense peak at *m/z* 1386.7, whereas the full-length encoding peptide generated a peak at *m/z* 1336.7. The peptide fragments derived from PED gave a series of peaks at *m/z* 1180.6, 1109.6, 1038.5, 925.5, 868.5, 740.4, and 683.3. From the masses of this peptide ladder, the sequence of the encoding peptide was determined as RAAMGKGRGBBLLM-resin, indicating that the inverted peptide had the sequence of resin-MLLBBE(hydroxymethylbenzoyl)RAAMGKG-COOH. All together, we sequenced 423 beads in this work and obtained 376 unambiguous sequences (89% success rate) (Table 2). The rest of the beads (11%) either had poor

Table 2: Success Rate for Sequencing Resin-Bound Peptides by PED/MS

trial no.	screening target	no. of beads analyzed by PED/MS	no. of complete sequences obtained
1	none	14	13 (93%)
2	NHERF-PDZ1	84	67 (80%)
3	CIPP-PDZ2	93	82 (88%)
4	CIPP-PDZ3	85	80 (94%)
5	CIPP-PDZ4	147	134 (91%)
total (average)		423	376 (89%)

Table 3: Selected Sequences for PDZ1 domain from NHERF1 (67 total)[a]

Class I			Class II
YDCRF	WSNRF	YTCML	HHQHP
LICRF	TKNRF	QTARM	RKHFH
DWCRF	NWSRF	KDCRM	VHGHR
AHCRF	RHSHF	LDIRM	VHKYH
NYERF	AFTYL	QWSRM	TRHIH
MSHYF	CRTYL	VFHRM	LHNNH
GNTYF	FNCRL	VAHRM	HKHKL
CSTYF	AGCRL	VYTRM	YKHRH
YHTSF	RRFRL	RHSVM	MHYIH
MRATF	FPHRL	YSSYM	KHFHT
RISKF	LHNRL	IHHRM	
LKTLF	QDSRL	HWTAM	
YWGRF	TTSRL	WITKM	
WFHRF	IGTRL	YYCKM	
YWHRF	QWSAL	HTSRI	
LHTCF	FHTAL	QYTRI	
YCHRF	RYTAL	QCTRC	
YVHRF	LITHL	TRFRP	
HIKRF	DFTIL	NQTRV	

[a] C, (*S*)-2-aminobutyric acid; M, norleucine.

ionization in MS or missed one or more of the ladder peaks, preventing complete sequence assignment.

Identification of Binding Ligands for PDZ Domains. We chose to test the validity of our library method on the PDZ1 domain of NHERF1. NHERF1 and its homologue NHERF2 represent a family of adaptor proteins characterized by the presence of two N-terminal PDZ domains and a C-terminal domain that binds to cytoskeleton proteins ezrin, radixin, and moesin (*50*). The PDZ domains of NHERF1 and NHERF2 bind to an array of transmembrane and soluble proteins, including ion channels, transcription factors, and cell surface receptors. A previous library study has shown that the NHERF1 PDZ1 domain binds to peptides of the consensus eX(T/s)(R/y)(L/f)-COOH, where X is any amino acid and the lower-case letters represent less frequently selected amino acids (*24*). To avoid any potential bias during library screening exerted by the positively charged Arg at position –6 (relative to the C-terminal residue, which is defined as position 0) in library I, we synthesized another library (library II), resin-MLLBBEAAX^{5}X^{4}X^{3}X^{2}X^1-COOH, in which the arginine was replaced by an alanine. Screening of a total of 20 mg of library II (~60 000 beads) produced 84 positive beads, which were sequenced by PED/MS to give 67 complete sequences (Table 3). Among these sequences, 57 clearly belong to one class (class I) with a consensus motif of XX(T/s/c/h)(R/y)(F/L/M)-COOH (Figure 4). Thus, our results are in excellent agreement with the earlier study, in which a genetically encoded peptide library was screened against the PDZ1 domain (*24*). The 10 class II peptides each contained at least two histidine residues (Table 3). Since

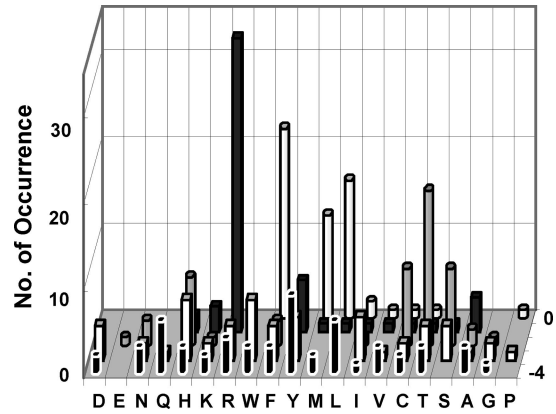

FIGURE 4: Sequence specificity of NHERF1 PDZ1 domain. Displayed are the amino acids identified at each position (−4 to 0). Number of occurrence on the *y* axis represents the number of selected sequences that contained a particular amino acid at a certain position.

similar sequences were also obtained from the screening experiments against other PDZ domains and later binding assays showed that they had no detectable binding affinity to the PDZ domains (vide infra), we conclude that the class II peptides were derived from nonspecific binding of unknown origin.

We next applied the library method to determine the sequence specificity of CIPP PDZ domains (*51*). CIPP is a 612-amino acid scaffolding protein containing four PDZ domains. It has been shown to bind to the C-termini of several ion channel proteins and cell surface receptors (*51, 52*), but otherwise little is known about the specificity of its PDZ domains. We expressed the four PDZ domains individually in *E. coli*; however, only PDZ2, PDZ3, and PDZ4 domains gave soluble proteins. Thus, library II (100 mg) was screened against each of the three PDZ domains (at 0.5−1.0 μM), resulting in 82, 80, and 134 binding sequences, respectively (Tables S1, S2, and S3 in Supporting Information). We have previously shown that the recognition motif(s) of a protein domain can be determined unambiguously by screening only a fraction of the library, and we usually screen 10% of a complete library (*6, 14*). Inspection of the selected sequences revealed that PDZ2 and PDZ3 domains have similar specificities, with consensus sequences of XX(S/h)(Y/R)V-COOH and XX(y/s/h/v/i/l/t)(Y/f/m/l)V-COOH, respectively (Figure 5). A notable difference between these two domains is the preference for an Arg at the –2 position by PDZ2 but not by PDZ3 domain. PDZ4 domain has a very different specificity profile. Unlike the stringent requirement for a Val at position 0 by PDZ2 and PDZ3 domains, PDZ4 domain accepts Ile, Leu, Val, and Nle at this position. On the other hand, PDZ4 domain has a more stringent requirement for a Thr (and to a less extent Ser) at the –2 position as compared to PDZ2 and PDZ3 domain (Figure 5). At positions –3 and –4, PDZ4 domain has a general preference for hydrophobic residues, whereas PDZ2 and PDZ3 domains prefer hydrophilic and small amino acids. Thus, PDZ4 domain has a consensus of $\varphi\varphi$(T/s)X(I/L/V/M), where φ represents hydrophobic amino acids. Consistent with our findings, acid-sensitive ion channel 3 (ASIC3), which has a C-terminal sequence of LVTRL-COOH, binds selectively to the PDZ4 domain of CIPP (*52*), whereas potassium ion channels Kir4.1 (which has a C-terminal sequence RISNV-COOH) and Kir4.2 (QQSNV-

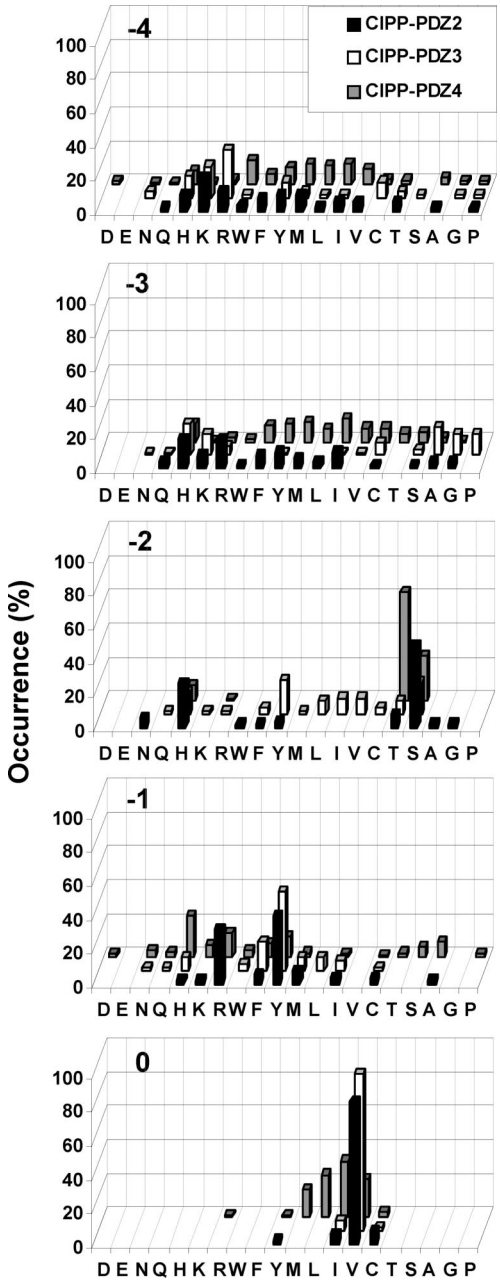

FIGURE 5: Sequence specificity of CIPP PDZ domains. Displayed are the amino acids identified at each position (−4 to 0). Occurrence on the *y* axis represents the percentage of selected sequences that contained a particular amino acid at a certain position.

COOH) and glutamate receptors NR2A (IESDV-COOH), NR2B (IESDV-COOH), NR2C (LESEV-COOH), and NR2D (LESEV-COOH) bind to the PDZ2 and/or PDZ3 domains (*51*).

Binding Affinity between Selected Peptides and CIPP PDZ Domains. To confirm the on-bead screening results, two representative peptides selected against each of the CIPP PDZ domains were synthesized, and their dissociation constants against the three PDZ domains were determined by surface plasmon resonance (SPR). The PDZ domains were biotinylated and immobilized onto a streptavidin-coated sensorchip. Binding analysis was carried out by flowing different concentrations of the synthetic peptides (0−1500 μM) over the sensorchip. Two peptides selected against the PDZ4 domain, YFTRL and VVTHI, bound to the PDZ4 domain with high affinities (K_D = 1.7 and 3.1 μM, respectively) but had no detectable binding toward the PDZ2 or PDZ3 domain (Table 4). Likewise, the two peptides selected against the PDZ3 domain (KHYYV and KKVYV) bound selectively to the PDZ3 domain, although the binding affinities (K_D = 18 and 32 μM, respectively) were an order of magnitude lower than that of PDZ4 and its cognate ligand (see Figure S2 in Supporting Information for actual sensograms). Interestingly, the peptides selected against the PDZ2 domain (VFHRV and YLSRV) bound weakly to all three PDZ domains (K_D = 70−2400 μM). Finally, we tested two of the class II peptides (RLHRH and HRHIR) and found that they had no significant binding affinity for any of the PDZ domains. It has previously been reported that PDZ domains generally have lower affinity to their cognate ligands, as compared to other modular domains (e.g., Src homology 2 domains), and our results are well within the range of K_D values reported earlier (*21, 29, 41, 42*).

Table 4: Dissociation Constants (K_D) of Selected Peptides toward CIPP PDZ Domains

peptide[a]	K_D (μM) PDZ2	PDZ3	PDZ4
YFTRL-OH	NB	NB	1.7 ± 0.1[b]
VVTHI-OH	NB	NB	3.1 ± 1.7[b]
KHYYV-OH	NB	19 ± 1[b]	NB
KKVYV-OH	NB	32 ± 2[b]	NB
VFHRV-OH	2400 ± 1200[b]	580 ± 60	1100 ± 400
YLSRV-OH	630 ± 540[b]	1200 ± 460	70 ± 40
RLHRH-OH	NB	NB	NB
HRHIR-OH	NB	NB	NB

[a] All peptides contained an Ac-YAA or Ac-YSS at their N-termini. [b] The PDZ domain to which the peptide was selected; NB, no significant binding was detected.

Database Search of Potential CIPP-Binding Proteins. We searched Protein Information Resource database (Web site: http://pir.georgetown.edu/) for potential CIPP-binding proteins. We restricted our search to mouse proteins for simplicity. For the PDZ2 domain, motif [HTS][RFY][V]> (where ">" restricts the search to protein C-terminus) was used for the pattern search. Similarly, motifs [HYLIVTS][HFYLI][V]> and [GAVLIMPFYW][GAVLIMPFYW][ST]X[IVML]> were used for the PDZ3 and PDZ4 domains, respectively. These searches recovered 15, 41, and 104 potential targets for the PDZ2 (Table 5), PDZ3 (Table 5), and PDZ4 (Table S4 in Supporting Information), respectively, after eliminating any redundant, fragment, or hypothetical proteins. Among the predicted targets for PDZ2 and PDZ3 domains, neuroligin 3 and Nrxn3 proteins have previously been shown to interact with the PDZ domains (*51*). For the PDZ4 domain, as expected, acid sensitive ion channel 3, a known PDZ4 domain-binding protein (*52*), was recovered from the database search. It is highly probable that some of the other predicted binding proteins, especially the cell surface receptors, adhesion molecules, and channel proteins, will prove to be bona fide CIPP-interacting proteins. A few previously reported CIPP-binding proteins [e.g., glutamate receptors (*51*)] were not recovered by the database search, because their C-terminal sequences contain an acidic amino acid at the –1 position (Asp or Glu), which were not preferred by the PDZ2 or PDZ3 domain. On the basis of our results, we predict that the C-termini of these proteins (IESDV and LESEV) would bind only very weakly to CIPP PDZ2 and/or PDZ3 domains. It is possible that CIPP is engaged in additional interactions with more upstream regions of the receptor proteins. Alterna-

Table 5: Potential CIPP-Binding Proteins from Database Search

	C-terminal Sequence	ID no.	protein
pdz2	CCSFV	Q9JKM7	Ras-related protein Rab-37
	EESFV	Q05CF3	Insc protein
	EWTRV	Q61152	tyrosine-protein phosphatase nonreceptor type 18 (EC 3.1.3.48)
	FLSFV	Q8K4E5	2′−5′ oligoadenylate synthetase 2
	FSTYV	A2CG52	Kalirin, RhoGEF kinase
	KPSRV	Q8CIV5	testis-specific leucine zipper protein nurit
	LITRV	A2RSL2	nicrorchidia 4
	PGTYV	Q3TEL6	RING finger protein 157
	PNTFV	Q61121	probable G-protein coupled receptor 19
	RITFV	Q3UDI8	minichromosome maintenance deficient 7
	SHHYV	Q8BKW4	zinc finger CCHC domain-containing protein 4
	STSYV	Q8BXA6	claudin-17
	STTRV	A2AGI2	neuroligin 3[a]
	VDTRV	Q925N4	claudin-16 (paracellin-1)
	YNTFV	Q8K2K3	E130014J05Rik protein
pdz3	AAVFV	Q4QQL3	cathepsin O
	AKVFV	P58404	striatin-4 (zinedin)
	EESFV	Q05CF3	Insc protein
	EFIFV	Q6EDY6	CARMIL
	EGIYV	Q8CJ96	Ras association domain-containing protein 8
	ERLLV	Q9D0T1	NHP2-like protein 1 (high mobility group-like nuclear protein 2 homologue 1)
	FHTLV	Q9D0P8	putative GTP-binding protein RAY-like (Rab-like protein 4)
	FLSFV	Q8K4E5	2′−5′ oligoadenylate synthetase 2
	FSTYV	A2CG53	kalirin, RhoGEF kinase
	HWLLV	Q61327	sodium-dependent dopamine transporter (DA transporter) (DAT)
	IASIV	Q7TRG2	olfactory receptor Olfr845 (olfactory receptor 845)
	KESYV	P41233	ATP-binding cassette subfamily A member 1 (ATP-binding cassette transporter 1)
	KEYYV	Q6P9K9	Nrxn3 protein[a]
	KGIHV	Q80TY4	suppression of tumorigenicity protein 18
	KTSHV	Q8CI59	metalloreductase STEAP3 (EC 1.16.1.-)
	KVSHV	Q8C739	protein FAM110B
	LQILV	A2ANS3	heat shock protein 12B
	LWLFV	Q0VE84	UDP-*N*-acetyl-alpha-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase 6
	MGLLV	Q61235	beta-2-syntrophin (59 kDa dystrophin-associated protein A1 basic component 2)
	NGLLV	P19253	60S ribosomal protein L13a (Transplantation antigen P198) (Tum-P198 antigen)
	NVTLV	Q7TSN7	cell adhesion molecule JCAM
	PASHV	Q06054	zinc finger protein
	PETLV	O54834	Rho GTPase-activating protein 6 (Rho-type GTPase-activating protein RhoGAPX-1)
	PGSLV	Q148W7	Lnx2 protein
	PGTYV	Q3TEL6	RING finger protein 157
	PNTFV	Q61121	probable G-protein coupled receptor 19
	QGTLV	P01741	Ig heavy chain V region (anti-arsonate antibody)
	QKILV	A2ASU5	olfactory receptor 1022
	QTSIV	Q80WG6	UROP11–110
	QYSFV	Q91ZM3	Pro-rich, PH and SH2 domain-containing signaling mediator alpha

[a] Proteins that have been shown to interact with corresponding PDZ domains (*51*).

tively, the interaction between CIPP and its partner proteins may involve oligomerization of the binding partners.

DISCUSSION

A variety of library methods have previously been applied to determine the recognition motifs of PDZ domains. One of the biological methods employed was phage display, which involved the fusion of a random peptide sequence to the C-terminus of bacteriophage gene VIII product (*21, 22*). After affinity purification, the identity of a binding peptide was identified by sequencing the phage DNA. In another analogous system, random sequences were fused to the C-terminus of *LacI* repressor protein, which bound to its cognate plasmid DNA with extremely high affinity (*23, 24*). Affinity purification against the PDZ domain retained repressor proteins that contained high-affinity C-terminal peptides and the corresponding plasmid DNAs encoding the repressor proteins. Subsequent sequencing of the DNAs revealed the identity of the binding peptides. In a FRET assay, a PDZ domain and a library of random peptides were fused to a pair of fluorescent donor and acceptor proteins (*25*). Binding between the PDZ domain and a peptide resulted in fluorescence energy transfer, and the cells containing the interacting pairs were sorted by FACS. Finally, the yeast two-hybrid screen has also been adapted to identify PDZ-binding proteins (*43*). These biological methods can generate libraries of high diversity (generally up to 10^9, which is limited by bacterial transformation efficiency) and identify individual binding sequences. Further, in the case of phage display and *LacI* repressor fusion, the libraries may be iteratively amplified and screened to converge on the most potent ligands. Their main limitation is that the libraries can only contain the 20 proteinogenic amino acids; this is adequate for PDZ domains, but not for other proteins that require modified amino acids for binding and/or catalysis (vide infra). Among the chemical methods, Cantley's solution-phase library screening coupled with pooled sequencing gives information on the relative preference for certain amino acids at a given position, but not individual sequences (*27*). The method also selects for both affinity and abundance and may overlook a minor class of tight-binding ligands. Its main advantage is its simplicity, which permits high-throughput

analysis of a large number of PDZ domains in a short period of time. Boisguerin et al. synthesized inverted peptides corresponding to the C-termini of 6223 human proteins on cellulose membranes and screened the SPOT library against Erbin PDZ domain (*29, 30*). Since the peptides are derived directly from human proteins, this method is excellent in predicting the protein partners of PDZ domains. However, due to its limited (10^3-10^4) and biased sequence coverage [e.g., some C-terminal sequences render proteins unstable and are underrepresented in the genome (*53, 54*)], the SPOT analysis did not reveal the complete specificity profile of the PDZ domain. It also required a robotic system to synthesize the library and is thus not accessible to an average biochemical laboratory. MacBeath and co-workers developed a protein microarray containing 157 mouse PDZ domains and measured their binding affinities to 217 genome-encoded peptides (*42*). The data set was then used to "train" the computer to predict the overall specificity profile of a PDZ domain. The major advantage of the microarray method is that quantitative measurements can be made in a high-throughput manner and it has the potential to provide a global comparison of the specificity profiles of all PDZ domains in the genome. The drawback of the approach is that the predicted specificity profiles depend on the choice of the initial data set; for PDZ domains without any known ligands, it is difficult to select the initial set of peptides for testing. The 217 C-terminal peptides selected in the recent report (*42*) represented only a small fraction of mouse protein C-termini and a very small fraction of all possible sequence space.

We have now developed a general methodology to synthesize and screen OBOC combinatorial peptide libraries containing free C-termini and demonstrated its validity by determining the recognition motifs of four different PDZ domains. The new method was built upon three previously developed concepts/techniques: peptide inversion to generate support-bound peptides with free C-termini, spatial segregation of microbeads to separate library molecules from encoding tags, and high-throughput peptide sequencing by PED/MS. Because our libraries are chemically synthesized, posttranslationally modified amino acids and unnatural building blocks may be readily accommodated. This feature will be useful for screening other proteins that require modified amino acids for binding/catalysis or developing metabolically stable inhibitors against PDZ domains. Theoretically, OBOC libraries may be synthesized with infinite diversity; in practice, however, they are limited by the amount of resin that can be conveniently handled in a research laboratory. For the most popular TentaGel microbeads (90 μm) used in this work, the library size is limited to 10^7 members (corresponding to ~4 g of resin). Although not yet demonstrated, our system should be compatible with 30-μm TentaGel beads; this would allow one to generate OBOC libraries of 10^9 members. Since most PDZ domains recognize only four to five C-terminal residues, these chemical libraries can readily cover the entire sequence space of these domains and are therefore able to determine their complete specificity profiles. Our method provides individual binding sequences, from which a consensus sequence(s) is readily derived. Furthermore, our method does not require any special equipment, is relatively inexpensive, and can be performed in any biochemical laboratory.

Our method should be readily applicable to any protein/enzyme that recognizes the C-terminus of a peptide or protein. In addition to PDZ domains, 14-3-3 proteins have been shown to recognize the C-termini of their partner proteins (*55*). Unlike the PDZ domains, however, binding by the 14-3-3 proteins strictly requires phosphorylation of Ser or Thr at the C-termini (*56*). To our knowledge, little is currently known about the specificity of these proteins (in terms of C-terminal binding), and no library method has been developed for this class of proteins. There are also a large number of enzymes that catalyze the modification of protein C-terminus, including C-terminal specific proteases (*53, 54*) and C-terminal lipidation enzymes (*57, 58*). Application of our library method to these challenging targets is already underway in this laboratory.

ACKNOWLEDGMENT

We thank Dr. Mike Zhu for providing the plasmids that encode the PDZ domain proteins.

SUPPORTING INFORMATION AVAILABLE

MS spectra, SPR sensograms, and peptide sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Chen, J. K., Lane, W. S., Brauer, A. W., Tanaka, A., and Schreiber, S. L. (1993) Biased combinatorial libraries: Novel ligands for the SH3 domain of phosphatidylinositol 3-kinase. *J. Am. Chem. Soc. 115*, 12591–12592.
2. Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993) SH2 domains recognize specific phosphopeptide sequences. *Cell 72*, 767–778.
3. Muller, K., Gombert, F. O., Manning, U., Grossmuller, F., Graff, P., Zaegel, H., Zuber, J. F., Freuler, F., Tschopp, C., and Baumann, G. (1996) Rapid identification of phosphopeptide ligands for SH2 domains: Screening of peptide libraries by fluorescence-activated bead sorting. *J. Biol. Chem. 271*, 16500–16505.
4. Dooley, C. T., Ny, P., Bidlack, J. M., and Houghten, R. A. (1998) Selective ligands for the mu, delta, and kappa opioid receptors identified from a single mixture based tetrapeptide positional scanning combinatorial library. *J. Biol. Chem. 273*, 18848–18856.
5. Smith, H. K., and Bradley, M. (1999) Comparison of resin and solution screening methodologies in combinatorial chemistry and the identification of a 100 nM inhibitor of trypanothione reductase. *J. Comb. Chem. 1*, 326–332.
6. Sweeney, M. C., Wavreille, A. S., Park, J., Butchar, J. P., Tridandapani, S., and Pei, D. (2005) Decoding protein-protein interactions through combinatorial chemistry: Sequence specificity of SHP-1, SHP-2 and SHIP SH2 domains. *Biochemistry 44*, 14932–14947.
7. Peng, L., Liu, R. W., Marik, J., Wang, X. B., Takada, Y., and Lam, K. S. (2006) Combinatorial chemistry identifies high-affinity peptidomimetics against alpha4 beta1 integrin for in vivo tumor imaging. *Nat. Chem. Biol. 2*, 381–389.
8. Matthews, D. J., and Wells, J. A. (1993) Substrate phage: Selection of protease substrates by monovalent phage display. *Science 260*, 1113–1117.
9. Meldal, M., Svendsen, I., Breddam, K., and Auzanneau, F.-I. (1994) Portion-mixing peptide libraries of quenched fluorogenic substrates for complete subsite mapping of endoprotease specificity. *Proc. Natl. Acad. Sci. U. S. A. 91*, 3314–3318.
10. Peterson, J. J., and Meares, C. F. (1998) Cathepsin substrates as cleavable peptide linkers in bioconjugates, selected from a fluorescence quench combinatorial library. *Bioconjugate Chem. 9*, 618–626.

11. Hu, Y. J., Wei, Y., Zhou, Y., Rajagopalan, P. T., and Pei, D. (1999) Determination of substrate specificity for peptide deformylase through the screening of a combinatorial peptide library. *Biochemistry 38*, 643–650.
12. Wu, J., Ma, Q. N., and Lam, K. S. (1994) Identifying substrate motifs of protein kinases by a random library approach. *Biochemistry 33*, 14825–14833.
13. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. U. S. A. 97*, 7754–7759.
14. Garaud, M., and Pei, D. (2007) Substrate profiling of protein tyrosine phosphatase PTP1B by screening a combinatorial peptide library. *J. Am. Chem. Soc. 129*, 5366–5367.
15. Miller, S. J., Copeland, G. T., Papaioannou, N., Horstmann, T. E., and Ruel, E. M. (1998) Kinetic resolution of alcohols catalyzed by tripeptides containing the N-alkylimidazole substructure. *J. Am. Chem. Soc. 120*, 1629–1630.
16. Copeland, G. T., Jarvo, E. R., and Miller, S. J. (1998) Minimal acylase-like peptides: Conformational control of absolute stereospecificity. *J. Org. Chem. 63*, 6784–6785.
17. Berkessel, A., and Herault, D. A. (1999) Discovery of peptide-zirconium complexes that mediate phosphate hydrolysis by batch screening of a combinatorial undecapeptide library. *Angew. Chem., Int. Ed. 38*, 102–105.
18. Francis, M. B., Jamison, T. F., and Jacobsen, E. N. (1998) Combinatorial libraries of transition-metal complexes, catalysts and materials. *Curr. Opin. Chem. Biol. 2*, 422–428.
19. Hoveyda, A. H. (1998) Catalyst discovery through combinatorial chemistry. *Chem. Biol. 5*, R187–R191.
20. Copeland, G. T., and Miller, S. J. (2001) Selection of enantioselective acyl transfer catalysts from a pooled peptide library through a fluorescence-based activity assay: An approach to kinetic resolution of secondary alcohols of broad structural scope. *J. Am. Chem. Soc. 123*, 6496–6502.
21. Fuh, G., Pisabarro, M. T., Li, Y., Quan, C., Lasky, L. A., and Sidhu, S. S. (2000) Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J. Biol. Chem. 275*, 21486–21491.
22. Laura, R. P., Witt, A. S., Held, H. A., Gerstner, R., Deshayes, K., Koehler, M. F., Kosik, K. S., Sidhu, S. S., and Lasky, L. A. (2002) The Erbin PDZ domain binds with high affinity and specificity to the carboxyl termini of delta-catenin and ARVCF. *J. Biol. Chem. 277*, 12906–12914.
23. Stricker, N. L., Christopherson, K. S., Yi, B. A., Schatz, P. J., Raab, R. W., Dawes, G., Bassett, D. E., Jr., Bredt, D. S., and Li, M. (1997) PDZ domain of neuronal nitric oxide synthase recognizes novel C-terminal peptide sequences. *Nat. Biotechnol. 15*, 336–342.
24. Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B., and Li, M. (1998) Peptide binding consensus of the NHE-RF-PDZ1 domain matches the C-terminal sequence of cystic fibrosis transmembrane conductance regulator (CFTR). *FEBS Lett. 427*, 103–108.
25. You, X., Nguyen, A. W., Jabaiah, A., Sheff, M. A., Thorn, K. S., and Daugherty, P. S. (2006) Intracellular protein interaction mapping with FRET hybrids. *Proc. Natl. Acad. Sci. U. S. A. 103*, 18458–18463.
26. Alsina, J., and Albericio, F. (2003) Solid-phase synthesis of C-terminal modified peptides. *Biopolymers 71*, 454–477.
27. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science 275*, 73–77.
28. Kania, R. S., Zuckermann, R. N., and Marlowe, C. K. (1994) Free C-terminal resin-bound peptides: Reversal of peptide orientation via a cyclization/cleavage protocol. *J. Am. Chem. Soc. 116*, 8835–8836.
29. Hoffmuller, U., Russwurm, M., Kleinjung, F., Ashurst, J., Oschkinat, H., Volkmer-Engert, R., Koesling, D., and Schneider-Mergener, J. (1999) Interaction of a PDZ protein domain with a synthetic library of all human protein C-termini. *Angew. Chem., Int. Ed. 38*, 2000–2004.
30. Boisguerin, P., Leben, R., Ay, B., Radziwill, G., Moelling, K., Dong, L., and Volkmer-Engert, R. (2004) An improved method for the synthesis of cellulose membrane-bound peptides with free C termini is useful for PDZ domain binding studies. *Chem. Biol. 11*, 449–459.
31. Davies, M. (1997) C-terminally modified peptides and peptide libraries-another end to peptide synthesis. *Angew. Chem., Int. Ed. 36*, 1097–1099.
32. Davies, M., Bonnat, M., Guillier, F., Kilburn, J. D., and Bradley, M. (1998) Screening an inverted peptide library in water with a guanidinium-based tweezer receptor. *J. Org. Chem. 63*, 8696–8703.
33. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) A new type of synthetic peptide library for identifying ligand-binding activity. *Nature 354*, 82–84.
34. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Pept. Protein Res. 37*, 487–493.
35. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., and Cuervo, J. H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature 354*, 84–86.
36. Hung, A. Y., and Sheng, M. (2002) PDZ domains: Structural modules for protein complex assembly. *J. Biol. Chem. 277*, 5699–5702.
37. Nourry, C., Grant, S. G., and Borg, J. P. (2003) PDZ domain proteins: Plug and play! *Sci. STKE 2003*, RE7.
38. Morais Cabral, J. H., Petosa, C., Sutcliffe, M. J., Raza, S., Byron, O., Poy, F., Marfatia, S. M., Chishti, A. H., and Liddington, R. C. (1996) Crystal structure of a PDZ domain. *Nature 382*, 649–652.
39. Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996) Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ. *Cell 85*, 1067–1076.
40. Basdevant, N., Weinstein, H., and Ceruso, M. (2006) Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. *J. Am. Chem. Soc. 128*, 12766–12777.
41. Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R., and Oschkinat, H. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol. 343*, 703–718.
42. Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A., and MacBeath, G. (2007) PDZ domain binding selectivity is optimized across the mouse proteome. *Science 317*, 364–369.
43. Song, E., Gao, S., Tian, R., Ma, S., Huang, H., Guo, J., Li, Y., Zhang, L., and Gao, Y. (2006) A high efficiency strategy for binding property characterization of peptide-binding domains. *Mol. Cell. Proteomics 5*, 1368–1381.
44. Joo, S. H., Xiao, Q., Ling, Y., Gopishetty, B., and Pei, D. (2006) High-throughput sequence determination of cyclic peptide library members by partial Edman degradation/mass spectrometry. *J. Am. Chem. Soc. 128*, 13000–13009.
45. Sweeney, M. C., and Pei, D. (2003) An improved method for rapid sequencing of support-bound peptides by partial Edman degradation and mass spectrometry. *J. Comb. Chem. 5*, 218–222.
46. Thakkar, A., Wavreille, A. S., and Pei, D. (2006) Traceless capping agent for peptide sequencing by partial Edman degradation and mass spectrometry. *Anal. Chem. 78*, 5935–5939.
47. Wavreille, A. S., Garaud, M., Zhang, Y., and Pei, D. (2007) Defining SH2 domain and PTP specificity by screening combinatorial peptide libraries. *Methods 42*, 207–219.
48. Yin, J., Straight, P. D., McLoughlin, S. M., Zhou, Z., Lin, A. J., Golan, D. E., Kelleher, N. L., Kolter, R., and Walsh, C. T. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U. S. A. 102*, 15815–15820.
49. Liu, R. M., Lam, J., and Kit, S. (2002) A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries. *J. Am. Chem. Soc. 124*, 7678–7680.
50. Weinman, E. J., Hall, R. A., Friedman, P. A., Liu-Chen, L. Y., and Shenolikar, S. (2006) The association of NHERF adaptor proteins with g protein-coupled receptors and receptor tyrosine kinases. *Annu. Rev. Physiol. 68*, 491–505.
51. Kurschner, C., Mermelstein, P. G., Holden, W. T., and Surmeier, D. J. (1998) CIPP, a novel multivalent PDZ domain protein, selectively interacts with Kir4.0 family members, NMDA receptor subunits, neurexins, and neuroligins. *Mol. Cell. Neurosci. 11*, 161–172.
52. Anzai, N., Deval, E., Schaefer, L., Friend, V., Lazdunski, M., and Lingueglia, E. (2002) The multivalent PDZ domain-containing protein CIPP is a partner of acid-sensing ion channel 3 in sensory neurons. *J. Biol. Chem. 277*, 16655–16661.

53. Keiler, K. C., Waller, P. R., and Sauer, R. T. (1996) Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science 271*, 990–993.
54. Pallen, M. J., and Wren, B. W. (1997) The HtrA family of serine proteases. *Mol. Microbiol. 26*, 209–221.
55. Coblitz, B., Wu, M., Shikano, S., and Li, M. (2006) C-terminal binding: An expanded repertoire and function of 14–3-3 proteins. *FEBS Lett. 580*, 1531–1535.
56. Ganguly, S., Weller, J. L., Ho, A., Chemineau, P., Malpaux, B., and Klein, D. C. (2005) Melatonin synthesis: 14–3-3-dependent activation and inhibition of arylalkylamine N-acetyltransferase mediated by phosphoserine-205. *Proc. Natl. Acad. Sci. U. S. A. 102*, 1222–1227.
57. Zhang, F. L., and Casey, P. J. (1996) Protein prenylation: Molecular mechanisms and functional consequences. *Annu. Rev. Biochem. 65*, 241–269.
58. Cook, T. A., Ghomashchi, F., Gelb, M. H., Florio, S. K., and Beavo, J. A. (2000) Binding of the delta subunit to rod phosphodiesterase catalytic subunits requires methylated, prenylated C-termini of the catalytic subunits. *Biochemistry 39*, 13516–13523.

BI7023628