# 3 Modification of the SPOT Synthesis Technique to Produce Peptides with Free C-Termini for the Study of PDZ Domain Interactions

## 3.1 Peptide Libraries for Proteomics Studies

Many combinatorial peptide library-based approaches have been developed to analyze the binding or substrate specificities of protein domains (reviewed in [131]). These libraries can be classified as either 'display' or 'pool peptide libraries'. The display libraries employ a variety of carriers for individual peptide presentation, including pins, phages, beads, DNA binding proteins or ribosomes. The display techniques involve multiple rounds of enrichment steps, and the identities of the selected peptides can be ascertained by DNA or peptide sequencing. Display peptide libraries generally require isolation and sequencing of individual peptides or the respective DNA to decode the binding motif, which is time-consuming and costly. In contrast, the pool peptide libraries rely on the comparison of binding affinities of different pools of randomly chosen peptides, in search for the binding consensus motif. In the oriented peptide library method [132], soluble pools of random peptides are 'oriented' towards a certain peptide motif *via* a central 'fixed' amino acid (e.g.: $xxxB_1B_2xxx$). The soluble oriented peptide library approach still depends on peptide sequencing, and large amounts of purified proteins are generally needed for the assay.

Recently, peptide arrays established by SPOT synthesis are reported to be useful to map specificity, for example of antibodies, protein domains and protein kinases [133-135].

### 3.1.1 SPOT Synthesis

SPOT synthesis is an easy-to-handle and flexible technique for parallel chemical synthesis on membrane supports (for detailed experimental description see [112, 113]). The most frequent application is the synthesis of peptides on cellulose membranes, which initially involves attachment of 9-fluorenyl-methoxycarbonyl-β-alanine (Fmoc-β-Ala-OH) to the hydroxyl functions of a commercially available filter paper (e.g. Whatman, Maidsone, UK). Fmoc-β-Ala-OH, other 'amino compounds' are used to generate either membrane modified with aminopropylether [136] or N-modified cellulose-amino-hydroxypropyl ether

(N-CAPE) [111, 137] as support. After the subsequent removal of the Fmoc-group, the free amino group is then available for the next steps in the synthesis.

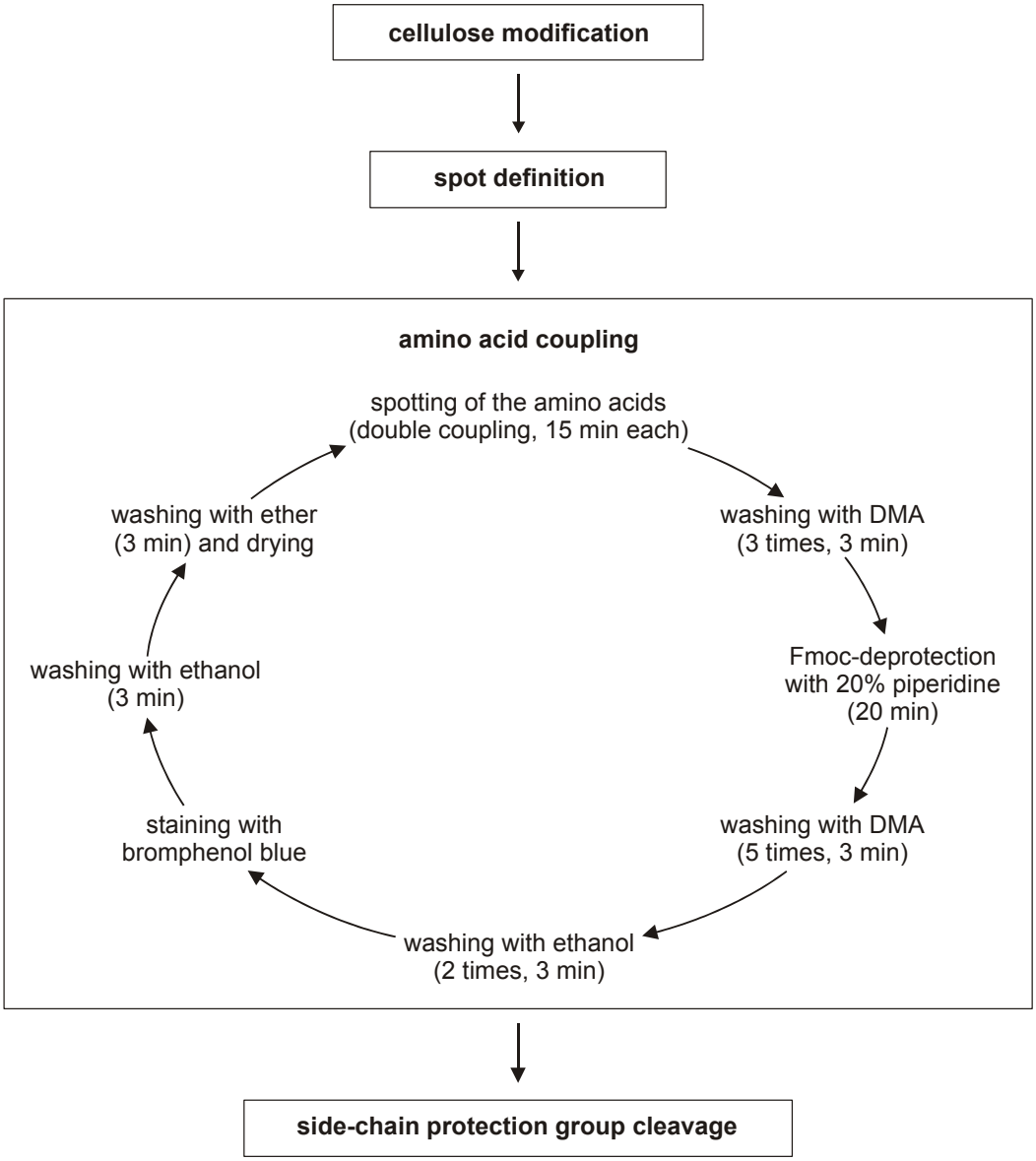


**Figure 3.1 Outline of the Standard SPOT Synthesis Procedure.**
*Cellulose modification:* The surface of the cellulose membrane is first uniformly converted to an 'amino'-group exposing support employing different 'amino-compounds' to accessible hydroxyl functions on a cellulose membrane.
*Spot definition:* In the second derivatisation step, the array of spot reactors is generated by spotwise coupling of Fmoc-β-Ala-OPfp, and all residual amino functions in between spots are blocked by acetylation.
*Amino acid coupling:* 0.2-0.3 M solutions of protected amino acid derivatives are coupled twice. After several washing steps, the Fmoc-group is cleaved with a piperidine solution. The staining with bromophenol blue (BPB) as an indicator of free amino functions helps to visualize the amino-spots on the membrane.
*Side-chain protection group cleavage:* Final peptide deprotection is carried out with 50-90% of trifluoroacetic acid in dichloromethane with 3% triisobutylsilane and 2% water as scavengers.
The detailed protocol is given in [112, 113].

Synthesis areas (spots) are defined by spotting Fmoc-β-alanine-pentafluorophenyl ester solution (Fmoc-β-Ala-OPfp) to distinct sites on the membrane. This is done either by a pipetting robot (Abimed GmbH, Langenfeld, Germany) or manually. Blocking (acetylating) the remaining amino functions between the spots provides up to 8000 discrete reaction sites on a 20 x 30 cm membrane for further standard, solid phase peptide synthesis using amino acid pentafluorophenyl esters (Fmoc-aa-OPfp). Using the Fmoc-aa-OPfp standard protocol (Figure 3.1), the carboxyl groups are activated and the synthesized peptides are therefore C-terminally attached to the cellulose membrane. An activation of the amino-group for synthesizing N-terminally attached peptides would lead to a racemate formation of ~50%.

Routinely, peptides up to a length of 20 residues, but also longer peptides [137] can be synthesized with sufficient fidelity. The peptides can then be used for binding and enzymatic assays directly on the cellulose membranes or cleaved from the solid support by the use of ammonia or other reagents [111, 136].

The extended SPOT synthesis concept can be used to create libraries for various screening options such as substitutional analyses, combinatorial libraries, length analyses, integration of non-natural building blocks or other chemical entities. In recent years, peptide arrays prepared by the SPOT technique have become popular tools for studying protein-protein interactions [138].

### 3.1.2 Previous Method of Inverted Peptides

Some protein domains, such as PDZ domains (reviewed in [27, 139]) bind to peptides only if the ligands have a free carboxyl-terminus (C-terminus). Unfortunately, peptides synthesized according to the standard SPOT synthesis protocol [112, 113] (see Chapter 3.1.1) are attached by their C-termini to the cellulose support. In principle, free C-termini can be obtained after standard SPOT synthesis by reversing the peptide orientation.

Methods for synthesizing free C-terminal resin-bound [140-142] and cellulose-bound [104] peptides have been published. Yields and purities of the inverted cellulose-bound peptides are generally low and their generation is extremely time-consuming.

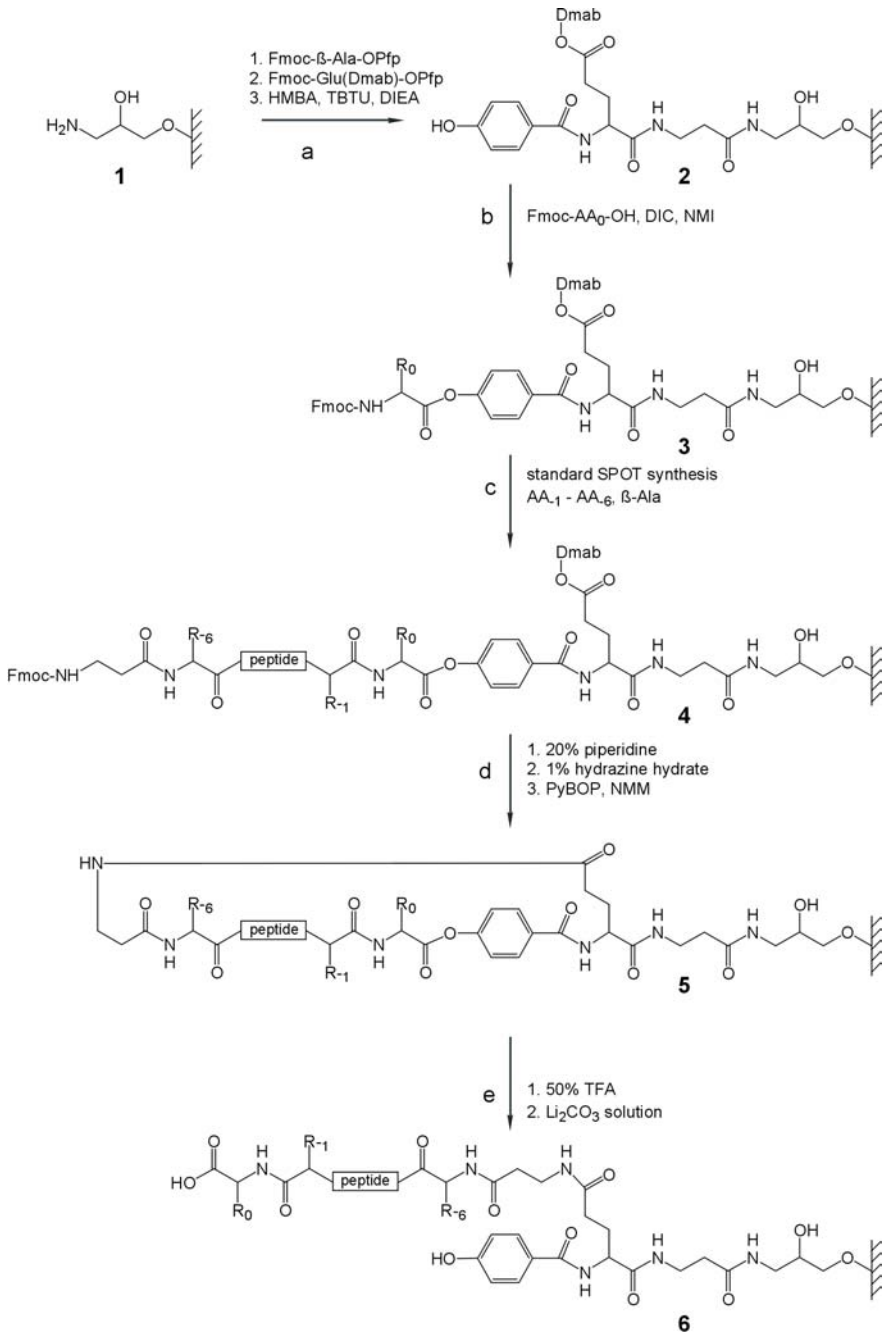


**Figure 3.2 Outline of the SPOT Synthesis Procedure Published in Hoffmüller *et al.* [104].**
Aminopropylether modified cellulose [136] **(1)** was used as support. (a) β-alanine was used as spacer and to define the spot areas, Dmab-glutamic acid as bivalent linker and HMBA as a base-labile cleavage site to yield in **(2)**. (b) Fmoc-aa-OH as intended C-terminal amino acid was introduced as ester **(3)**. (c) Standard SPOT synthesis and β-Ala coupling as spacer to yield the β-Ala-coupled peptide **(4)**. (d) The Fmoc-protection-group at the N-terminus and the Dmab-protection-group at the side chain of the glutamic acid were cleaved. Subsequently, the cyclisation was performed using PyBOP and NMI to achieve the cyclic peptide **(5)**. (e) After cleavage of the side chains, the ester bound was hydrolyzed with $Li_2CO_3$ resulting in a linear peptide with free C-terminus **(6)**.
Dmab = α-4-[N-{1-(4,4-dimethyl-2,6-dioxocyclohexyliden)- 3-methylbutyl}amino]benzyl, Fmoc = 9-fluorenyl-methyloxycarbonyl-, HMBA = 4-hydroxymethylbenzoic acid, TBTU = O-(benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium tetrafluoroborate, DIEA = N,N-diisopropylethylamine, DIC = diisopropylcarbodiimide, NMI = N-methylimidazole, PyBOP = [benzotriazol-1-yloxy]tripyrrolidinophosphonium hexafluorophosphate, NMM = N-methylmorpholine, TFA = trifluoro acetic acid.
Details of the SPOT synthesis procedure are given in Hoffmüller *et al. [104]*.

Coupling problems are reflected in low yields as well as a high background of peptides with incorrect sequences. Due to those problems, it was necessary to develop an improved protocol, which allows the rapid and robust synthesis of large peptide arrays with free C-termini. Short reaction times, together with high coupling efficiencies at each step during SPOT synthesis, fewer side reactions and the possibility of automating the SPOT synthesis process were the criteria that needed to be matched. Therefore, a robust and more efficient protocol for the preparation of cellulose membrane-bound inverted peptide arrays was developed in the context of this thesis. This protocol is much better suited for extensive experimental projects mapping many different PDZ domain interactions. It is based on the protocol published by Hoffmüller *et al.* [104] which was ameliorated to make it less time consuming and more adapted for fully automatic SPOT synthesis. Using the method of Hoffmüller *et al.*, the following compounds had to be coupled four times for satisfying yields: the Dmab-glutamic acid as bivalent linker, the hydroxymethylbenzoic acid (HMBA) as base-labile cleavage site and the Fmoc-aa-OH as intended C-terminal amino acid (Figure 3.2). Furthermore, these compounds had to be freshly activated at each reaction-step.

## 3.2 Results

### 3.2.1 A Modified SPOT Synthesis Strategy of Cellulose Membrane-Bound Inverted Peptides

The reaction scheme for the new SPOT synthesis of cellulose membrane-bound inverted peptides is shown in Figure 3.3. Compared to the standard SPOT synthesis protocol [111, 112, 143], synthesis of inverted peptides was performed on a cellulose membrane carrying a stable N-modified cellulose-amino-hydroxypropyl ether (N-CAPE) linker [111, 137] (bold numbers **1-9** refer to Figure 3.3). Key compounds in the synthesis are the Fmoc-amino acid 3-bromopropyl esters (Fmoc-aa-OPBr) **(3)**, the membrane-bound mercaptopropionyl cysteine adduct **(4)**, the matrix-bound amino acid ester derivative **(5)** and the cyclic peptide **(8)**. Critical reaction steps are the formation of both the cleavable ester bond and the cyclic peptide (Figure 3.3, d and g, respectively).

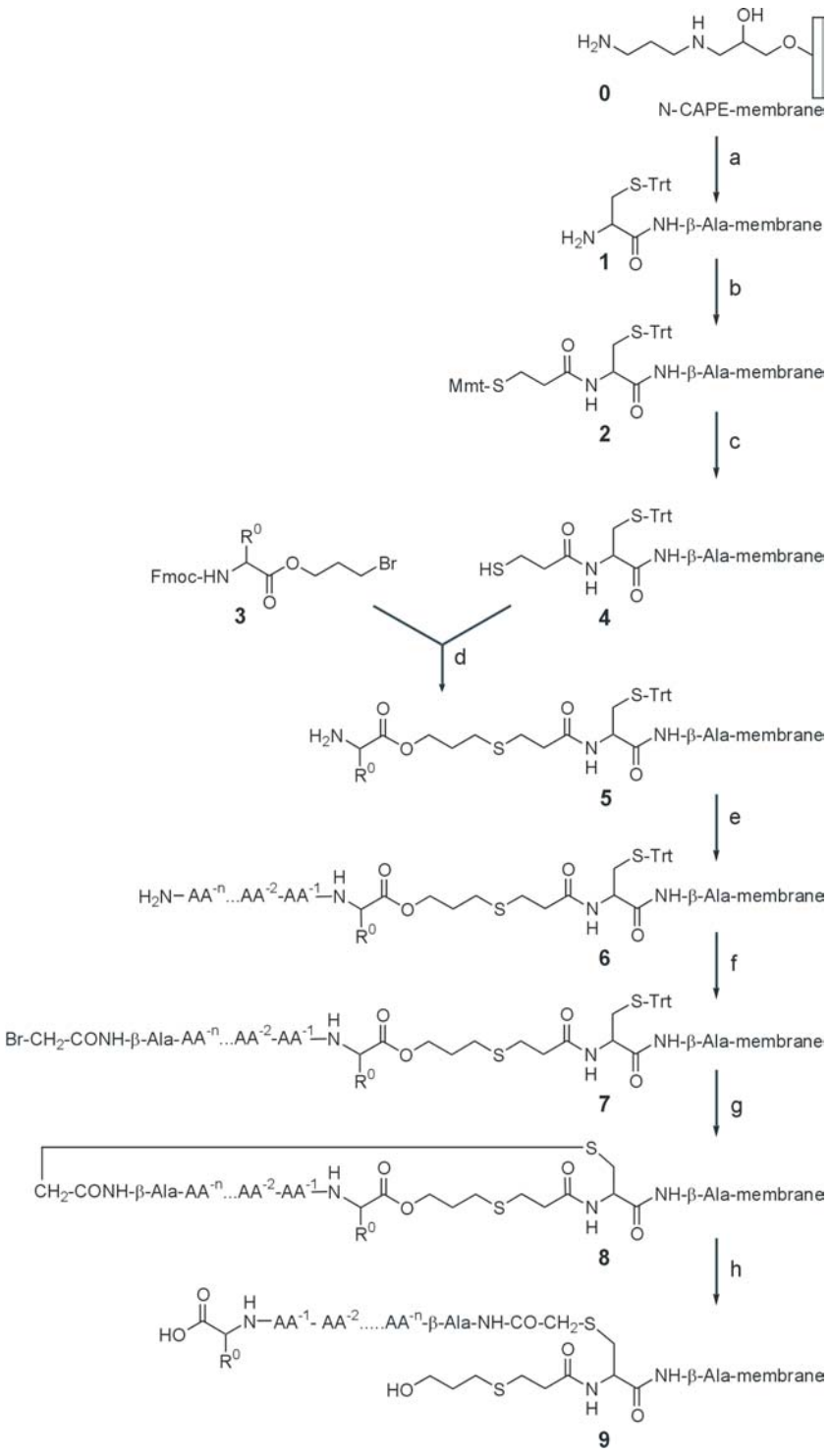


**Figure 3.3 Reaction Scheme for the Synthesis of Inverted Peptides on Cellulose Membranes.**
N-modified cellulose-amino-hydroxypropyl ether (N-CAPE) membrane [111, 137] **(0)** was used as support. (a) Coupling of Fmoc-β-Ala-OPfp in DMSO followed by Fmoc-cysteine-(Trt)-OPfp in NMP yield the bivalent linker **(1)**; (b) Mmt-S-$CH_2$-$CH_2$-COOH in DMF preactivated with HATU and NMI; were used to achieve the base-labile compound **(2)** (c) Mmt-cleavage with dichloroacetic acid, TFA, TIBS, DCM; (d) activation of the SH-group of **(4)** by aqueous solution of $Cs_2CO_3$ followed by the coupling of Fmoc-aa-OPBr **(3)** yielding the peptide **(5)**; (e) peptide synthesis using the standard SPOT synthesis protocol [112] to elongate the peptide **(6)**; (f) Fmoc-β-Ala-OPfp in NMP and 2,4-dinitrophenyl-bromoacetate in NMP were coupled to the compound **(7)**; (g) TFA, TIBS, DCM and subsequently aqueous solution of $Cs_2CO_3$ to prepare the cyclic peptide **(8)**; (h) Used of saturated aqueous solution of $Li_2CO_3$ to obtain a linear peptide with free C-terminus **(9)**.

The inverted and N-terminally fixed peptides **(9)** display a free C-terminus resulting from the reversal of the peptide orientation, and were achieved by successive thioether-cyclisation/ester-cleavage (Figure 3.3, g-h). The crucial step during the synthesis of the key compound **(4)** is the selective removal of the 4-methoxytriphenyl (Mmt) protection group (Figure 3.3, c). The acidic mixture used in the Mmt-deprotection step has to be adjusted carefully in order to prevent cleavage of the triphenylmethyl (Trt) protection group of the cysteine residue.

As shown in Table 3.1, selective removal of the Mmt-group is achieved by applying a mixture of 10% dichloroacetic acid, 0.5% trifluoroacetic acid (TFA), followed by three-times repeated incubation of 10% dichloroacetic acid, 0.5% trifluoroacetic acid (TFA) and 5% triisobutylsilane (TIBS) in dichloromethane (DCM). The key compound **(5)**, containing the C-terminal amino acid, was synthesized subsequently by caesium salt-supported S-alkylation of **(4)** and the coupling of Fmoc-aa-OPBr **(3)**.

**Table 3.1 Selective Cleavage of S-Mmt Group.**

| | Protecting Group | |
|---|---|---|
| Acidic Conditions* | Mmt | Trt |
| <0.5% TFA in DCM | – | – |
| 0.5 –1% TFA in DCM | + | – |
| ≥ 1% TFA in DCM | + + | + |
| < 8% Dichloracetic acid in DCM | – | – |
| 8 – 15% Dichloroacetic acid in DCM | + | – |
| ≥ 15% Dichloroacetic acid in DCM | + + | + |
| 50% Acetic acid | – | – |
| 10% Dichloroacetic acid, 0.5% TFA in DCM | + + | – |

**Footnotes:** * = Mixtures contain additional 5% TIBS; % = v/v; - = no cleavage; + = cleavage; ++ = immediate and fast cleavage

Compounds **(3)** (except the arginine derivative) were prepared in high yields by O-acylation of 1-bromo-3-propanol with Fmoc-amino acid fluorides [144] in a “one-pot” reaction. The reaction is characterized by its lack of side reactions. Racemic products could not be detected, which is in good accordance with the literature [145]. Only the corresponding Fmoc-amino acids were observed as impurities in the range of 10-25%. The Fmoc-arginine 1-bromo-3-propanol ester was synthesized according to standard

esterification protocols [146]. Yields of the prepared Fmoc-aa-OPBr were determined by analyzing the crude reaction products using reversed phase high performance liquid chromatography (RP-HPLC) and electrospray ionisation-mass spectrometry (ESI-MS) as shown in Table 3.2.

**Table 3.2 Yields and Characteristics of the Fmoc-Amino Acid 3-Bromopropyl Esters (3).**

| Fmoc-aa-R<br>R: O-$(CH_2)_2$-$CH_2Br$ | Yields | | | ESI Mass Spectrometry Data | |
|---|---|---|---|---|---|
| | Fmoc-aa-F | DCC/NMI | CDI | expected $[M+H]^+/[M+H]^{++}$ | measured $[M+H]^+/[M+H]^{++}$ |
| Fmoc-Ala-R | 82% | | | 432.0811/434.0793 | 432.0834/434.0791 |
| Fmoc-Arg(Pbf)-R | - | 76% | | 769.2271/771.2257 | 769.2255/771.2269 |
| Fmoc-Asn(Trt)-R | 91% | 43% | | 717.1964/719.1953 | 717.1986/719.2010 |
| Fmoc-Asp(OtBu)-R | 89% | | | 532.1335/534.1319 | 532.1354/534.1350 |
| Fmoc-Gln(Trt)-R | 88% | | | 731.2120/733.2109 | 731.2167/733.2137 |
| Fmoc-Glu(otBu)-R | 86% | | | 546.1491/548.1475 | 546.1503/548.1425 |
| Fmoc-Gly-R | 81% | | | 418.0654/420.0636 | 418.0654/420.0641 |
| Fmoc-His(Trt)-R | 84% | | | 740.2124/742.2113 | 740.2128/742.2101 |
| Fmoc-Ile-R | 82% | | | 474.1280/476.1263 | 474.1295/476.1285 |
| Fmoc-Leu-R | 86% | 77% | | 474.1280/476.1263 | 474.1281/476.1269 |
| Fmoc-Lys(Boc)-R | 89% | | | 589.1913/591.1898 | 589.1934/591.1940 |
| Fmoc-Met-R | 83% | | | 492.0844/494.0826 | 492.0819/494.0846 |
| Fmoc-Phe-R | 76% | 58% | | 508.1123/510.1107 | 508.1127/510.1104 |
| Fmoc-Pro-R | 82% | 75% | | 458.0967/460.0950 | 458.0928/460.0916 |
| Fmoc-Ser(tBu)-R | 81% | | | 504.1385/506.1369 | 504.1361/506.1364 |
| Fmoc-Thr(tBu)-R | 79% | | | 518.1542/520.1526 | 518.1523/520.1449 |
| Fmoc-Trp(Boc)-R | 82% | | | 647.1743/649.1743 | 647.1812/649.1785 |
| Fmoc-Tyr(tBu)-R | 76% | 52% | | 580.1699/582.1684 | 580.1716/582.1689 |
| Fmoc-Val-R | 78% | | 52% | 482.0162/484.0285 | 482.0221/484.2285 |

During the reaction between the Fmoc-aa-OPBr **(3)** and the membrane-bound mercaptopropionyl cysteine adduct **(4)** (Figure 3.3, d) the C-terminal amino acid of the peptide and the ester cleavage site, both necessary to reverse the peptide orientation, were incorporated simultaneously. Due to problems caused by the ubiquitous cellulose hydroxyl functions, we chose the S-alkylation reaction for selective formation of the ester linkage. Coupling efficiencies of the Fmoc-aa-OPBr **(3)** are shown in Figure 3.4. Low coupling efficiencies were found for the Fmoc-aa-OPBr derivatives of histidine, asparagine, glutamine (Trt-protected side-chains) and arginine (Pbf-protected side-chain), probably due to sterical hindrance by the bulky protection groups. However, increasing concentrations of the critical amino acid derivatives up to 0.8 M in dimethylformamide (DMF) and repeated coupling (3x) significantly increased the coupling efficiency (Figure 3.4). Furthermore, the thiolate formation of compound **(4)** is stable when activated by 10% caesium carbonate after drying the cellulose membrane. Furthermore, no significant

decrease in coupling efficiency was observed, even when the S-alkylation reaction started three hours later. This result is important, particularly since one coupling step with the SPOT robot-supported preparation of large peptide arrays may require up to three hours.

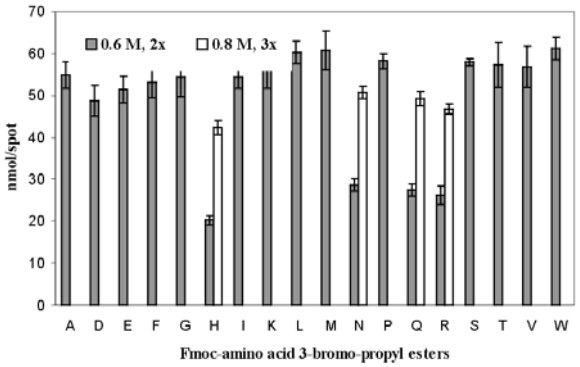


**Figure 3.4 Coupling Efficiency of Fmoc-Amino Acid 3-Bromo-Propyl esters (2).**
The 19 Fmoc-aa-OPBr were checked for coupling to membrane-bound caesium thiolate **(3)** (Figure 3.3, d). The coupling efficiency is over 50nmol/spot with a 0.6 M solution. Only for H, N, Q and R it is necessary to apply a 0.8 M solution with 3 coupling steps to obtain satisfying results.
Gray bars indicate a double coupling approach using 0.6 M solutions. White bars indicate the use of 0.8 M solutions and a three times coupling approach.

Before preparation of the key compound **(8)**, the completed peptide **(6)** was N-terminally elongated with Fmoc-β-Ala-OPfp and bromoacetic acid 2,4-dinitrophenyl ester, respectively. Subsequently, its side-chains were deprotected (Figure 3.3, g). The cyclization of the linear peptide **(7)**, which resulted in the formation of the thioether bond (Figure 3.3, g), was performed with a dilute aqueous solution of caesium carbonate. The formation of cyclic peptides using thioethers has been reported to be independent of peptide sequences and to deliver cyclic peptides in high yields and excellent purity [147]. Finally, the generation of peptides with reverse sequence-orientation **(9)** was achieved by cleaving the ester bond of the cyclic compound **(8)**, using a saturated aqueous solution of lithium carbonate (Figure 3.3, h) as described [104]. The formation of cellulose membrane-bound inverted peptides depends on the rate of peptide thioether cyclization. We tested the efficiency of thioether-cyclization with the peptide $NYKQTSV_{COOH}$ which binds to the PSD-95 PDZ3-domain [148]. The bromoacetylated peptide **(10)** (bold numbers **10-12** refer to Figure 3.5 (A)) was synthesized on several cleavable spots (cleavability is conferred by incorporating the Fmoc-Rink linker).

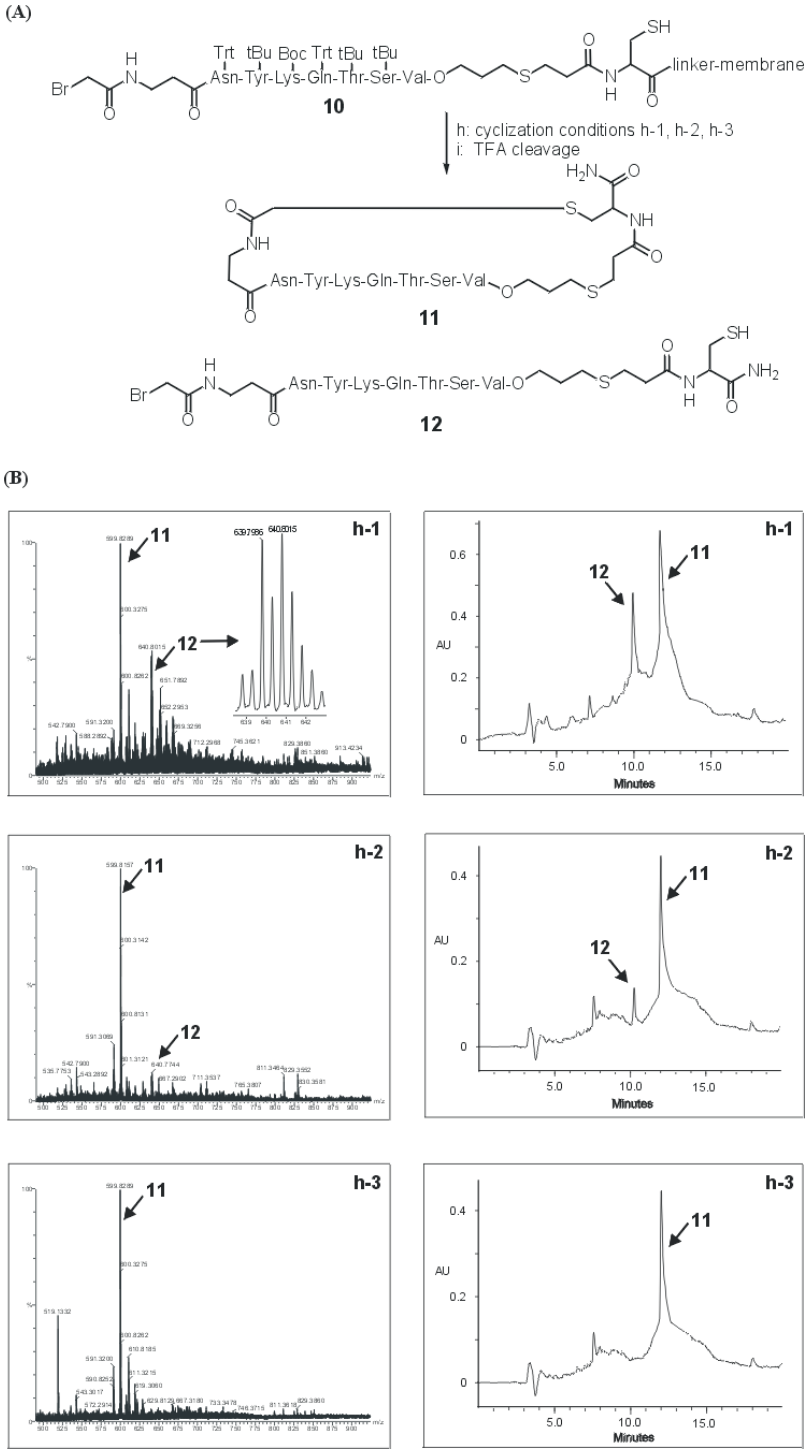


**Figure 3.5 Determination of the Cyclization Conditions.**
HPLC and MS analysis of the crude reaction product indicated that cyclization is quantitative using the condition h-3 (1 mM $Cs_2CO_4$ in DMF). Conditions h-1 and h-2 lead to incomplete cyclization of **(10).** Depending on the conditions used, various amounts of the linear peptide **(12)** were detected.
The efficiency of the thioether-cyclization reaction is given using a solution of caesium carbonate (1 mM) in DMF (condition h-3).
(**A**) The cyclic peptide **(11)** was generated according to Figure 3.3 with β-alanine replaced by the Fmoc-Rink linker (for details see Materials and Methods 2.2.3). (h) Conditions used for cyclization: (h-1) DMF, 3 h; (h-2) caesium thiolate, DMF, 3 h; (h-3) $Cs_2CO_4$ in DMF (1mM), 3 h; (i) TFA, water, TIBS, DCM, 3 h.
(**B**) ESI mass spectrometry (left) and RP-HPLC (right) of the crude products resulting from the conditions h-1, h-2 and h-3 used in cyclization of the membrane bound peptide **(10)**.

Three conditions were used for the cyclization: one array was incubated with DMF (Figure 3.5, h-1), the second was first treated with a dilute aqueous solution of caesium carbonate and then incubated with DMF (Figure 3.5, h-2), and the third array was incubated with a solution of caesium carbonate (1 mM) in DMF (Figure 3.5, h-3). After 3 hours the peptide products h-1, h-2 and h-3 were cleaved from the spots using TFA and analyzed by HPLC and ESI-MS (Figure 3.5 (B)). Analysis of the reaction products clearly demonstrated the efficiency of the thioether-cyclization reaction, using condition h-3. To determine the influence of the peptide sequence on the cyclization/cleavage step, two further membrane bound bromoacetylated peptides, $EFHAALGSYV_{COOH}$ and $QHIDSQKKA_{COOH}$, were synthesized. During the cyclization-steps, we used a dilute aqueous solution of caesium carbonate (5%, 12 h) and analyzed them subsequently by ESI-MS. In both cases, quantitative and reproducible thioether-cyclization can be observed, as shown in the spectrum of condition h-3 (1 mM caesium carbonate in DMF).

### 3.2.2 Evaluation of the Novel Strategy

To assess the novel method for synthesizing cellulose membrane-bound inverted peptides, a library of 6223 C-termini of human proteins (Swiss-Prot database, release 40) was screened. We synthesized the C-terminal peptides as 11-mers to be sensitive to a putative interaction involving the amino acids in ligand positions below -6. This library, called 6223-Humlib (Figure 3.6), was incubated with the SNA1 PDZ domain and the signal intensities were subsequently measured in Boehringer Light Units (BLU). The results were compared with previously published data from Hoffmüller *et al.* [104]. In this publication, a library comprising 3514 C-termini of human proteins (Swiss-Prot database, release 34) was synthesized as 7-mer peptides and incubated under similar conditions with the SNA1 PDZ domain (called hereafter 3514-Humlib).

The BLU intensities of each spot of the two libraries (3514-Humlib and 6223-Humlib) were determined to compare both binding experiments accurately. To deduce the peptide sequences that could be counted as PDZ domain binders, the signal intensity of the background was quantified using 20 spots randomly located on the cellulose membrane. Peptide sequences showing signal intensities higher than the background intensity plus the

double standard deviation were defined as PDZ domain binders (3514-Humlib: background = $3x10^3$ BLU; standard deviation = $1x10^3$ BLU; PDZ domain binder $\geq 5x10^3$ BLU; 6223-Humlib: background = $9x10^3$ BLU; standard deviation = $1x10^3$ BLU; PDZ domain binder $\geq 11x10^3$ BLU) (A.A. Weiser and R. Volkmer-Engert, personal communication). Taken together, we obtained signal intensities for peptide sequences that bind to the SNA1 PDZ domain with values ranging between $5x10^3$ - $28x10^3$ BLU for the 3514-Humlib and between $11x10^3$ - $519x10^3$ BLU for the 6223-Humlib. The first 30 spots of each library with the highest intensities showed values ranging between $12x10^3$ - $28x10^3$ BLU for the 3514-Humlib and between $237x10^3$ - $519x10^3$ BLU for the 6223-Humlib (Table 3.3).

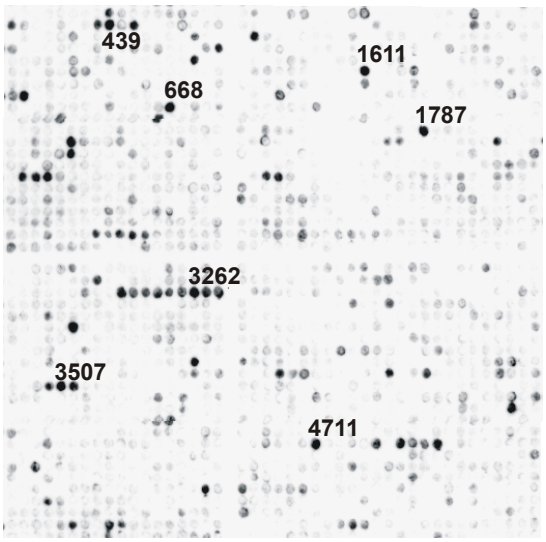


**Figure 3.6** ***In vitro*** **Identification of Putative SNA1 PDZ Domain Ligands.**
Selected region of the library of 6223 C-termini (11-mers) of human proteins from the Swiss-Prot database incubated with GST-SNA1 PDZ domain (cysteine is replaced by serine). The numbers (positioned with the first numeral above/under the corresponding spot) represent the peptide sequences showing the highest signal intensity measured in BLU (see Table 3.3 and Table 0.3 in the Appendix)

**Table 3.3 Evaluation of the SNA1 PDZ Domain Interactions.**

| Acc.No. | Protein description | Sequence | 6223-Humlib x10$^3$ BLU | 3514-Humlib x10$^3$ BLU |
|---|---|---|---|---|
| P41595 | 5-hydroxytryptamine 2B receptor | GDKTEEQVSYV | 83 | 13 |
| P28335 | 5-hydroxytryptamine 2C receptor | SSVVSERISSV | **319** | 10 |
| P25100 | Alpha-1D adrenergic receptor ** | ADYSNLRETDI | 84 | 28 |
| P08913 | Alpha-2A adrenergic receptor | ILCRGDRKRIV | 21 | 13 |
| P11766 | Alcohol dehydrogenase class III chi chain | SGKSIRTVVKI | 23 | 15 |
| P50052 | Type-2 angiotensin II receptor (AT2) | SSLREMETFVS | **361** | 6 |
| P18440 | Arylamine N-acetyltransferase 1 | VPKHGDRFFTI | 45 | 16 |
| Q01814 | Calcium-transporting ATPase plasma membrane, brain isoform 2 | GSPIHSLETSL | **283** | * |
| Q9UKP5 | Adam-TS 6 precursor | SSCNLAKETLL | **333** | * |
| P08588 | Beta-1 adrenergic receptor | CRPGFASESKV | **281** | 8 |
| O60241 | Brain-specific angiogenesis inhibitor 2 precursor | EPPDGDFQTEV | **248** | * |
| O60242 | Brain-specific angiogenesis inhibitor 3 precursor | DVQEGDFQTEV | **259** | * |
| Q16581 | C3A anaphylatoxin chemotactic receptor | NVISERNSTTV | **254** | * |
| P07766 | T-cell surface glycoprotein CD3 epsilon chain precursor | DLYSGLNQRRI | 30 | 16 |
| P22459 | Voltage-gated potassium channel protein KV1.4 | CSNAKAVETDV | 149 | 14 |
| P22460 | Voltage-gated potassium channel protein KV1.5 | LCLDTSRETDL | 106 | 19 |
| P35499 | Sodium channel protein, skeletal muscle alpha-subunit | TVRPGVKESLV | 197 | 16 |
| P53618 | Coatomer beta subunit | KINLSQKETSI | 218 | 14 |
| P81408 | Cote1 protein | SLNGGSRETGL | **238** | * |
| P33402 | Guanylate cyclase soluble, alpha-2 chain | IGTMFLRETSL | **359** | 28 |
| O00341 | Excitatory amino acid transporter 5 | TIQISELETNV | **270** | * |
| P23610 | Factor VIII intron 22 protein | LHLVLQETISP | **288** | 6 |
| P07148 | Fatty acid-binding protein, liver | DIVFKRISKRI | 14 | 24 |
| Q08379 | Golgin-95 | DENDEVKITVI | **238** | * |
| Q9UI32 | Glutaminase, liver isoform, mitochondrial precursor | ALSKENLESMV | **485** | * |
| P49683 | Probable G protein-coupled receptor GPR10 | HGQNMTVSVVI | **238** | 9 |
| P11169 | Glucose transporter type 3, brain | IEPAKETTTNV | **332** | 6 |
| Q9Y2T3 | Guanine deaminase | GKQVVPFSSSV | **260** | * |
| P01776 | IG heavy chain V-III region WAS | FDVFGQGTLVT | **227** | 12 |
| O43711 | Homeobox protein HOX-11L2 | SSKVPAVTSLV | **254** | * |
| P01589 | Interleukin-2 receptor alpha chain precursor | QRRQRKSRRTI | 16 | 26 |
| P15248 | Interleukin-9 precursor | KEKMRGMRGKI | 17 | 15 |
| P48049 | Inward rectifier potassium channel 2 | EPRPLRRESEI | 195 | 17 |
| P48050 | Inward rectifier potassium channel 4 | DNISYRRESAI | 197 | 13 |
| P78508 | ATP-sensitive inward rectifier potassium channel 10 | GSALSVRISNV | **306** | * |
| Q99712 | ATP-sensitive inward rectifier potassium channel 15 | LRTLLLQQSNV | **408** | * |
| P18084 | Integrin beta-5 precursor | KFNKSYNGTVD | 55 | 13 |
| P11801 | Putative serine/threonine-protein kinase H1 | VDPGARMTALQ | **439** | 5 |
| P01702 | IG lambda chain V-I region NIG-64 | FGGGTRVTVLG | 128 | 12 |
| P06887 | IG lambda chain V-I region MEM | FGTGTKVTVLR | **254** | 8 |
| P01705 | IG lambda chain V-II region NEI | FGGGTRVTVLS | **230** | 12 |
| P27816 | Microtubule-associated protein 4 | TLDSQIQETSI | 128 | 20 |
| P04201 | MAS proto-oncogene | NCNTVTVETVV | **235** | 12 |
| P33993 | DNA replication licensing factor MCM7 | VNASRTRITFV | 172 | 28 |
| O60669 | Monocarboxylate transporter 2 | QSVTSERETNI | **268** | * |
| Q13614 | Myotubularin-related protein 2 | AQCVTPVQTVV | **432** | * |
| P48645 | Neuromedin U-25 precursor | PRNGRRSAGFI | 34 | 16 |
| Q01959 | Sodium-dependent dopamine transporter | RQFTLRHWLKV | 22 | 14 |
| Q29459 | Platelet-activating factor acetylhydrolase IB beta subunit | ETPEEKQTTIA | **238** | * |
| P35080 | Profilin II | ELALYLRRSDV | 59 | 17 |
| P25800 | Rhombotin-1 | QLNGTFESQVQ | **261** | - |
| P47872 | Secretin receptor precursor | QSQGTCRTSII | 147 | 18 |
| P30872 | Somatostatin receptor type 1 | NGTCTSRITTL | **258** | 18 |
| P01135 | Transforming growth factor alpha precursor | RTACCHSETVV | 210 | 14 |
| P04437 | T-cell receptor alpha chain V region CTL-l17 precursor | FGTGTRLQVTL | **334** | 5 |
| P01733 | T-cell receptor beta chain V region YT35 precursor | TFGSGTRLTVV | **519** | 23 |
| Q99437 | Acuolar ATP synthase 21 kDA proteolipid subunit | ILQTSRVKMGD | **317** | * |

**Footnotes:** * = sequence not present in Hoffmueller *et a.l* [104]; ** = new sequence entry in Swiss-Prot;
**-** = non-binder; binder = background + 2x standard deviation
3514-Humlib: total: 5x10$^3$ BLU - 28x10$^3$ BLU; 30 strong: 12x10$^3$ BLU - 28x10$^3$ BLU;
6223-Humlib: total: 10x10$^3$ BLU - 519x10$^3$ BLU; 30 strong: 237x10$^3$ BLU - 519x10$^3$ BLU;

The new method produced a 20-fold increase in signal intensity but only a 3-fold increase in the background signal, despite the usage of longer peptides. Altogether, we obtained a total of 60 binding peptides, of which 17 sequences from the 6223-Humlib were not represented in the 3514-Humlib. Furthermore, 3 peptide sequences were identical in both libraries. Of the resulting 40 comparable sequences, all strong binders from the 3514-Humlib were also found as binders in the 6223-Humlib and *vice versa*. The only exception was the rhombotin-1 peptide, which was detected as a strong binder ($261x10^3$ BLU) in the 6223-Humlib, but showed no signal in the 3514-Humlib. This may be due to different peptide length and could be excluded by additional experiments, such as substitutional analysis or surface plasmon resonance measurements.

Comparing all sequences defined as SNA1 PDZ binder, a good correlation between both libraries was obtained. In accordance with previously published results, strong signals were detected for the peptide derived from the soluble α-2 chain of the guanylate cyclase (3514-Humlib: $28x10^3$ BLU; 6223-Humlib: $359x10^3$ BLU) [104], from muscle sodium channel protein type IV alpha subunit (3514-Humlib: $16x10^3$ BLU; 6223-Humlib: $197x10^3$ BLU) [34], from the stress-activated protein kinase-3 (3514-Humlib: sequence not represented; 6223-Humlib: $186x10^3$ BLU) [83], from glutaminase-L (3514-Humlib: sequence not represented; 6223-Humlib: $485x10^3$ BLU) [84] and a weaker signal of the aquaporin-4 peptide (3514-Humlib: $7x10^3$ BLU; 6223-Humlib: $72x10^3$ BLU) [85].

### 3.2.3 Analysis of novel ERBIN PDZ Domain/Ligands

The newly synthesized library of 6223 C-termini from human proteins was incubated with the ERBIN PDZ domain (Figure 3.7 (A)). The PDZ binders were determined using the same procedure as in Chapter 3.2.2 (background = $4x10^4$ BLU, standard deviation = $1x10^4$ BLU, PDZ domain binder ≥ $6x10^4$ BLU, signal range: $6x10^4$ - $676x10^4$ BLU). The sequences of the 100 strongest interacting peptides are given in Table 0.2 in the Appendix. This list contains the C-termini of many membrane proteins and receptors including the well-known ERBIN PDZ domain interaction partner of the armadillo repeat protein deleted in velo-cardio-facial syndrome (ARVC; spot no.: 407) [75]. Surprisingly, an interaction with the muscle sodium channel protein type IV alpha subunit (CIN4; spot no.: 1027) and with the soluble α-2 chain of the guanylate cyclase (CYG4; spot no.: 1351) were found,

which are both described as SNA1 PDZ domain ligands [34, 104]. We selected for further investigations by substitutional analysis the voltage-gated potassium channel protein Kv1.4 (CIK4, peptide $CSNAKAVETDV_{COOH}$), the voltage-gated potassium channel protein Kv1.5 (CIK5, peptide $LCLDTSRETDL_{COOH}$), the plasma membrane Ca(2+) ATPases, isoform 1 (ATB1, peptide $GSPIHSLETSL_{COOH}$) (Figure 3.7 (B-D)) and the breakpoint cluster region protein (BCR, peptide $KRQSILFSTEV_{COOH}$) (Figure 3.8 (A)). In those assays, each residue of the ligand is substituted by all natural L-amino acids (cysteine omitted), enabling the identification of the key residues of the ligand. The observed substitution patterns for the chosen peptides were generally very similar to each other and typical for the canonical PDZ domain/peptide interaction. In all cases, the ERBIN PDZ domain shows a clear preference for valine in ligand position 0. However, it also tolerates leucine or isoleucine. Position -1 shows fewer restrictions but some amino acids such as glycine, isoleucine and valine are not tolerated. For position -2, the ERBIN PDZ domain prefers threonine and serine, as typical for a class I PDZ domain. On the other hand, ligands such as the receptor protein tyrosine kinase ERB2 peptide ($NPEYLGLDVPV_{COOH}$) with a hydrophobic amino acid at position -2 (class II) can also interact with this PDZ domain [70]. In position -3, ERBIN preferred a negatively charged amino acid such as glutamate or aspartate, but also tolerates substitutions with small or hydrophobic amino acids. The substitutional analysis showed less selectivity for ligand positions beyond residue -3 (positions -4 to -9). Normally, residues N-terminal to ligand position -3 do not contribute substantially to affinity or specificity of PDZ domains, with some exceptions [31-33].

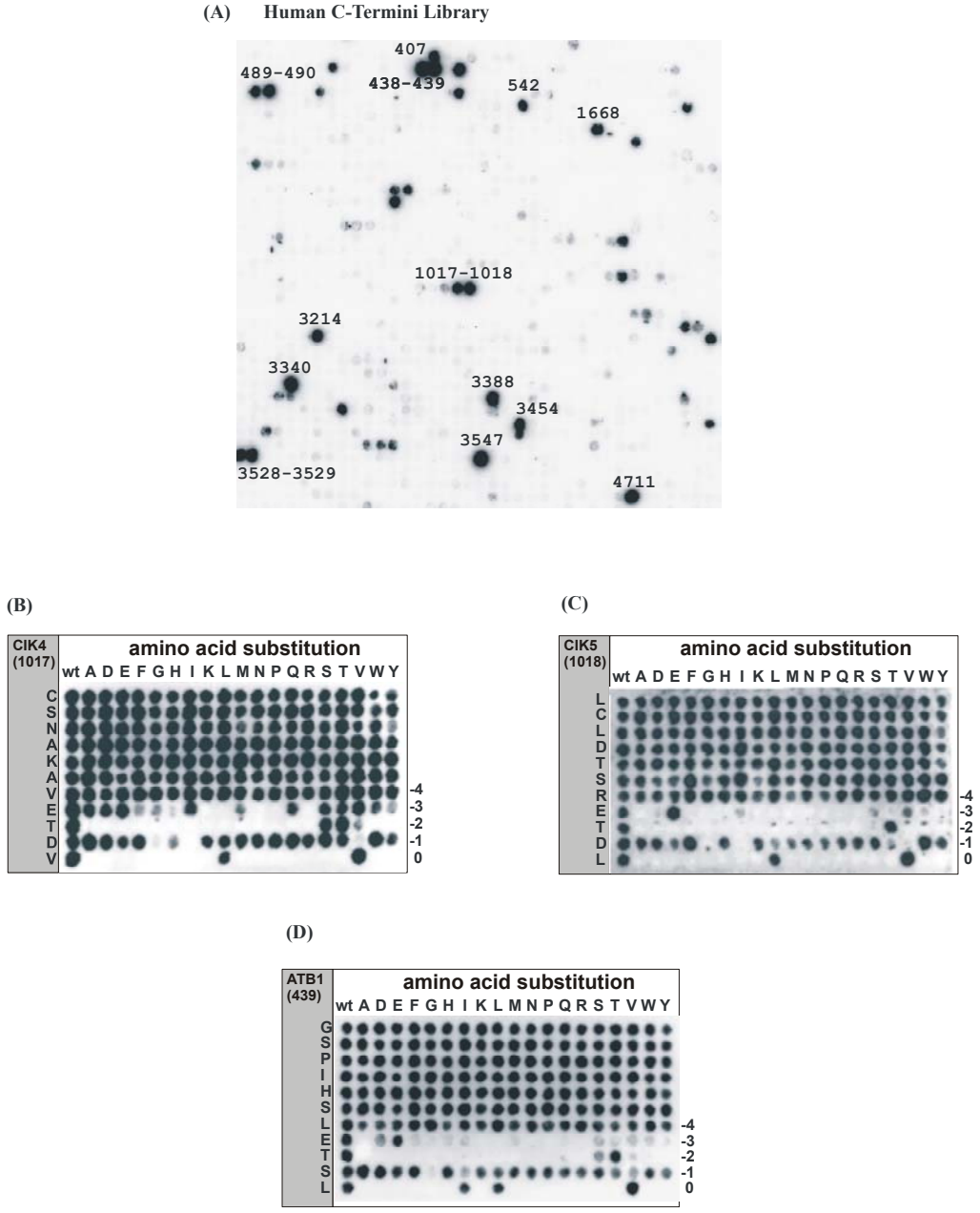


**Figure 3.7 Inverted Peptide Arrays on Cellulose Membranes for Protein Binding.**
(**A**) Selected region of the library of 6223 C-termini (11-mers) of human proteins from the Swiss-Prot database incubated with the GST-labeled ERBIN PDZ domain (cysteine is replaced by serine). The 100 strongest binders are listed in Table 0.2 in the Appendix.
Substitutional analyses of the peptide derivates **(B)** of the voltage-gated potassium channel protein Kv1.4 (CIK4), **(C)** of the voltage-gated potassium channel protein Kv1.5 (CIK5) and **(D)** of the calcium-transporting ATPase plasma membrane, isoform 1 (ATB1) incubated with GST-ERBIN-PDZ (spot numbers are given in parentheses). Each residue of the peptide ligand is substituted by 19 naturally occurring L-amino acids (cysteine omitted). All spots in the left column are identical and represent the wild type (wt) peptide (boxed in grey). All other spots represent the substitution of one amino acid against the amino acid in the respective column. Hence, each spot bears a single substitution compared to the sequence of the wild type. For example, V in position 0 can only be substituted by L and in some case by I.

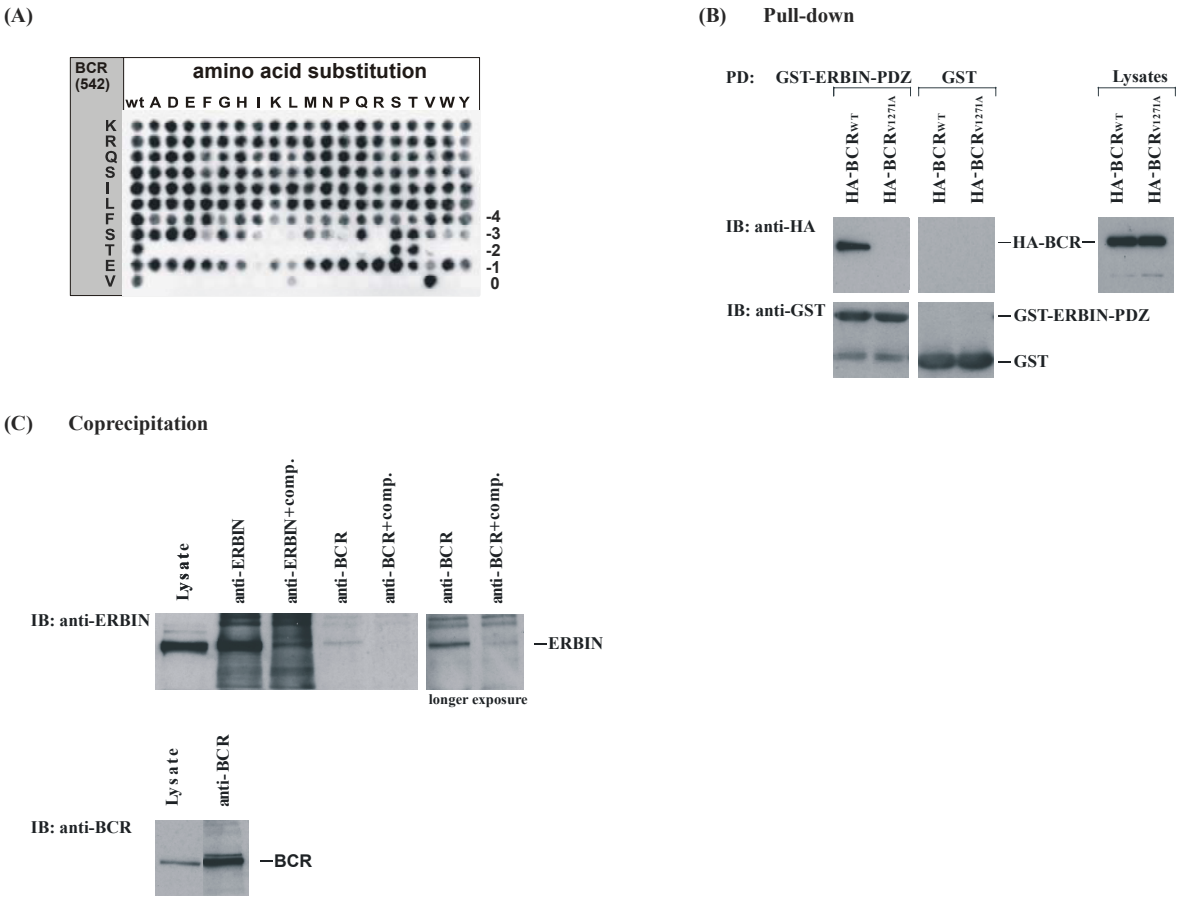


**Figure 3.8 Characterization of BCR as Ligand of the ERBIN PDZ Domain.**
The interaction between the C-terminus of BCR and the ERBIN PDZ domain deduced from the 6223-Humlib could be further confirmed through a substitutional analyses and a so-called pull-down experiment. Additionally, this interaction is also detectable between endogenous ERBIN and BCR in cells.
(**A**) Substitutional analysis of the C-terminal peptide derived from the breakpoint cluster region (BCR, spot number is given in parentheses).
(**B**) Pull-down (PD): HEK293 cells transfected with plasmids encoding HA-$BCR_{WT}$ and HA-$BCR_{V1271A}$ were lysed and incubated with GST-ERBIN-PDZ bound to GSH-agarose beads. Proteins bound to GST fusion proteins were fractioned by SDS-PAGE and immunoblotted (IB) with an anti-HA antibody and an anti-GST antibody, respectively. As a control for the expression of the BCR proteins, direct lysates of the 293 cells were immunoblotted with an anti-HA antibody.
(**C**) Coprecipitation of endogenous ERBIN and BCR in MKN-7 cells. The lysate of MKN-7 cells was incubated with an anti BCR antibody in the absence (top panel, lanes 4, 6) or presence of BCR peptide (+ comp.) (top panel, lanes 5, 7) followed by immunoblotting with anti-ERBIN. Expression of endogenous ERBIN and BCR was verified by immunoprecipitation with anti-ERBIN (top panel, lanes 2) and with anti-BCR (bottom panel, lane 2) and by analysis of the direct lysate with anti-ERBIN (top panel, lane 1) and with anti-BCR (bottom panel, lane 1).

### *3.2.3.1 Measurements of Dissociation Constants*

To give an overview of the affinity of the 40 strongest ERBIN PDZ domain interacting ligands from the 6223-Humlib, the dissociation constants ($K_d$) were determined by surface plasmon resonance (Biacore) measurements for selected peptides (spot no.: 407, 438, 542, 1018 and 1027). We obtained $K_d$ values between $8.0 \pm 3.1$ µM for ARVC and $82.7 \pm 1.3$ µM for CIN4, which are typical for PDZ domain interactions (Table 3.4).

**Table 3.4 Dissociation Constants of ERBIN PDZ domain/peptide complexes.**

| Spot | Sequence | Kd [µM] | Protein description |
|---|---|---|---|
| 407 | DAKPQPVDSWV | 8.0 ± 3.1 | armadillo repeat protein deleted in velo-cardio-facial syndrome* |
| 438 | GSPLHSLETSL | 24.5 ± 5.0 | plasma membrane calcium-transporting ATPase, isoform 1 |
| 542 | KRQSILFSTEV | 36.0 ± 5.2 | breakpoint cluster region protein |
| 1018 | LCLDTSRETDL | 79.4 ± 2.8 | voltage-gated potassium channel protein KV1.5 |
| 1027 | TVRPGVKESLV | 82.7 ± 1.3 | sodium channel protein, skeletal muscle alpha-subunit |

**Footnotes:** Peptide sequence with * is published in [75]. Cysteine is replaced by serine.

### *3.2.3.2 Verification of the BCR/ERBIN Interaction in Mammalian Cells*

To confirm that the BCR peptide interaction observed *in vitro* may take place between BCR and ERBIN in mammalian cells, we performed a pull-down assay by using the PDZ domain of ERBIN fused to GST. The GST-ERBIN PDZ and the GST protein alone as a control were incubated with lysates of HEK293 cells overexpressing $BCR_{WT}$ and a C-terminal mutant $BCR_{V1271A}$, both as HA (influenza hemaglutinin-epitope)-tagged fusions. As predicted, $BCR_{WT}$ but not the mutant $BCR_{V1271A}$ bound to the ERBIN PDZ domain (Figure 3.8 (B), lanes 1, 2), whereas the GST protein alone did not bind (Figure 3.8 (B), lanes 3, 4). Direct lysates of the 293 cells were immunoblotted with an anti-HA antibody as a control for the expression of the $BCR_{WT}$ and $BCR_{V1217A}$ proteins. This study was extended to substantiate the biological relevance of the interaction between full-length ERBIN and BCR as endogenous proteins. BCR and ERBIN were precipitated with the help of specific antibodies from MKN-7 cells where both proteins are well expressed (Figure 3.8 (C)). Indeed, ERBIN coprecipitated with BCR (Figure 3.8 (C), lanes 4, 6). Competition of the BCR and the ERBIN antibody with their respective antigen reduced immunoprecipitation, supporting the specificity of the interaction (Figure 3.8 (C), lanes 3,

5, 7). As a control for the expression of the BCR protein, the direct lysates of the 293 cells were also immunoblotted with an anti-HA antibody.

## 3.3 Discussion

The SPOT synthesis concept has been widely used to prepare peptide arrays for proteomics studies. The overall goal of this work was to extend the application range of the SPOT synthesis concept by creating N-terminally fixed peptide arrays (inverted peptides), enabling the display of free C-termini on planar cellulose supports. The new steps introduced into the SPOT synthesis concept are the synthesis of Fmoc-aa-OPBr, the selective cleavage of the Mmt-protection-group in presence of the Trt-protection-group, the formation of the cleavable ester bond and thioether-cyclisation/ester-cleavage. The developed method is now robust and suitable for proteomics studies. It could be applied to create non-redundant statistical libraries or libraries based on database sequences. The improved methodology, in particular with regard to automatic, robot assisted SPOT synthesis, enables the screening of a large range of divergent peptide sequences.

The coupling problems of the previous method of Hoffmüller *et al.* [104], where all compounds had to be activated stepwise, required up to four coupling-steps. With the new method, we obtained a better signal-to-noise graduation, which is important for detecting PDZ domain ligand interactions with low or intermediate affinities. The new method allowed thus a better discrimination of binding strength in the micromolar range for the investigated PDZ domains.

An obvious advantage of the improved method is due to the fact that only successfully cyclized peptides will be reversed in their sequence-orientation. The uncyclized peptides are released during the ester-cleavage procedure. In addition, side-reactions between neighboring peptide sequences that perhaps occur during the cyclization-step, e.g. cross-linking via a disulfide bridge or thioether cross-linking, do not disturb the reversal of peptide-orientation. Such disulfide cross-linked peptides were cleaved from the membrane during hydrolysis with saturated lithium carbonate solution. On the other hand, the thioether cross-linked products resulted also in inverted peptides. Taken together, all these steps are in principle additional purification steps, as evident from the RP-HPLC and MS

spectra. The spectra showed inverted peptides with high purity, clearly revealing the superiority of the improved thioether-cyclisation/ester-cleavage method.

The screen of 6223 C-termini of human proteins revealed four new peptide ligands for the ERBIN PDZ domain: the voltage-gated potassium channel proteins Kv1.4 and Kv1.5 and the plasma membrane Ca(2+) ATPases, isoform 1, which could be potential ERBIN PDZ domain ligands [73]. Further *in vivo* experiments will be needed to determine the biological relevance of these interactions.

ERBIN is the founding member of the LRR and PDZ domain (LAP) protein family, which is characterized by 16 LRRs at the N-terminus and one to four PDZ domains at the C-terminus [76]. Besides the PDZ-specific interactions with ERB2, δ-catenin, ARVC or p0071 [70, 75, 149, 150], ERBIN is known to interact with the small GTPase Ras and Rho through the LRRs [73, 151]. The novel interaction partner of ERBIN, BCR is a multidomain protein, containing a serine/threonine protein kinase domain, a Guanine Nucleotide Exchange Factor (GEF) function and a GTPase-activating protein (GAP) domain. The GEF and GAP domains can modulate the activity of Rho-type GTPases [152, 153]. Based on these results, it could be suggested that ERBIN is a scaffolding protein that links BCR and Rho-type GTPases through its PDZ domain and its LRRs, respectively, into a single complex.

Taken together, these data suggest that the improved method for generating cellulose membrane-bound inverted peptides may be a powerful tool to generate peptides with free C-termini in a large-scale manner for proteomics studies. It is a feasible method to find new ligands for PDZ domains and, furthermore, this technique allows the parallel screening of different PDZ domains to determine their specificity and selectivity towards a wide ligand sequence space.