Journal of Peptide Science, Vol. 3, 123–132 (1997)

# Synthesis and Screening of an Indexed Motif-Library Containing Non-proteinogenic Amino Acids

**SØREN ØSTERGAARD and ARNE HOLM**

Research Center for Medical Biotechnology, Chemistry Department Royal Veterinary and Agricultural University, Frederiksberg, Denmark

*Received 4 February 1996*
*Accepted 14 June 1996*

Abstract: In an effort to increase the probability of finding novel peptides in resin-bound combinatorial libraries displaying affinity to various macromolecular targets, we increased the diversity of a solid-phase library considerably by synthesizing multiple structures on each bead – a motif-library – including 45 building blocks. The building blocks consist of L-aa, D-aa and eight hydrophobic non-proteinogenic α-amino acids. A library with the format O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin was synthesized giving the four motifs OOOXX, OZOOXX, OOZOXX, OZOZOXX corresponding to 364.500 different motifs ($45^3 \times 4$ theoretical combinations). The positions O are defined amino acids while Z represents three mixtures Π, Ω, $\varphi$, where Π is a mixture of polar and charged residues, Ω is a mixture of aliphatic residues and $\varphi$ is a mixture of aromatic residues. X represents a mixture of all 45 residues. The library was screened with the macromolecular target streptavidin which served as a model receptor. Binding peptides were sequenced by microsequencing. We included small amounts of norvaline and norleucine in the library, which served as index residues to be able to distinguish between LD-amino acids and other residues with the same retention time in the HPLC system. Beads that interact with the receptor were found, and the binding motifs that appeared had no homology to known binding motifs found in either L-aa or D-aa libraries, instead motifs with the non-proteinogenic residues L-phenylglycine, *O*-benzyl-L-hydroxyproline and *O*-benzyl-L-tyrosine dominated. The novel peptides inhibit binding of biotin to streptavidin but do not bind to avidin, and the affinity is higher than the peptides found in linear all L-aa peptide libraries. © 1997 European Peptide Society and John Wiley & Sons, Ltd

Keywords: Indexed combinatorial peptide library; motif-library; nonproteinogenic amino acids; streptavidin

## INTRODUCTION

Over the past few years it has become increasingly feasible to screen very large combinatorial libraries of random structures and to identify rare members having binding activity. The synthetic libraries can be divided into two broad categories. One category of libraries is based on support-bound compounds, such as the pin method [1], resin-beads [2], cotton [3] or cellulose [4], while the other category involves the generation of soluble libraries [5]. Ligands found with all these methods may be useful agents either as lead components or used directly as drug candidates; for reviews see [6,7].

Peptides and oligonucleotides were initially used in the synthetic combinatorial experiments because they are easily synthesized on the solid support and the structure is readily deduced by microsequencing or PCR. More recently, the generation of synthetic non-peptide or peptidomimetic libraries has proven effective in the search for new reagents [8–21]. We have constructed a support-bound non-proteinogenic peptide library based on the one-bead one-motif approach [22] and used it to search for novel ligands for various receptors. The libraries were essentially constructed by the portioning-mix procedure [2, 5, 23] and incorporated both the L- and D-amino acids in addition to eight non-proteinogenic lipophilic α-amino acids. We have previously found that screening synthetic peptide libraries with

Address for correspondence: Søren Østergaard, Department of Chemistry, Royal Vet. & Agr. University, Thorvaldsensvej 40, Frederiksberg, Denmark.


CCC 1075-2617/97/020123-10 $17.50

various target receptors often resulted in a number of motifs where the most critical residues that appear in the motifs are either aliphatic or aromatic and frequently surrounded or adjacent to charged residues (unpublished results). This observation motivated us to include a larger repertoire of lipophilic residues in the peptide libraries. With the Fmoc/tBu strategy this can very easily be accomplished by selecting Fmoc-amino acids with side-chain functionalities compatible with Fmoc-chemistry, e.g. benzyl-protected residues that remain after TFA treatment.

Since the L- and D-amino acids are inseparable in the HPLC system of the sequencer we used an index to be able to determine whether the residues were in the L or D configuration. Ten per cent norvaline and 10% norleucine were added to the solutions of D-aa and L-aa respectively and thereby the appearance of a small peak of either norvaline or norleucine in addition to the specific retention times identified the residue. This approach was also used for those of the eight non-proteinogenic amino acids which have coinciding retention times.

To demonstrate the utility of this method, we synthesized a motif library with the format O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin, where O indicate 45 defined positions synthesized by the split-procedure and Z indicates three different groups of mixtures: a group Π comprising only polar or charged residues, a group Ω comprising aliphatic residues and a group $\varphi$ comprising aromatic amino acids. X is a mixture of all 45 amino acids and thus the length of the library varies between five and seven residues. We have characterized the library by the use of the model receptor streptavidin since this target has extensively been used in the screening of peptide libraries [15,24–33]. Thus novel ligands can be compared to already known binders. In this paper we describe the synthesis and screening of a non-proteinogenic peptide library to identify novel ligands to the protein target streptavidin and report the identification of peptides that have higher affinity than peptide ligands found in linear L-aa peptide libraries.

## MATERIAL AND METHODS

### Synthesis of O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin

One gram of TentaGel resin (Rapp Polymere, Tübingen) was coupled with a mixture X of 45 α-amino acids (1.1 eq., 2 h) using 1.1 eq. DIC and 1.1 eq HOBt followed by a further coupling (overnight) and after Fmoc deprotection this was repeated to give XX-resin. The mixture X comprised: L,D-Ala; L,D-Asn(trt); L,D-Gln(trt); L,D-Arg(pmc); L,D-Lys(Boc); L,D-His(trt); L,D-Asp(tBu); L,D-Leu; L,D-Val; LIle; L,D-Met; L,D-Pro; L,D-Ser(tBu); L,D-Trp(Boc); L,D-Phe; L,D-Tyr(tBu); Gly; L-cyclohexylalanine (Cha); *O*-benzyl-L L-phenylglycine [Phg]; L-Lys (Ac); L-Lys (Z) (all purchased from Novabiochem); D-Ile; $N_{im}$-benzyl-L-histidine (HBzl) *O*-benzyl-L-hydroxyproline (HypBzl); β-2-naphthyl-L-alanine (Nal2) (all purchased from Bachem). The XX-resin was then equally distributed into 32 reaction vessels and either single amino acids or pairs of amino acids were added using 4 eq. of amino acids with 4 eq. DIC and 4 eq. HOBt and coupling overnight. The following pairs of amino acids were used: (L/M), (l/m), (I/V), (i/v), (A/G), (a,G), (D/E), (d/e), (Q/N), (q/n), (R/K), (r/k), (S/T), (s/t); all in 1 : 1 ratio except for S/T and s/t which were 2 : 3. Single amino acids were: Y, y, F, f, W, w, H, h, P, p, Cha, Phg, Lys(z), Lys(Ac), YBzl, Nal2, HypBzl, HBzl. To each of the individual or pairs of naturally L-amino acids and YBzl, Fmoc-norleucine-OH was added to a final content of 10%, and to the D-amino acids and Cha, 10% Fmoc-norvaline-OH was added. To Phg, H(Bzl) and Lys(Ac) both norleucine and norvaline was added to give a content of 5% of each. The amino acids Lys(Z), Nal2 and HypBzl were coupled without index. After recombining the 32 portions coupling of Z (mixtures Π, Ω, $\varphi$) was accomplished by dividing the OXX-resin into four fractions. Fraction 1 was left uncoupled, while to the other three fractions the mixtures Π = [D, d, E, e, N, n, Q, q, K, k, R, r, S, s, T, t]; Ω = [V, v, I, i, L, l, M, m, P, p, A, a, G, Cha, Lys(Ac)] and $\varphi$ = [W, w, F, f, Y, y, H, h, Lys(Z), Phg, H(Bzl), Nal2, Hyp(Bzl), Tyr(Bzl)] were added respectively (1.1 eq.) followed by a double coupling overnight. The resins were combined and repeated split synthesis was performed giving a final library O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin. Deprotection of side-chain protecting groups was done by using 20 ml TFA containing 5% $H_2O$, 5% phenol, 5% thioanisole and 2.5% ethandithiol for 2.5% h. The resin was then washed with ethanol and subsequently washed with 5% AcOH and lyophilized.

### Synthesis of Dual Sub-library [L/l/M/m/Q/q/N/n]-HypBzl-O-Phg-O-O-resin + F-p-YBzl-O-O-resin

TentaGel resin weighing 0.6 g (approximately $1.5 \times 10^6$ beads) was divided into 45 reactions vessels (Position 0) and coupled with single Fmoc-amino acids. The 45 amino acids used in the synthesis were the same as described above except that L-Lys(Z) was replaced with D-Phg. Again 10%

norleucine and 10% norvaline were used as indexes. The split-synthesis was repeated and the O-O-resin was divided into two portions. The sub-library F-p-YBzl-O-O-resin was synthesized on *ca.* 40,000 beads while the sub-library [L/l/M/m/Q/q/N/n]-HypBzl-O-Phg-O-O-resin was synthesized on $1.5 \times 10^6$ beads. The two sub-libraries were then mixed and side-chain deprotection was performed as described above.

### Screening of Peptide Libraries

The primary library O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin (approximately $1.5 \times 10^6$ beads) was washed with TRIS washing buffer (25 mM Tris-HCl; pH = 7.4; 0.2 M NaCl; 0.05% Tween 20) followed by incubation in TRIS blocking buffer (25 mM Tris; pH = 7.4; 0.2 M NaCl, 0.1% gelatine; 0.05% Tween 20) for 1 h. Streptavidin–alkaline phosphatase (Sigma) was added (diluted 1 : 20,000) and incubated for 1 h, followed by washing with TRIS washing buffer and TRIS staining buffer (50 mM TRIS; pH = 9.3; 0.1 M NaCl; 10 mM $MgCl_2$). The substrates BCIP and NBT were added as described by the procedure by Lam and Lebl [25] and after 30 min the substrate was removed by washing with TRIS washing buffer. Approximately 300–400 beads were stained and the most intensively stained beads were removed for microsequencing. Screening of the dual sub-library was performed as described above except that streptavidin–alkaline phosphatase was diluted 1 : 100,000.

### Sequencing

The positive beads were removed with a capillary tube and washed extensively with $H_2O$. The beads were then submitted to sequencing in microsequencer (Procise, Applied Biosystems, Perkin-Elmer).

### Synthesis of Peptides for Binding Analysis

Synthesis of binding motifs from the screenings (see text) for binding analysis on beads was done as described before, except that X and Z mixtures were substituted by Fmoc-L-Ala-OH, giving the peptides beads: L/M-HypBzl-A-Phg-A-A-resin, F-*p*-YBzl-A-A-resin. Free peptides F-p-YBzl-S-S-$NH_2$, M-HypBzl-S-Phg-S-S-$NH_2$, Phg-M-n-Q-R-HypBzl-$NH_2$, M-HypBzl-n-Phg-R-Q-$NH_2$ were synthesized on TentaGel resin with Rink amide linker using HOBt/DIC, while peptide FSHPQNT was synthesized on KA-Resin (Novabiochem). The peptides were cleaved and side-chain deprotected as described in the synthesis of libraries. Peptides were precipitated in diethylether, washed five times with diethylether and dissolved in 10% acetic acid and lyophilized. Waters 600E equipped with Waters 996 photodiode array detector controlled by Millenium$^{TM}$ software was used for analytical HPLC on a C18 column (Delta-Pak 100 Å, 15 μ). If necessary preparative purification on C18 column were performed. Identification of peptides were confirmed by MS-MALDI Time-Of-Flight obtained on a Fisons Instrument, VG Tofspec E.

### Competitive ELISA

For the ELISA experiments peptides L-HypBzl-S-Phg-S-S-$NH_2$, M-HypBzl-S-Phg-S-S-$NH_2$, F-p-YBzl-S-S-$NH_2$, M-Hyp-Bzl-n-Phg-R-Q-$NH_2$ and Phg-M-n-Q-R-HypBzl-$NH_2$ were synthesized and dissolved in 40% DMSO, while FSHPQNT-$NH_2$ was dissolved in PBS. BSA coupled with biotinamidocaproate *N*-hydroxysuccinimide ester (1 eq.) was coated to maxisorp (Nunc, Denmark) overnight in $NaHCO_3$, pH = 9.5 at 4 °C. Serial dilutions of peptides and biotin were incubated with streptavidin–peroxidase (Sigma) diluted 1 : 8000 for 60 min and transferred to BSA–biotin coated maxisorp for 30 min. The substrate 3,3′,5,5′-tetramethylbenzidine (TMB tablets, Sigma) in 0.1 M citrate-buffer, pH = 5.0, and $H_2O_2$ were added and incubated for 20 min and absorbance was read at 450 nm.

## RESULTS AND DISCUSSION

A motif-library of peptides with the format O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin was synthesized using indexed combinatorial synthesis. O designates a defined position while Z is three different mixtures. Π comprising polar and charged residues. Ω comprising a mixture of aliphatic residues and φ comprising a mixture of aromatic residues, and the two C-terminal X is a mixture of all 45 amino acids. The use of cluster libraries have been reported by Blake and Litzi-Davis [34]. The 45 monomers comprise 18 L- and 18 D-amino acids (L-, D-cysteine excluded), glycine and the eight non-proteinogenic α-amino acids β-cyclohexyl-L-alanine (Cha), L-phenylglycine (Phg), L-Lys(Z), L-Lys(Ac), *O*-Benzyl-L-hydroxyproline (HypBzl), $N_{im}$-benzyl-L-histidine (HBzl), β-(2-naphthyl)-L-alanine (Nal2), O-benzyl-L-tyrosine (YBzl) (see Table 1). The defined positions O were synthesized by the portioning, mix procedure with

the slight modification that for some structurally related residues these were added as pairs, e.g. instead of adding isoleucine and valine into two separate steps they were added as a pair in one step (Table 1). Thereby the number of dividing steps is decreased but the diversity is maintained. To obtain near equal incorporation of the amino acids in the mixtures X or Z, only 1.1 eq. of each mixture relative to resin substitution was coupled but followed by a further coupling (1.1 eq.) overnight.

Based on previous results with peptide libraries it was found that hydrophobic residues very often appear as critical residues and charged amino acids are frequently observed. Criteria such as hydrophobicity and compatibility with Fmoc-chemistry were considered when we chose the eight non-proteinogenic building blocks, but a wider range of other α-amino acids are commercially available, making other or even larger peptide libraries possible.

### Structure Determination

The identification of each α-amino acid in a peptide by microsequencing is based on the retention time in the HPLC system and consequently L- and D-amino acids cannot be directly identified, neither can other unnatural α-amino acids if their retention time coincides with others. For some residues this problem can in principle be solved by changing HPLC conditions; however, changing gradients has no effect on the separation of D- and L-amino acids and for some other residues only small effects are observed. A solution to this problem is to tag residues by adding an index residue that will be co-sequenced. A small subset of indexes allows both the L-aa and D-aa and also non-proteinogenic α-amino acids to be identified by both the retention time and the index residue, even if residues happen to have coinciding retention times. This approach requires that the index residues are α-amino acids which can be used without side reactions in peptide synthesis and have specific retention times different from the residues that comprise the building blocks in the library. Norvaline and norleucine fulfil these requirements and we decided to index the genetically encoded L-amino acids with 10% norleucine while 10% norvaline was used for all the D-amino acids (Figure 1). Some of the eight non-proteinogenic amino acids had retention times coinciding other L- and D-amino acids: HBzl ~ L,D-lysine; L-Phg ~ L,D-phe; Lys(Ac) ~ L,D-alanine, and for these residues a 1:1 mixture of 5% norleucine and 5% norvaline was used. Lys(Z), Nal2 and HypBzl could be identified solely by their retention time and an index was not necessary, while Cha and YBzl were indexed with 10% norvaline and 10% norleucine respectively (Table 1).

The motif library described O-$Z_{0-1}$-O-$Z_{0-1}$-O-XX-resin above contained the following four motifs: OOOXX-resin, OOZOXX-resin, OZOOXX-resin, OZOZOXX-resin. Septov *et al.* have described the synthesis of a more extensive motif library with a fixed length; however, this requires more split synthesis and is more laborious, although it ensures that all motifs with a given length and with three defined positions (O) are synthesized [22]. We decided to reduce the number of motifs in order to include all motifs statistically. During split synthesis

Table 1 α-Amino Acids and the Corresponding Indexes Used in the Synthesis of the Non-proteinogenic Peptide Motif-Library

| 10% norleucine | 10% norvaline | 5% norvaline + 5% norleucine | No index |
|---|---|---|---|
| L-Phe | D-Phe | L-Phg | L-Lys(Z) |
| L-Trp | D-Trp | L-HBzl | L-Nal2 |
| L-Tyr | D-Tyr | L-Lys(Ac) | L-HypBzl |
| L-His | D-His | | |
| L-Pro | D-Pro | | |
| L-(Arg/Lys) | D-(Arg/Lys) | | |
| L-(Ala/Gly) | D-(Ala/Gly) | | |
| L-(Asp/Glu) | D-(Asp/Glu) | | |
| L-(Ile/Val) | D-(Ile/Val) | | |
| L-(Leu/Met) | D-(Leu/Met) | | |
| L-(Ser/Thr) | D-(Ser/Thr) | | |
| L-(Asn/Gln) | D-(Asn/Gln) | | |
| L-Tyr(Bzl) | L-Cha | | |

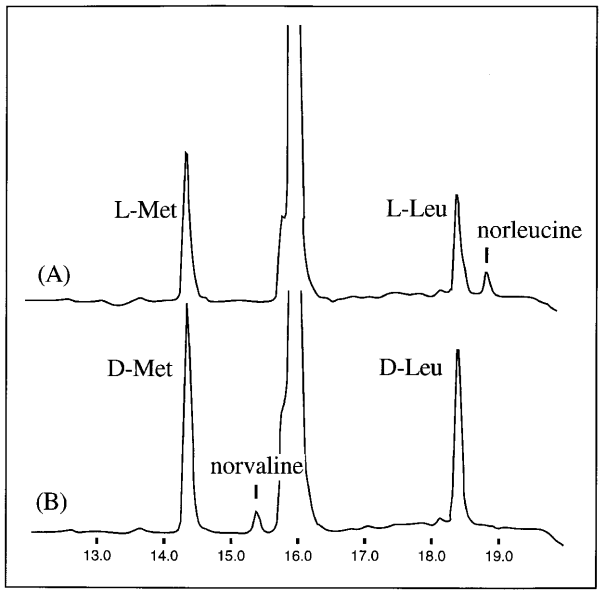


Figure 1 HPLC chromatogram from the sequencing of two beads having L-Met, L-Leu indexed with norleucine (A) and D-Met, D-Leu indexed with norvaline (B).

the resin is divided into 32 fractions (45 monomers) to synthesize position O, theoretically giving 131,072 ($32^3 \times 4$) different beads and by including $1.5 \times 10^6$ beads we ensure that more than $>99\%$ of all motifs are synthesized [35]. Since some residues O are added as a pair (see Materials and methods), some beads contain more than one motif and the theoretical number of different motifs in the library is therefore 364,500 ($45^3 \times 4$).

In order to test our library we used streptavidin as the macromolecular target. The motif library contained 1.5 million beads and approximately 300 coloured beads were observed and 15 were isolated for micro-sequencing. Results are shown in Table 2. Two related binding motifs l/L/m/M-HypBzl-_-Phg-_ _ and q/Q/n/N-HypBzl-_-Phg-_ _ were found and in addition two beads with the sequence F-p-YBzl-_ _ were also identified (underscores represent a non-conservative residue). Determination of the sequence is straightforward as can be seen from the HPLC chromatogram in Figure 1, where the major peaks correspond to the L-aa or D-aa while the minor peaks are index residues norvaline and norleucine respectively.

The sequences found in this non-proteinogenic peptide library are very different from peptide sequences found in all L- or D-amino acid libraries. From the screening of linear all L-aa libraries with streptavidin the binding motif HPQ- is found [2, 29]. Peptides with other motifs but having lower affinity than peptides containing HPQ-motif can be found in all L-aa libraries, e.g. if L-His is omitted in the synthesis of a pentapeptide library the dominating motifs W_A_ _ and YME_W are selected [25]. Peptides that bind streptavidin with significant higher affinity can be found in cyclic all L-aa peptide libraries on phages [33]. These peptides also contain the HPQ motif but are structurally constrained and crystal structure supports that the increase in affinity is primarily entropic since the disulphides do not interact with the protein [36, 37].

If streptavidin is screened with a linear all D-aa library the motif wy_ _a or wh_ _a appears [24,30]. In addition Lam *et al.* have reported the screening of a pentapeptide library where positions 1,3,5 were synthesized with only L-aa while positions 2 and 4 were only D-aa. In this library the motif _ _WpH dominated [24]. In the latter library the presence of either L- or D-amino acids were fixed; however, it is not possible to foresee which composition of L- or D-amino acids will be optimal for a ligand. Libraries where L- or D-amino acids or other residues are randomly intermixed are therefore preferable. This is accomplished in our indexed motif library where each residue is distributed equally randomly in the sequences, and not surprisingly the motifs found in

Table 2 Sequences that interact with streptavidin

| Group 1A[a] | Group 1B | Group 2 |
|---|---|---|
| L/m-HypBzl-Π-Phg (3) | Q/N-HypBzl-Π-Phg (4) | F-p-Y(Bzl) (2) |
| L/M-HypBzl-Ω-Phg (2) | q/n-HypBzl-Π-Phg | |
| L/M-HypBzl-Π-Phg (2) | | |
| L/M-HypBzl-$\varphi$-Phg | | |

[a]HypBzl = *O*-benzyl-L-hydroxyproline, Y(Bzl) = *O*-benzyl-L-tyrosine, Phg = L-phenylglycine, Π = mixture of polar and charged residues, Ω = a mixture of aliphatic residues, $\varphi$ = a mixture of aromatic residues (see Materials and methods). Numbers in parentheses are the number of beads detected with that sequence.

Table 3 Binding Assay on TentaGel Beads

| | Peptide Bead[a] | |
|---|---|---|
| Recepter | L/M/HypBzl-A-Phg-AA | F-p-Y(Bzl)-A-A |
| Streptavidin–AP | + + | + + + |
| Streptavidin–AP + biotin (10 μM) | – | – |
| Avidin–AP | – | – |

[a]L/M-HypBzl-A-Phg-AA and FpYBzl-AA were synthesized on TentaGel resin and assayed with streptavidin–alkaline phophatase (1:20,000) in the absence or presence of 10 μM biotin. Avidin was only assayed in the absence of biotin. – denotes no colour reaction; + + denotes moderate colour reaction; + + + denotes strong colour reaction.

the indexed non-proteinogenic motif library are different from other reported motifs. The motif from groups 1A and 1B (Table 2) had HypBzl and L-Phg in positions 2 and 4 as the most critical residues. The first position could be either l/m, L/M or q/n, Q/N. Position 3 was a mixture of residues but there was a slight preference for polar and charged residues, since 10 out of 13 beads had mixture Π in this position. Another motif F-p-YBzl was also identified on two beads and apparently this motif does not resemble the former motif. Both motifs were re-synthesized on resin to determine whether the binding between the peptides and streptavidin could be inhibited by biotin and whether avidin could bind to the peptides as well. The peptide mixture L/M-HypBzl-A-Phg-A-A-resin, where A (L-Ala) substitutes the X and Z mixtures and the sequence F-p-YBzl-A-A-resin were synthesized and screened on beads and tested for binding to streptavidin in the absence and presence of biotin (Table 3). The binding of the peptides to streptavidin was reconfirmed and the binding could be completely inhibited in the presence of 10 μM biotin, indicating that the novel motifs may bind to the biotin-binding site in streptavidin. Although streptavidin and avidin both bind very strongly to biotin, no binding could be observed when avidin was tested.

Competitive ELISA was used to compare the binding affinities in solution of the non-proteino-

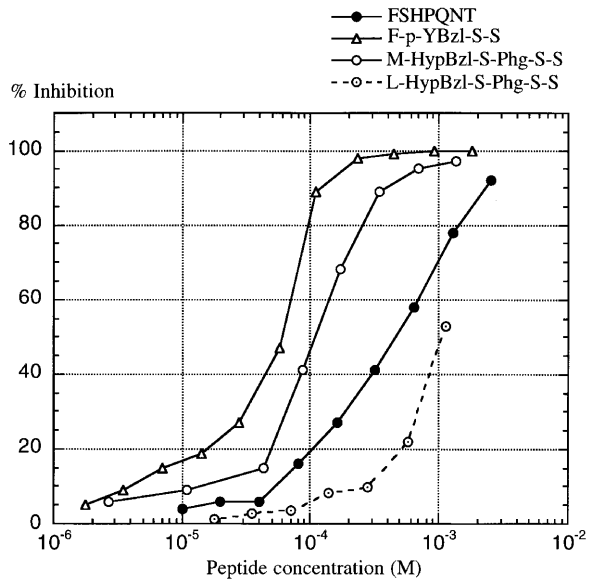


Figure 2 Inhibition of streptavidin binding to immobilized biotin by the peptides F-p-YBzl-S-S-$NH_2$ (△), M-HypBzl-S-Phg-S-S-$NH_2$ (○), L-HypBzl-S-Phg-S-S-$NH_2$ (⊙), FSHPQNT (●) as measured by ELISA.

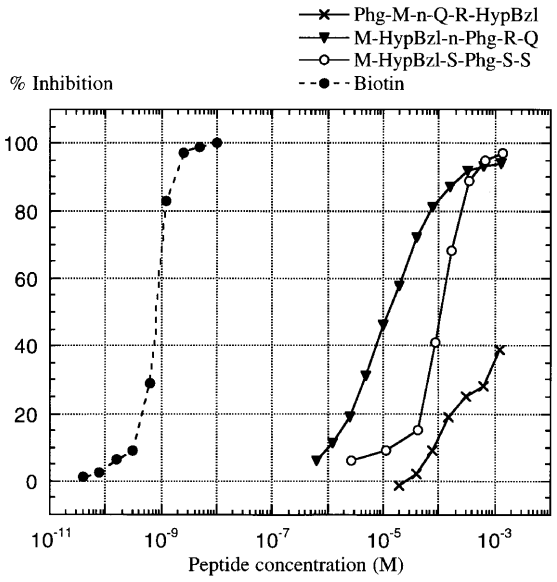


Figure 3 Inhibition of streptavidin binding to immobilized biotin by the peptide M-HypBzl-S-Phg-S-S-$NH_2$ (○) from the primary library and M-HypBzl-n-Phg-R-Q-$NH_2$ (▼) from the sub-library. Scrambled peptide Phg-M-n-Q-R-HypBzl-$NH_2$ (×) served as control and the natural ligand biotin (●) was also included.

genic binding motifs to the well-known all L-aa binding motif FSHPQNT. The peptide FSHPQNT was synthesized since this particular sequence has been used in binding experiments with streptavidin ($K_D \sim 0.13$ mM) and the crystal structure of the peptide binding to streptavidin is known [31, 33, 36]. The two non-proteinogenic motifs bind approximately four and eight times better than the all L-aa motif HPQ. $IC_{50}$ of FSHPQNT was determined to be 460 μM whereas $IC_{50}$ for peptide F-p-YBzl-S-S-$NH_2$ and peptide M-HypBzl-S-Phg-S-S-$NH_2$ were 60 and 110 μM respectively while the peptide L-HypBzl-S-Phg-S-S-$NH^2$ binds with $IC_{50} > 1$ mM.

Since only three positions could be identified in the primary library we constructed a sub-library in order to find additional conserved residues and to find out what residue (L/l/M/m/Q/q/N/n) contributes the most to binding in the [L/l/M/m/Q/q/N/n]-HypBzl-_-Phg-_ _ motif. The two motifs F-p-YBzl-_ _-resin and (L,l,M,m,Q,q,N,n)-HypBzl-_-Phg-_ _-resin was used in the construction of the dual sub-library and had the format: F-p-YBzl-O-O-Resin + [L/l/M/m/Q/q/N/n]-HypBzl-O-Phg-O-O-Resin, where O comprises 45 monomers. The sublibrary was synthesized without adding amino acids in pairs as in the primary library and Lys(Z) was replaced with D-phenylglycine (D-Phg) since it was observed that the Z-group was partially cleaved by TFA exposure. Approximately 40,000 beads with the motif FpYBzl-O-O-resin and $\sim 1.5 \times 10^6$ beads with the motif [L/l/M/m/Q/q/N/n]-HypBzl-O-Phg-O-O-resin were mixed and screened simultaneously with streptavidin. Approximately 2000–3000 beads were stained to various degrees, but only the most intensively stained beads were removed for microsequencing (see Table 4). All beads sequenced had the motif M-HypBzl-[n/q/d]-Phg-[r/R/k/K]-_-resin. Only L-Met was found as the N-terminal residue and only D-Asn, DGln, and D-Asp were found in third position confirming the results from the primary library (Table 2) that a polar or charged residue was preferred in this position. In addition a positively charged residue in either the L- or D-configuration was also significant in position 5. Peptide M-HypBzl-n-Phg-R-Q-$NH_2$ was synthesized and tested in ELISA, and showed increased binding affinity with an $IC_{50} = 12$ μM, which is a 10-fold increase compared to M-HypBzl-S-Phg-S-S-$NH_2$ and approximately 40-fold better than FSHPQNT. In comparison the natural ligand biotin bind with an $IC_{50} = 0.8$ nM.

In principle a very large number of building blocks can be included in the combinatorial library if the number of index residues is increased. As in our library, norvaline and norleucine, two index residues A, B may code for three building blocks $X_1$, $X_2$, $X_3$, e.g. A-$X_1$, B-$X_2$, AB-$X_3$, and expanding this to three index residues A, B, C such as norvaline, norleucine and α-aminobutyric acid, it would be possible to include seven building blocks which have coinciding retention time, e.g. A-$X_1$, B-$X_2$, C-$X_3$, AB-$X_4$, A C-$X_5$, BC-$X_6$, ABC-$X_7$. This allows even very large numbers of monomers to be included in the defined positions O and in principle it would be conceivable to identify the structure of 100–200 monomers by direct microsequencing of the peptide. The incorporation of norleucine or norvaline or both account for approximately of 10% of the couplings and of course this will result in a decrease of the full-length peptide. For a peptide of five residues and 10% incorporation of index in each step allows 59% ($0.9^5$) of the full-length pentapeptide to be synthesized; however, the index residues do not terminate the synthesis. Adding 10% index residues in the

Table 4 Sequences Interacting with Streptavidin Found in the Dual Sub-library F-p-YBzl-O-O-resin + [L/l/M/m/Q/q/N/n]-HypBzl-O-Phg-O-O-resin, where O comprises 45 monomers (see Materials and methods)

| M-HypBzl-n-Phg-r/R-X[a] | M-HypBzl-q-Phg-r/K-X | M-HypBzl-d-Phg-r/R-X |
|---|---|---|
| **M**-HypBzl-**n**-Phg-**r**-E | **M**-HypBzl-**q**-Phg-**r**-Q | **M**-HypBzl-**d**-Phg-**r**-G |
| **M**-HypBzl-**n**-Phg-**r**-P | **M**-HypBzl-**q**-Phg-**K**-P | **M**-HypBzl-**d**-Phg-**r**-h |
| **M**-HypBzl-**n**-Phg-**r**-q | **M**-HypBzl-**q**-Phg-**K**-H | **M**-HypBzl-**d**-Phg-**R**-P |
| **M**-HypBzl-**n**-Phg-**R**-E | | |
| **M**-HypBzl-**n**-Phg-**R**-Q | | |
| **M**-HypBzl-**n**-Phg-**R**-H | | |
| **M**-HypBzl-**n**-Phg-**k**-h | | |

[a]HypBzl = *O*-Benzyl-L-hydroxyproline; Phg = L-phenylglycine; D-amino acids in lower case.

three defined positions O as in our motif library will not seriously affect the synthesis of a motif, since approximately 73% ($0.9^3$) of a motif will be synthesized on each bead.

In comparison, the encoded library strategy [38], based on the parallel synthesis of a coding sequence, has been used to construct libraries containing (nonsequenceable) unnatural building blocks [9, 16,38–41]. By this strategy the ligand structure and the coding tag, which may be either a peptide or DNA, are associated either on the bead or with a linker. A problem is, however, that the chemistry used for the alternating synthesis of the tag and ligand must be compatible and thereby may restrict the use of chemistry associated with, for example, peptide synthesis. Additionally, there is a risk that the macromolecular target may bind to the coding part, although this can be avoided by synthesizing the encoding part only inside the bead and the ligand structure on the outside [42].

Using the index strategy, no alternating co-synthesis of tag is used and the building of a well-defined sequentially connected tagging structure is thus avoided. Index libraries using electrophoretic tags have been successfully employed and peptides as well as non-peptidic structures have been synthesized by this method [10,13,18]. Alternative structure determination can also be accomplished using MS-MALDI TOF, where small amounts of deletion peptides are generated during synthesis by simple capping procedure and subsequently analysed after release from beads [43]. This technique is, however, not suitable for the analysis of motif-libraries since these libraries contain *multiple* structures on each bead and consequently determination of the mass difference between deletion peptides is unfeasible.

## CONCLUSIONS

In this paper we used a parallel to the electrophoretic tagging strategy to synthesize a non-proteinogenic peptide library. The two structurally related residues norvaline and norleucine are randomly distributed in the peptides and the synthesis of a unique tagging structure is thus avoided, making it unlikely that the indexes will interfere with the screening assay.

In summary the results obtained from the synthesis and screening of the library with the receptor streptavidin verify that a chemically inert index residue such as norvaline and norleucine is useful for the practical synthesis and structure determination of non-proteinogenic peptides on solid supports. The results also demonstrate that it is rational to include a wider repertoire of lipophilic residues in the combinatorial peptide libraries based on the one-bead multiple-structure approach. The novel linear peptide ligands identified had higher affinity to the receptor than linear peptides found in linear all L-aa peptide libraries. Such diverse non-proteinogenic peptide libraries may be of advantage in the searching for novel ligands for other macromolecular targets when all L-aa or all D-aa peptide libraries do not afford suitable ligands.

## ACKNOWLEDGEMENTS

We thank Dr Stanley Brown for critical comments on the paper.

## REFERENCES

1. H. M. Geysen, S. T. Rodda and T. J. Mason (1986). *A priori* delineation of a peptide which mimics a discontinuous antigenic determinant. *Mol. Immunol. 23*, 709–715.
2. K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski and R. J. Knapp (1991). A new type of synthetic peptide library for identifying ligand-binding activity. *Nature 354*, 82–84.
3. M. Stanková, S. Wade, K. S. Lam and M. Lebl (1994). Synthesis of combinatorial libraries with only one representation of each structure. *Peptide Res. 7*, 292–298.
4. A. Kramer, R. Volkmer-Engert, R. Malin, U. Reineke and J. Schneifer-Mergener (1993). Simultaneous synthesis of peptide libraries on single resin and continuous cellulose membrane supports: Examples for the identification of protein, metal and DNA binding peptide mixtures. *Peptide Res. 6*, 314–319.
5. R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo (1991). Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature 354*, 84–86.
6. M. A. Gallop, R. W. Barett, W. J. Dower, S. P. A. Fodor and E. M. Gordon (1994). Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries. *J. Med. Chem. 37*, 1233–1251.
7. C. Pinilla, J. R. Appel, S. E. Blondelle, C. T. Dooley, J. Eichler, J. M. Ostresh and R. A. Houghten (1994). Versatility of positional scanning synthetic combinatorial libraries for the identification of individual

compounds. *Drug Dev. Res. 33*, 133–145. 8. N. K. Terret, D. Bojanic, D. Brown, P. J. Bungay, M. Gardner, D. W. Gordon, C. J. Mayers and J. Steele (1995). The combinatorial synthesis of a 30,752-compound library: Discovery of SAR around the endothelin antagonist, FR-139,317. *Bio. Med. Chem. Lett. 5*, 917–922.

9. V. Krchnák, A. S. Weichsel, D. Cabel and M. Lebl (1995). Linear presentation of variable side-chain spacing in a highly diverse combinatorial library. *Peptide Res. 8*, 198–205.
10. J. J. Baldwin, J. J. Burbaum, I. Henderson and M. H. L. Ohlmeyer (1995). Synthesis of a small molecule combinatorial library encoded with molecular tags. *J. Am. Chem. Soc. 117*, 5588–5589.
11. S. H. DeWitt, M. C. Schroeder, C. J. Stankovic, J. E. Strode and A. W. Czarnik (1994). DIVERSOMER™ technology: Solid phase synthesis, automation, and integration for the generation of chemical diversity. *Drug Dev. Res. 33*, 116–124.
12. J. M. Ostresh, G. M. Husar, S. E. Blondelle, B. Dörner, P. A. Weber and R. A. Houghten (1994). 'Libraries from libraries': Chemical transformation of combinatorial libraries to extend the range and repertoire of chemical diversity. *Proc. Natl Acad. Sci. USA 91*, 11138–11142.
13. M. H. J. Ohlmeyer, R. N. Swandon, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi and M. Wigler (1993). Complex synthetic chemical libraries indexed with molecular tags. *Proc. Natl Acad. Sci. USA 90*, 10922–10926.
14. R. N. Zuckermann, E. J. Martin, D. C. Spellmeyer, G. B. Stauber, K. R. Shoemaker, J. M. Kerr, G. M. Figliozzi, D. A. Goff, M. A. Siani, R. J. Simon, S. C. Banville, E. G. Brown, L. Wang, L. S. Ritcher and W. H. Moos (1994). Discovery of nanomolar ligands for the 7-transmembrane G-protein-coupled receptors from a diverse *N*-(substituted)glycine peptoid library. *J. Med. Chem. 37*, 2678–2685.
15. M. Stanková, O. Issakova, N. F. Sepetov, V. Krchnák, K. S. Lam and M. Lebl (1994). Application of one-bead one structure approach to identification of nonpeptidic ligands. *Drug Dev. Res. 33*, 146–156.
16. J. M. Kerr, S. C. Banville and R. N. Zuckermann (1993). Encoded combinatorial peptide libraries containing non-natural amino acids. *J. Am. Chem. Soc. 115*, 2529–2531.
17. J. Eichler, A. W. Lucka and R. A. Houghten (1994). Cyclic template combinatorial: Synthesis and identification of chymotrypsin inhibitors. *Peptide Res. 7*, 300–307.
18. A. Borchardt and W. C. Still (1994). Synthetic receptor binding elucidated with an encoded combinatorial library. *J. Am. Chem. Soc. 116*, 373–374.
19. M. C. Pirrung and J. Chen (1995). Preparation and screening against acetylcholinesterase of a non-peptide 'indexed' combinatorial library. *J. Am. Chem. Soc. 107*, 1240–1245.
20. B. A. Bunin, M. J. Plunkett and J. A. Ellman (1994). The combinatorial synthesis and chemical and biological evaluation of a 1,4-benzodiazepine library. *Proc. Natl Acad. Sci. USA*, 4708–4712.
21. D. A. Campbell, J. C. Bermak, T. S. Burkoth and D. V. Patel (1995). A transition state analogue inhibitor combinatorial library. *J. Am. Chem. Soc. 117*, 5381–5382.
22. N. F. Sepetov, V. Krchnák, M. Stanková, S. Wade, K. S. Lam and M. Lebl (1995). Library of libraries: Approach of synthetic combinatorial library design and screening of 'pharmacophore' motifs. *Proc. Natl Acad. Sci. USA 92*, 5426–5430.
23. À. Furka, F. Sebestyán, M. Asgedom and G. Dibó (1991). General method for rapid synthesis of multicomponent peptide mixtures. *Int. J. Peptide Protein Res. 37*, 487–493.
24. K. S. Lam, M. Lebl, S. Wade, A. Stierandova, P. S. Khattri, C. N. and V. J. Hruby in: *Peptides: Chemistry, Structure and Biology*, R. S. Hodges and J. A. Smith, Es. p. 1005–1006, ESCOM, Leiden 1994.
25. K. S. Lam and M. Lebl (1992). Streptavidin and avidin recognize peptide ligands with different motifs. *Immunomethods 1*, 11–15.
26. T. G. M. Schmidt and A. Skerra (1993). The random peptide library assisted engineering of a C-terminal affinity peptide, useful for the detection and purification of a functional Ig Fv fragment. *Protein Eng. 6*, 109–122.
27. M. A. McLafferty, R. B. Kent, R. C. Lander and W. Markland (1993). M13 bacteriophage displaying disulfide-constrained microproteins. *Gene 128*, 29–36.
28. B. K. Kay, N. B. Adey, Y.-S. Hem J. P. Manfredi, A. H. Mataragnon and D. M. Fowlkes (1993). An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene 128*, 59–65.
29. J. J. Devlin, L. C. Panganiban and P. E. Devlin (1990). Random peptide libraries: A source of specific protein binding molecules. *Science 249*, 404–406.
30. S. Østergaard, H. P. Hansen, M. Olsen and A. Holm (1995). Novel avidin and streptavidin binding sequences found in synthetic peptide libraries. *FEBS Lett. 362*, 306–308.
31. P. C. Weber, M. W. Pantoliano and L. D. Thompson (1992). Crystal structure and ligand-binding studies of a screened peptide complexed with streptavidin. *Biochemistry 31*, 9350–9354.
32. B. Gissel, M. R. Jensen, K. Gregorius, H. I. Elsner, I. Svendsen and S. Mouritsen (1995). Identification of avidin and streptavidin binding motifs among peptides selected from a synthetic peptide library solely of D-amino acids. *J. Peptide Sci. 1*, 217–226.
33. L. B. Giebel, R. T. Cass, D. L. Milligan, D. C. Young, R. Arze and C. R. Johnson (1995). Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry 34*, 15430–15435.

34. J. Blake and L. Litzi-Davis (1992). Evaluation of peptide libraries: An iterative strategy to analyze the reactivity of peptide mixtures with antibodies. *Bioconjugate Chem. 3*, 510–513.
35. K. Burgess, N. Wang and A. I. Liaw (1994). Combinatorial technologies involving reiterative division/coupling/recombination: Statistical considerations. *J. Med. Chem. 37*, 2985–2987.
36. B. A. Katz (1995). Binding to protein targets of peptidic leads discovered by phage display: Crystal structures of streptavidin-bound linear and cyclic peptide ligands containing the HPQ sequence. *Biochemistry 34*, 15421–15429.
37. B. A. Katz, C. R. Johnson and R. T. Cass (1995). Structure-based design of high affinity streptavidin binding cyclic peptide ligands containing thioether cross-links. *J. Am. Chem. Soc. 117*, 8541–8547.
38. S. Brenner and R. A. Lerner (1992). Encoded combinatorial chemistry. *Proc. Natl Acad. Sci. USA 89*, 5381–5383.
39. M. Needles, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett and M. A. Gallop (1993). Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl Acad. Sci. USA 90*, 10700–10704.
40. J. Nielsen, S. Brenner and K. D. Janda (1993). Synthetic methods for the implementation of encoded combinatorial chemistry. *J. Am. Chem. Soc. 115*, 9812–9813.
41. V. Nikolaiev, A. Stierandová, V. Krchnák, B. Seligmann, K. S. Lam, S. E. Salmon and M. Lebl (1993). Peptide-encoding for structure determination of non-sequenceable polymers within libraries synthesized and tested on solid-phase supports. *Peptide Res. 6*, 161–170.
42. M. Lebl, V. Krchnák, N. F. Sepetov, B. Seligmann and S. Felder (1995). One-bead-one-compound combinatorial libraries. *Biopolym. (Peptide Sci.) 37*, 177–198.
43. R. S. Youngquist, G. R. Fuentes, M. P. Lacey and T. Keough (1994). Matrix-assisted laser desorption ionization for rapid determination of the sequences of biologically active peptides isolated from support-bound combinatorial peptide libraries. *Rapid Commun. Mass Spectrom. 8*, 77–81.

 