Peptide Research

Peptide-Encoding for Structure Determination of Nonsequenceable Polymers Within Libraries Synthesized and Tested on Solid-Phase Supports

V. Nikolaiev, A. Stierandová, V. Krchňák, B. Seligmann, K.S. Lam[1], S.E. Salmon[1] and M. Lebl
Selectide Corporation and
[1]Arizona Cancer Center

ABSTRACT

A method of indirectly determining the structure of nonpeptidic or nonsequenceable compounds that have been synthesized on individual particles of solid support is described. The technique permits the parallel synthesis of a compound that is not susceptible to Edman degradation (e.g., N-terminal-blocked peptide), or one containing components that cannot be identified by amino acid sequencing, together with a corresponding "coding" peptide. Each coupling step in the assembly of the nonsequenceable compound is followed by the coupling of an amino acid to a different attachment site of the same carrier particle, whereby the amino acid unambiguously codes for the previously coupled building block of the nonsequenceable compound. The rationale is to enable the sequence determination of a biologically active compound that has been identified through the screening of a library of nonsequenceable compounds, by translating the sequence of its "coding" peptide, determined by Edman degradation, into the structure of the active compound. The technique facilitates the construction and screening of nonpeptidic libraries for the discovery of important pharmaceutical compounds.

INTRODUCTION

The process of identifying new structures that interact with pharmacologically important macromolecular targets has been significantly accelerated by the introduction of so-called "library techniques" (4,6,9,15,20; for reviews see 10,22,24). These approaches are designed to generate and readily screen a multiplicity of structures that interact with a given acceptor (receptor, enzyme, antibody). There are in principle two approaches to screening the libraries: serial and parallel. In the serial approach [Geysen et al. (6) and Houghten et al. (9)], a multiplicity of semi-random libraries is generated and tested. These libraries consist of groups of peptide mixtures with one or two defined positions. The most active of these groups in a given bioassay is selected for an iterative process of synthesis and screening, during which all of the positions of the active sequence are successively defined. Since stepwise sequence-revealing is inherent to this approach, as well as to the concept of positional scanning libraries (23), which consists of n groups of peptide mixtures (n = number of residues in the sequence) with another position defined in each group, serial approaches can be readily applied to the generation and screening of libraries of nonpeptide or nonsequenceable compounds without modified methods for sequence determination.

The parallel approach, exemplified either by phage display screening (20, for review see 22) or by the Selectide technology (15), is based on simultaneous screening of millions of unique structures (prepared by simple modification (5,15) of the solid-phase peptide synthesis principle (18)) separated in space but contained in the same reaction vessel. After identifying and isolating the active particle, it is necessary to determine the structure of the molecule responsible for the given activity. The application of nonpeptide structures for library building is possible using the Selectide technology, but is limited by the ability to determine the structure of the active molecule, which is usually available in a limited amount (the amount present in one particle of the carrier).

The rationale for the construction and screening of nonpeptide libraries is obvious. The goal of most drug-discovery programs is the development of a specific ligand for a macromolecular target. The ability to mimic peptide–protein interaction with nonpeptide structures was demonstrated by several authors (7,8,19,25–27). Therefore, it would be desirable to construct and screen libraries based on building blocks capable of all major types of interactions (ionic, hydrogen bonding, hydrophobic, charge-transfer, chelation, aromatic, etc.), and assuming a wide variety of tertiary structures, in addition to those created by peptide bonds.

Coding for the structure of an active molecule is not itself a novel concept. The phage-display technique uses the nucleic acids of the phage to code for the sequence of the interacting peptide. Sequencing of nucleic acids attached

Abbreviations: The one-letter notation of amino acid residues is used in this manuscript for description of amino acid residues and peptide sequences. Derivatives of amino acids are represented by three-letter codes. Other abbreviations used: Boc, *tert*-butyloxycarbonyl; BOP, benzotriazolyloxy-trisdimethylamino-phosphonium hexafluorophosphate; CHA, cyclohexylamine; DCHA, dicyclohexylamine; DCM, dichloromethane; Ddz, 3,5-dimethoxy-dimethylbenzyloxycarbonyl; DIC, diisopropylcarbodiimide; DIEA, diisopropylethylamine; DMF, dimethylformamide; DMPU, 1,3-dimethyl-3,4,5,6-tetrahydro-2(1H)-pyrimidinone (dimethylpropyleneurea); Fmoc, fluorenylmethoxycarbonyl; HOBt, *N*-hydroxybenzotriazole; HPLC, high pressure liquid chromatography; PyBrOP, bromo-tris-pyrrolidinophosphonium hexafluorophosphate; SCAL, safety-catch amide linker; TFA, trifluoroacetic acid; Tfa, trifluoroacetyl.

directly to peptide molecules was suggested by Brenner and Lerner (2) for determination of a structure of a peptide in a library. However, successful co-synthesis of nucleic acids and peptides requires novel chemistry which has not yet been described. [The use of peptides to code for nonnatural amino acid-containing peptides in peptide libraries in solution was published (11) while this manuscript was in preparation.] Methods available for protein sequencing require only femtomolar quantities and are therefore well suited for the technology employing screening and sequencing of pure peptides synthesized on the polymeric beads (15). There are no other types of polymeric molecules besides DNA and peptides for which techniques of structure determination have been well developed. Biophysical techniques of molecular structure determination are significantly less sensitive (MS) and/or less powerful (NMR, IR, x-ray), and typically much slower and more labor intensive.

MATERIALS AND METHODS

Instrumentation

Fast atom bombardment (FAB) mass spectroscopic measurements were carried out on a ZAB EQ spectrometer (VG Analytical Ltd, Manchester, UK). ^{1}H NMR spectra were obtained on a General Electric QE 300 instrument (Fullerton, CA). Sequencing by Edman degradation was performed on an ABI 4778 protein sequencer (Applied Biosystems, Foster City, CA) and a Porton PI 3010 instrument (Porton Instruments, Tarzana, CA). Both analytical and preparative HPLC were carried out on a Millipore 625 LC system (Bedford, MA) with a Millipore 490E Programmable Multiwavelength Detector using Vydac Peptide and Protein C18 analytical (4.6 × 250 mm, 5 μm, 1 ml/min) and preparative (10 × 250 mm, 10 μm, 3 ml/min) columns (The Separations Group, Hesperia, CA), respectively. Analyses of mixtures released from one bead were performed on an Ultrafast Microprotein Analyzer (Michrom BioResources, Pleasanton, CA) using a Reliasil C18 column (5 μm, 300 Å, 1 × 150 mm). All spectra are reported in ppm relative to tetramethylsilane (δ) using either $CDCl_3$ or CD_3SOCD_3 as solvents. UV/VIS absorption spectra were recorded on a Hewlett Packard HP 8452A Diode-Array spectrophotometer (Palo Alto, CA) using a 1-cm quartz cuvette. Amino acid analyses were carried out on a D-500 system (Durrum, Palo Alto, CA).

General Procedures

Solid-phase synthesis was performed manually in polypropylene syringes as described by Krchňák and Vagner (13). Syntheses were performed on TentaGel (TG) resin (Rapp Polymere, Tubingen, Germany) (130 or 80 μm, 0.23 mmol/g) modified with SCAL handle (21) (safety-catch amide linker) or with an appropriate linker. Fmoc-protecting groups were cleaved with 50% piperidine/DMF for 1 × 10 min. Boc groups were cleaved for 20 min with 30% TFA/DCM containing 3% anisole. Ddz groups were cleaved for 30 min with 2% TFA/DCM. After Boc cleavage, a solution of DIEA/DCM (10%) was used for neutralization. A mixture of BOP/ HOBt/DIEA (1:1:2 eq) in DMF was used for the activation of both Nα-Fmoc and Boc amino acids. The completeness of each condensation reaction (1.5–40 h) was checked by the ninhydrin test or by the chloranil test in the cases of coupling to secondary amino groups. The coupling protocol included washing with DMF (6–8 times) [followed by washing with DCM in the case of Boc-protected amino acids] between coupling and deprotection, and between deprotection and coupling. The SCAL linker was reduced by 20% $(EtO)_2P(S)SH$ in DMPU for 2 h. Final cleavage was done by 95% TFA–5% water mixture.

Materials

Commercial-grade solvents were used without further purification. Protected amino acids were obtained from Bachem (Torrance, CA), Advanced ChemTech (Louisville, KY) or Propeptide (Vert-le-Petit, France).

Models of a Peptide Library Encoded by a Peptide Sequence

Boc-Lys(Fmoc)-OH was coupled as the first amino acid to SCAL-TG, the N^{ε}-Fmoc group was deprotected and Fmoc-Lys(Fmoc)-OH was coupled to the side chain of the first lysine. The N^{α}- and N^{ε}-Fmoc groups of lysine were cleaved and the resin was divided into three parts. Fmoc-Ala-OH, Fmoc-Phe-OH and Fmoc-Val-OH, respectively, were each coupled to one portion of the resin. The corresponding Boc amino acids (Gly, Tyr and Leu — Boc-Tyr-OH was used with the unprotected hydroxyl group) were coupled in the next step to the α-amino group of lysine after Boc deprotection, while the "Fmoc branch" remained protected. After completion of Boc amino acid condensations, all three portions of the resin were combined, and the "Fmoc branch" was deprotected. The following randomization was performed exactly the same way as the first one after the splitting of the resin into three equal portions. After randomization of three positions (coupling of three different amino acids in each position), the resin was divided into smaller parts and treated differently.

1. Cleavage of both N-terminal Fmoc- and Boc-protecting groups

Two completely deprotected beads were separately submitted for sequence analyses. Correct "complementary" amino acids were found in all three cycles in the expected ratio 2:1. Results (values in pmol): *1st bead:* 1st cycle: V 251, L 146; 2nd cycle: V 244, L 147; 3rd cycle: V 245, L 119. *2nd bead:* 1st cycle: A 102, G 39; 2nd cycle: V 121, L 59; 3rd cycle: F 125, Y 50.

Part of the resin (about 100 mg) was treated with 20% diethyldithiophosphate in DMPU (2 × 1 h shaking) to reduce the SCAL handle. The mixture of peptides was cleaved from the reduced SCAL with TFA/H_2O (95:5) for 1 h. The cleavage mixture was concentrated *in vacuo* and precipitated with Et_2O. The precipitate was collected by centrifugation and dried. The mixture of peptides was dissolved in 0.1% TFA/H_2O and analyzed by HPLC. First, a faster gradient 0%–100% of acetonitrile and 0.1% TFA in 100 min was run for orientation. A slow gradient 0%–50% of acetonitrile and 0.1% TFA over 200 min revealed that the expected 27 peaks were present. Several additional minor peaks were identified, the formation of which was attributed to the use of side-chain unprotected tyrosine during the synthesis. Because of possible losses of some hydrophobic sequences during the ether precipitation, the second cleavage of the

mixture avoided this step. The cleavage mixture of TFA and water was diluted by additional water, concentrated on an evacuated centrifuge and lyophilized. HPLC evaluation of the mixture demonstrated an equimolar representation of all expected peaks.

2. Deprotection of the N-terminal Fmoc group and acetylation of the "Fmoc branch"

The free N-terminal amino groups were acetylated with a 0.3 M solution of *N*-acetylimidazole in DMF for 20 min (ninhydrin test negative). N-terminal Boc groups on the other branch were deprotected after acetylation. Three randomly chosen beads were sequenced and provided the following readings (values in pmol): 1st cycle: Y 213 (bead 1), G 161 (bead 2), Y 201 (bead 3); 2nd cycle: L 165 (1), Y 166 (2), Y 205 (3); 3rd cycle: Y 188 (1), L 128 (2), G 162 (3). The readings were not contaminated by the amino acids present in the acetylated arm.

A part of the acetylated beads (about 100 mg) was treated the same way as described above (precipitation of the mixture by ethyl ether and/or evaporation of cleavage mixture and lyophilization) to reduce the handle and cleave the acetylated peptides. HPLC analysis under the same conditions has shown that during ether precipitation a significant proportion of the library was lost due to its solubility in ether. Evaporated and lyophilized sample provided the same number of peaks of approximately the same pattern as in the case of unprotected library, with retention times shifted to higher values.

3. Replacement of the Boc-protecting group by the Tfa group

The trifluoroacetyl group was introduced as a protecting group instead of the Boc group at the N-terminus in order to permit stepwise sequencing experiments. First, the N-terminal Boc group was cleaved from a resin sample (50 mg) while the "Fmoc branch" was left protected. The free amino groups were protected with trifluoroacetyl by treatment with 10 equivalents (0.14 mmol, 21 µl) of trifluroacetic acid anhydride in dichloromethane (0.5 ml) in the presence of DIEA (0.16 mmol, 28 µl). The reaction was complete after 20 min (ninhydrin test negative). After the trifluoroacetylation, the Fmoc group on the other branch was removed and three beads were submitted for sequencing. The sequence of the Fmoc branch was determined, support used for the sequencing was removed from the sequencer, and the support bead assembly was treated with a 0.2 M solution of NaOH (24 h, 20°C), dried and submitted for an additional three cycles of sequencing. The appropriate sequences predicted from the sequencing of the Fmoc branch were obtained. *1st bead* (values in pmol): 1. F (735), 2. V (643), 3. A (837); after TFA removal: 1. Y (207), 2. L (187), 3. G (76), sequence FVA/YLG; *2nd bead*: 1. A (215), 2. A (230), 3. F (193); after TFA removal: 1. G (88), 2. G (86), 3. Y (80), sequence AAF/GGY; *3rd bead*: 1. F (63), 2. F (67), 3. V (41); after TFA removal: 1. Y (15), 2. Y (12), 3. L (4), sequence FFV/YYL.

4. Cleavage of a peptide from one bead

Several beads of the resin containing fully deprotected sequences on the reduced SCAL handle (see point 1) were placed separately into small glass vials and treated overnight with 30 µl of neat TFA. Aliquots (3 µl) were withdrawn and diluted with H_2O to the total volume of 20 µl and analyzed by HPLC on a microbore HPLC (Michrom apparatus) (gradient 5%–60% of acetonitrile in 0.1% TFA in water over 20 min). Calculation based on the average extinction coefficient of peptides at 215 nm have shown that about 100–200 pmol of peptide was released from one polymeric bead.

Non-peptide Screening Libraries Encoded by a Peptide Structure

Couplings of amino acids were performed by a manual method using standard protocol at room temperature: protected amino acid (3 eq) in DMF was mixed with DIC (3 eq) (or DIC and HOBt [3 eq each] and the resin), and coupling was followed by analytical tests. Symmetric anhydrides were used where specified.

Fmoc-SCAL linker and Boc-Lys(Fmoc) were coupled first to the resin (TentaGel S NH_2, 1 g) using DIC and HOBt. After cleavage of the Fmoc group, Fmoc-Trp was coupled and Fmoc deprotected. The peptide-resin was divided into three equal portions, and three different bromoacids (one in each reaction vessel, 3 eq each) were coupled using DIC in DMF (3 eq). The three acids were bromoacetic, α-bromovaleric and α-bromo-p-toluic acid. The coupling of the last acid was repeated because of its low reactivity, using a 6-fold excess of both acid and DIC. Boc protection of the α-amino group of Lys was removed by TFA, and the first coding sequence Boc-protected amino acids (Gly, Ala, Leu) were coupled by DIC. Coding amino acids were chosen according to the molecular weight of the nonpeptide building blocks, so that the lightest block (bromoacetic acid, in this case) was coded by Gly; the heaviest one (bromotoluic acid) was coded by Leu; and medium-weight one (bromovaleric acid) by Ala.

The three resin parts were pooled together, washed thoroughly with DCM and deprotected by TFA/DCM in preparation for the coupling of the next coding amino acids. After deprotection, the resin was divided again into three portions. Couplings of Boc-protected amino acids (again Gly, Ala and Leu) were performed as usual by means of DIC. Two parts of the resin were treated with 2-M solutions of amines (benzylamine and 1-amino-4-methylpiperazine) in DMF overnight. The third part was treated with a 2-M solution of fluorenylmethyloxycarbonyl-2-aminoethanethiol, and after completion of the reaction the Fmoc group was removed. Coding of amines was based again on their molecular weights.

The resin was pooled together once more, mixed and divided into three parts for the final couplings. Carboxylic acids (cyclohexylacetic acid, phenyloxyacetic acid and 4-pyridylthioacetic acid) were coupled to amines obtained (primary and secondary) by DIC, and the coupling reactions were repeated twice using pre-formed symmetrical anhydrides in 3-fold to 5-fold excess. After obtaining a negative chloranil test, the three batches of resin were treated separately by TFA, neutralized, and the last coding Boc-protected amino acids were coupled using DIC and HOBt. Coding of the last carboxylic acids was based on the same scheme as before. Finally, all the resin was pooled together.

Two "unnatural" libraries have been completed using this general approach. The only difference between those two is the location of the SCAL linker. In

the first library, the SCAL linker is attached to the N^{E} of Lys attached directly to the resin, and therefore Trp amide was the last amino acid in all the compounds of this library. The coding peptides remained on the resin beads after cleavage. The synthesis of the second library started with attaching the SCAL linker to the resin, and the last amino acid in all the compounds was Lys. All the compounds released from this library included also their coding sequence peptides.

Beads from the first library (exclusive, A) were treated with the reducing agent and individual beads were picked up for separate cleavage and sequence analyses. Five beads were studied. After cleavage of the nonpeptide part, the beads were successfully sequenced (see Table 1) and the structure of the nonpeptide compound could be deduced. Solutions containing the cleaved compounds were analyzed on a micro-HPLC system.

A sample (800 mg) of the second library (inclusive, B) was treated with 95% TFA after reduction of the SCAL linker, freeze-dried, dissolved in water, and separated on a semi-preparative HPLC column yielding 44 peaks using a gradient of 0%–60% acetonitrile in 0.1% TFA in water over 200 min. (The higher-than-predicted number of peaks observed in the HPLC trace can be explained by synthetic problems experienced in the solid-phase nonpeptide couplings. This conclusion is supported by the analysis of products obtained in the synthesis of individual components of this model library.) The fractions representing the different peaks were lyophilized, and several peaks were analyzed by FAB MS and sequencing to show the correspondence between the structure predicted from the amino acid coding sequence and the molecular weight of the construct. Examples of the results: *Peak 4:* RT 25.31 min, sequencing: 1. Leu (364 pmol), 2. Gly (139), 3. Gly (422); FAB MS - 827.0 (building block combination 169); *Peak 8:* RT 28.69 min, sequencing: 1. Gly (261), 2. Leu (176), 3. Ala (225); FAB MS - 770.2 (building block combination 257 w/o block 7); *Peak 13:* RT 31.47 min, sequencing: 1. Leu (792), 2. Leu (551), 3. Gly (128); FAB MS - 921.0 (building block combination 159); *Peak 14:* RT 32.27 min, sequencing: 1. Leu (7930), 2. Gly (1810), 3. Ala (1763); FAB MS - 883.0 (building block combination 269); *Peak 15:* RT 32.77 min, sequencing: 1. Leu (784), 2. Ala (447), 3. Ala (360); FAB MS - 776.2 (building block combination 249 w/o block 9); *Peak 16:* RT 33.16 min, sequencing: 1. Leu (1286), 2. Ala (918), 3. Ala (688); FAB MS - 776.2 (building block combination 249 w/o block 9); *Peak 17:* RT 33.51 min, sequencing: 1. Leu (298), 2. Leu (280), 3. Ala (202); FAB MS - 826.2 (building block combination 259 w/o block 9); *Peak 19:* RT 34.80 min, sequencing: 1. Leu (641), 2. Gly (412), 3. Ala (460); FAB MS - 883.1 (building block combination 269); *Peak 20:* RT 36.66 min, sequencing: 1. Leu (150), 2. Leu (119), 3. Leu (80); FAB MS - 902.2 (building block combination 359 w/o block 9); *Peak 26:* RT 41.77 min, sequencing: 1. Gly (39), 2. Gly (38), 3. Gly (23); FAB MS - 744.1 (building block combination 167); *Peak 31:* RT 48.86 min, sequencing: 1. Ala (180), 2. Gly (98), 3. Ala (106); FAB MS - 824.0 (building block combination 268); *Peak 32:* RT 49.46 min, sequencing: 1. Leu (234), 2. Leu (320), 3. Ala (277); FAB MS - 826.1 (building block combination 259 w/o block 9); *Peak 33:* RT 50.70 min, sequencing: 1. Gly (152), 2. Gly (120), 3. Ala (94); FAB MS - 800.1 (building block combination 267).

Table 1. Structures Contained on Randomly Selected Beads from a Library of Nonpeptide Structures

Bead No.	Building Block Combination	M.W. (m/z)	Amino Acid Detected (pmol)		
			1st cycle	2nd cycle	3rd cycle
1	149	501.6	L (50)	A (55)	G (72)
2	169	471.6	L (34)	G (31)	G (29)
3	258	534.7	A (101)	L (83)	A (98)
4	147	474.6	G (45)	A (41)	G (25)
5	167	444.6	G (39)	G (30)	G (22)

Synthesis of Representative Compounds from the Nonpeptide Library

A component of the exclusive library A, compound I, was synthesized on 0.23 g of Knorr resin (Bachem, Bubendorf, Switzerland) (0.5 meq/g). Fmoc-Trp was coupled first according to the general protocol, using DIC and HOBt. After deprotection of the amino group, α-bromoacetic acid (50 mg) was coupled using DIC (50 µl) in DMF (0.5 ml). Benzylamine (100 µl) was dissolved in 0.5 ml of DMSO and the bromoresin was treated with this solution overnight. The final carboxylic acid, 4-pyridylthioacetic acid (80 mg), was dissolved in 0.85 ml of DMPU and preactivated with DIC (80 µl) and HOBt (80 mg), and coupled to the aminoresin for 10 h. The coupling was repeated using PyBrOP and DIEA for activation. Compound I was cleaved by 95% TFA. After cleavage, TFA was evaporated *in vacuo* and the residue was dissolved in 30% aqueous acetonitrile, and lyophilized. The product obtained after drying was redissolved in neat acetonitrile and precipitated by ether. This operation was repeated twice and an almost white precipitate was obtained. The product showed two peaks on RP HPLC from which the second one gave correct mass-spectrum. Yield after purification on semi-preparative RP HPLC — 18 mg. Formula: $C_{27}H_{27}N_5O_3S$; MS expected — 501.6; MS found — 502.2 $(M+H)^+$. ^1H NMR data (DMSO-d6): 10.804 d (1H, $N^{in}H$); 8.49 d (2H, pyridyl C_2H and C_6H); 8.35 d (1H, NH); 7.62 d (2H, pyridyl C_3H and C_5H); 6.9–7.7 mm (Bzl and Trp aromatic protons); 4.59 m (1H, Trp $C^{\alpha}H$); 3.75–4.65 m (aliphatic protons); 3.19 dd and 2.91 dd (2H, Trp $C^{\beta}H$).

The second component of the exclusive library, compound II, was synthesized according to the same scheme as compound I, using α-bromovaleric acid (40 µl), 4-methyl-1-aminopiperazine (100 µl) and cyclohexylacetic acid (80 mg) as building blocks. Formula: $C_{29}H_{44}N_6O_3$; MS expected — 524.7; MS found — 525.3 $(M+H)^+$, 558.2 $(M+Na)^+$ and 573.2 $(M+K)^+$.

The third component of the exclusive library, compound III, was synthesized according to a scheme similar to compound I, using α-bromo-p-toluic

acid (120 mg), fluorenylmethyloxycarbonyl-2-aminoethanethiol (280 mg) (deprotection after coupling with piperidin/DMF) and phenoxyacetic acid (80 mg) as building blocks. Formula: $C_{29}H_{30}N_4O_4S$, MS expected — 530.6; MS found — 553.0 $(M+Na)^+$.

Model Library Synthesis:

*Library: XXXX-Lys(XXXX)-Lys(ZZ)-*βAla-Gly-βAla-Gly-TG

The library was synthesized according to the following protocol.

1. Coupling of Fmoc-Lys(Boc) to H-βAla-Gly-βAla-Gly-TG; 2. Fmoc cleavage; 3. Coupling of Fmoc-Lys(Fmoc); 4. Boc cleavage; 5. The resin was divided into 9 parts and the following Ddz-protected amino acids were coupled: A,D,I,K,M,N,S,T,V; 6. Fmoc cleavage; 7. Coupling of nine Fmoc-protected amino acids: Y,G,F,L,H,P,Q,R,E (Y was coupled to that part of resin that had already attached A, etc.); 8. Resin combined and Ddz cleaved; 9. Repeat steps 5–7; 10. Fmoc cleavage; 11. Coupling of nine Fmoc-protected amino acids: Y,G,F,L,H,P,Q,R,E; 12. Repeat steps 10–11; 13. Fmoc cleavage; 14. Side-chain protecting groups and Ddz removed by mixture K (12).

One bead was submitted to four cycles of Edman degradation: *1st cycle:* Arg (64), Ile (67); *2nd cycle*: Gly (45), Thr (14); *3rd cycle:* Phe (42); *4th cycle:* Arg (35). Ile was coupled Ddz-protected and found in the first cycle. It coded for Phe, which was detected in the third cycle. In the second cycle Thr was detected as the amino acid that had been coupled Ddz protected. Arg, coded by Thr, was accordingly found in the fourth cycle of sequencing.

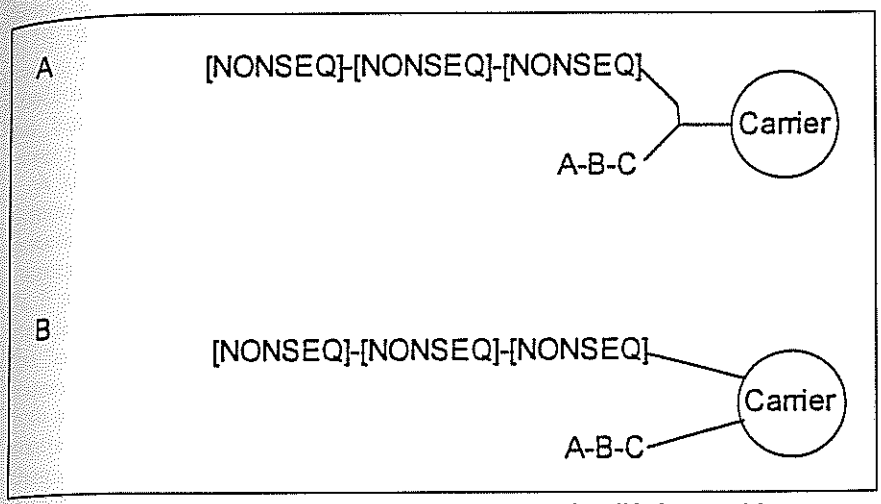


Figure 1. Alternative ways of attachment of coding sequence to the solid-phase particle.

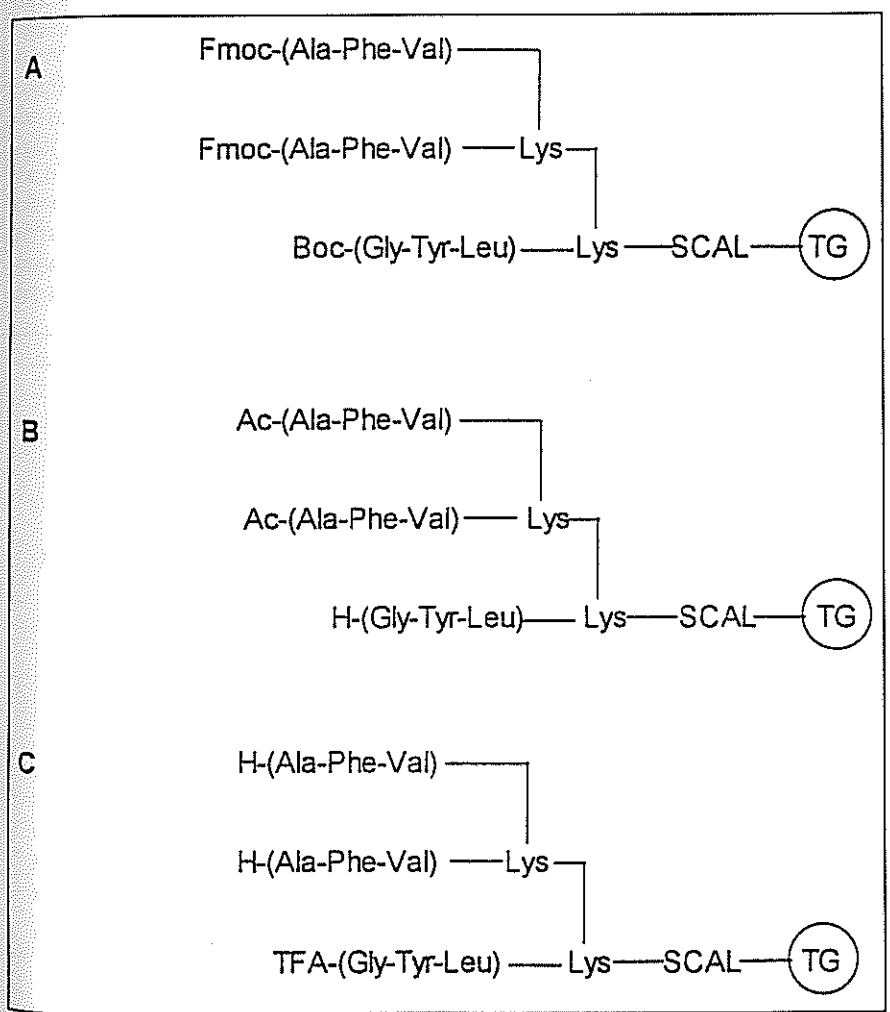


Figure 2. Structure of model libraries carrying the coding sequence.

Screening Protocol of the Library

The peptide library was screened according to the published procedure (14). The peptide beads were first mixed with incrementally increasing double-distilled water to remove the DMF. After extensive washing with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.2), the beads were coated with 0.05% gelatin (wt/vol) to block any nonspecific binding. The beads were then incubated with a 1:100 000 dilution of streptavidin-alkaline phosphatase at 2 mg/ml (Pierce, Rockford, IL) in 2× PBS/Tween/gelatin (2× PBS, 0.1% Tween-20 (vol/vol) and 0.05% gelatin (wt/vol). The beads were then thoroughly washed with TBS (137 mM NaCl, 2.7 mM KCl, 25 mM Tris base, pH 7.4) and the standard substrate 5-bromo-4-chloro-3-indolyl phosphate was added. The beads, together with the substrate, were then transferred to petri dishes for color development. After 30 min to 1 h, the colored beads were collected with the aid of a micropipet, washed with 6 M guanidine hydrochloride, pH 1.0, and subjected to sequencing as described (14,15).

The remaining library of colorless beads were then recycled with 6 M guanidine hydrochloride, pH 1.0, thoroughly washed with PBS, and incubated overnight with 60 pM biotinylated anti-β-endorphin (clone 3-E 7; Boehringer Mannheim, Mannheim, Germany) in 2× PBS/Tween/gelatin. After thorough washing, streptavidin-alkaline phosphatase was added. One hour later, the beads were washed, the substrate was added and color development proceeded as described above. The colored beads were then physically isolated and subjected to sequencing. In these two experiments, only the darkest beads were sequenced. The results are given in Figure 6.

RESULTS AND DISCUSSION

The principle of the technique for coding described in this paper is illustrated in Figure 1. The structure of the "screening" arm may be constructed from the unnatural (or nonsequenceable) amino acids, or the screening arm may be constructed by connecting various building blocks using different chemistries. In any case, given a composition of the screening arm that does not undergo Edman degradation, the sequence information from the coding sequence is sufficient to determine the structure of the screening arm. The building blocks in the screening arm must be unequivocally associated with the amino acids and positions in the coding arm. To cover a wider range of building blocks in the screening arm, more than one amino acid per building block can be used for coding. Thus, using doublets of 20 natural amino acids, 400 unnatural building blocks can be encoded. Selective cleavage of both arms may be performed by the use of specific linkers, and in this way the separate determination of any effect of the coding arm on the biological test system is possible.

There are several approaches to assembling the coding sequence. In the first one (Figure 1A), both screening and coding structures are built on the branched attachment to the solid support, realized, for example, by diamino carboxylic acid (lysine). Both "sequences" are present in the defined molar ratio and a defined spatial arrangement accessible to the acceptor molecule being screened. A second

Table 2. Combinations of Building Blocks Used in the Construction of a Model Nonpeptide Library

Combination	M.W. No Coding[a]	M.W. With Coding[b]	Coding Sequence
167	444.6	743.9	GGG
168	454.5	767.9	AGG
169	471.6	827.0	LGG
147	474.6	787.9	GAG
148	484.5	811.9	AAG
149	501.6	871.1	LAG
157	482.6	838.0	GLG
158	492.6	862.0	ALG
159	509.6	921.2	LLG
267	486.7	800.0	GGA
268	496.6	824.0	AGA
269	513.7	883.1	LGA
247	516.7	844.1	GAA
248	526.6	868.0	AAA
249	543.7	927.2	LAA
257	524.7	894.1	GLA
258	534.7	918.1	ALA
259	551.7	977.3	LLA
367	520.7	876.1	GGL
368	530.6	900.1	AGL
369	547.7	959.2	LGL
347	550.7	920.1	GAL
348	560.6	944.1	AAL
349	577.7	1003.3	LAL
357	558.7	970.2	GLL
358	568.7	994.2	ALL
359	585.7	1053.3	LLL

[a]From library in Figure 4B
[b]From library in Figure 4A

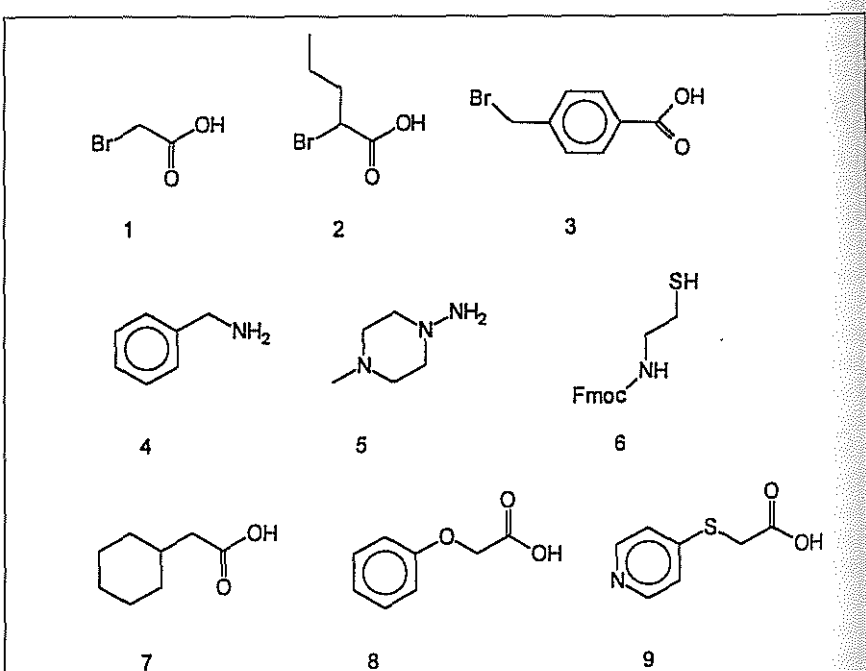


Figure 3. Building blocks used in the construction of nonpeptide library.

approach (Figure 1B) uses a statistical distribution of both structures on the polymeric bead. In this case, any possible ratio can be achieved, and the possibility of producing a cooperative effect of both sequences can be minimized. In the applications where the release of the screened peptide into solution is used [as described by us recently (17)], the localization of screening and coding compound on the bead is of no concern, since, due to the use of different linkers, the coding sequence is never released into the solution.

A simple scheme was followed to demonstrate the chemical synthesis of screening and coding sequences. A "screening" sequence was built from A, F and V. These amino acids were encoded by G, Y and L, respectively, in the "coding" sequence. The screening sequences were built in duplicate on both amino groups of lysine attached to another lysine side chain using Fmoc chemistry (see Figure 2). The resin was split into three parts and N^{α}-Fmoc protected amino acids were coupled to the screening branches, and the corresponding N^{α}-Boc (coding) amino acids were coupled to the remaining branch. All the resin was combined and redivided again into three parts. Deprotection of the N^{α}-Fmoc group and coupling of the subsequent N^{α}-Fmoc amino acid were performed in the presence of Boc protection on the other branch, stable under those conditions. In the next step, the N^{α}-Boc group was cleaved and the "complementary" N^{α}-Boc amino acids were coupled to the coding branch. The procedure of mixing, splitting and separate coupling of Fmoc and Boc amino acids was repeated once more. The synthesis was performed on a handle stable under conditions of both Boc and Fmoc strategies, which can, however, be cleaved under relatively mild acidolytic conditions after reduction of its sulfoxide moieties (21).

Sequencing of peptide-beads prepared in this manner demonstrated a molar ratio of 2:1 of screening to coding sequence and the appropriate correspondence of particular amino acids (Figure 2A). Using one aliquot of beads, the "screening" sequence was acetylated and a clean sequence reading was obtained from the "coding" sequence (Figure 2B). Using a different aliquot of beads, the "coding" sequence was blocked by a trifluoroacetyl group, sequencing of the "screening" branch was performed, the trifluoroacetyl group was cleaved from the sequenced beads and the sequence of the "coding" peptide was determined, confirming the results from the sequencing of the "screening" peptide (Figure 2C).

To verify that the synthetic strategy generates the predicted equimolar ratio of defined number of structures, the "mini-library" represented in an aliquot was cleaved from the support. Reversed-phase HPLC confirmed the presence of 27 different peptides. The peaks identified in the trace were collected and submitted for sequence analysis, which confirmed the purity of each peptide and its composition. Peptides were also cleaved from the single beads, and the feasibility of analysis of peptides released from only one bead was confirmed.

The second example demonstrates the construction of nonpeptide structures in parallel with the coding sequence. The building blocks chosen for this library are shown in Figure 3 and the structures of the two alternative libraries are shown in Figure 4. The difference between these libraries is the placement of the linker allowing for the selective cleavage of the product. In the first case (Figure 4A), the cleavage of the linker (SCAL – safety-catch amide linker [21]) results in the release of the nonpeptide screening structure X_3-X_2-X_1-Trp (Trp is attached for spectroscopic monitoring purposes) attached to its coding peptidic structure via a lysine moiety. Cleavage of the linker in the second case (Figure 4B) results in the release of the nonpeptide screening structure X_3-X_2-X_1-Trp without any attached coding peptide. Besides use in proving reliable coding, this chemistry permits one to determine if interaction with the acceptor is via the screening or coding sequence. Construction of the nonpeptide screening molecule

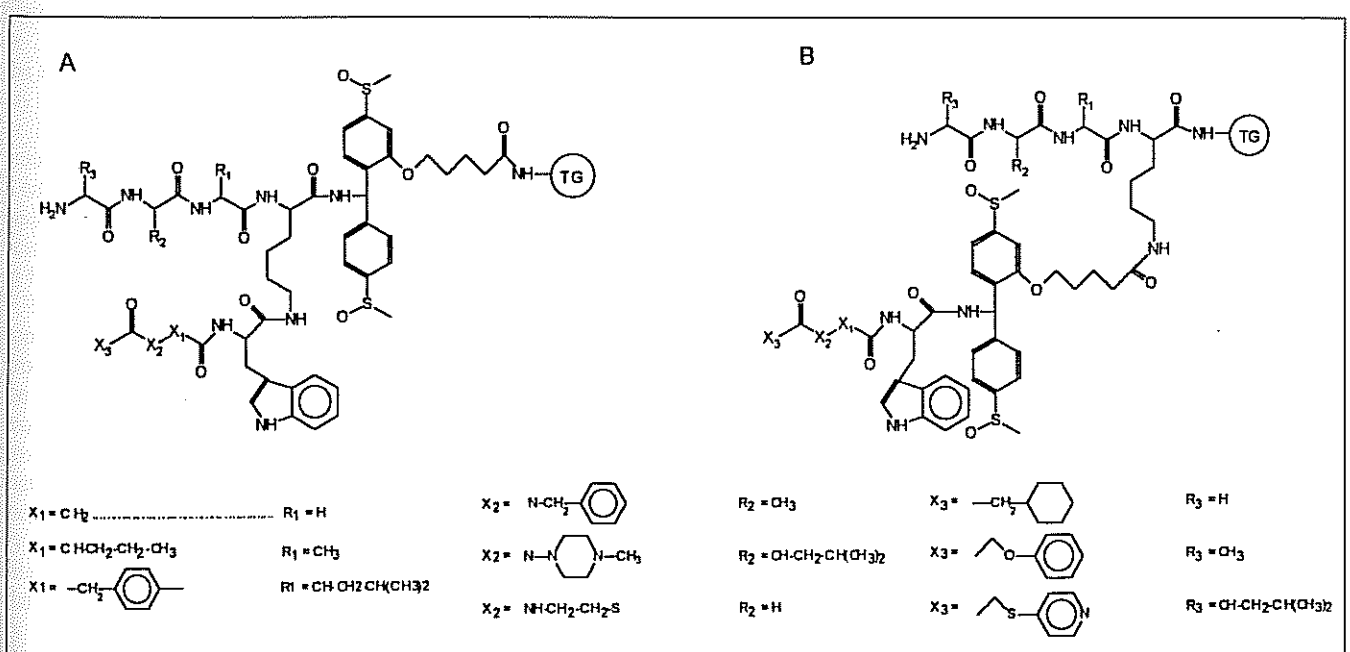


Figure 4. Alternative structures of nonpeptide libraries.

involved (i) attachment of a α-bromo-substituted carboxylic acid or bromomethylbenzoic acid to the available amino group on the solid carrier, (ii) alkylation of an amino (28) or thiol group of an amine or N-protected aminomercaptan, and (iii) acylation of a generated amino group by a derivative of carboxylic acid. We have selected the building blocks for this experiment in a manner that permits the assignment of the structure of the constructed screening molecules based solely on the molecular weight of the construct (see Table 1). Introduction of every unnatural building block to the screening structure was followed (or preceded) by the coupling of a coding amino acid to the other arm of the molecule. We have used only glycine, alanine and leucine for coding (these amino acids, therefore, encoded a different structural element in every step of the randomization). Assignment of these amino acids to a particular structural element is given in Figure 4. Alkylation of amines or thiol used in this experiment by 2-bromopentanoic acid attached to the polymeric matrix led to the generation of compounds with a chiral center, and therefore the number of structural combinations is not 27 but 36. However, only 27 different bead types are generated (with screening sequences of differing molecular weights), 9 of which contain a mixture of diastereoisomeric compounds. To simplify the analysis of the mixtures and to demonstrate the feasibility of this type of synthesis on a polymeric carrier, three of the possible structures were resynthesized as individual compounds (Figure 5), using the same chemistry as in the synthesis of the model library.

The generated mixtures were cleaved from the carrier, after the reduction of the SCAL linker, and analyzed by reversed-phase HPLC. The number of peaks obtained corresponds approximately to the predicted number of 36. Individual peaks from the first type of library (Figure 4A) were collected, part of the solution was subjected to Edman degradation and part was analyzed by mass spectroscopy. Results obtained confirm the correlation of sequence determination with molecular weight determination by mass spectroscopy, confirming the viability of the principle of coding by peptide sequence.

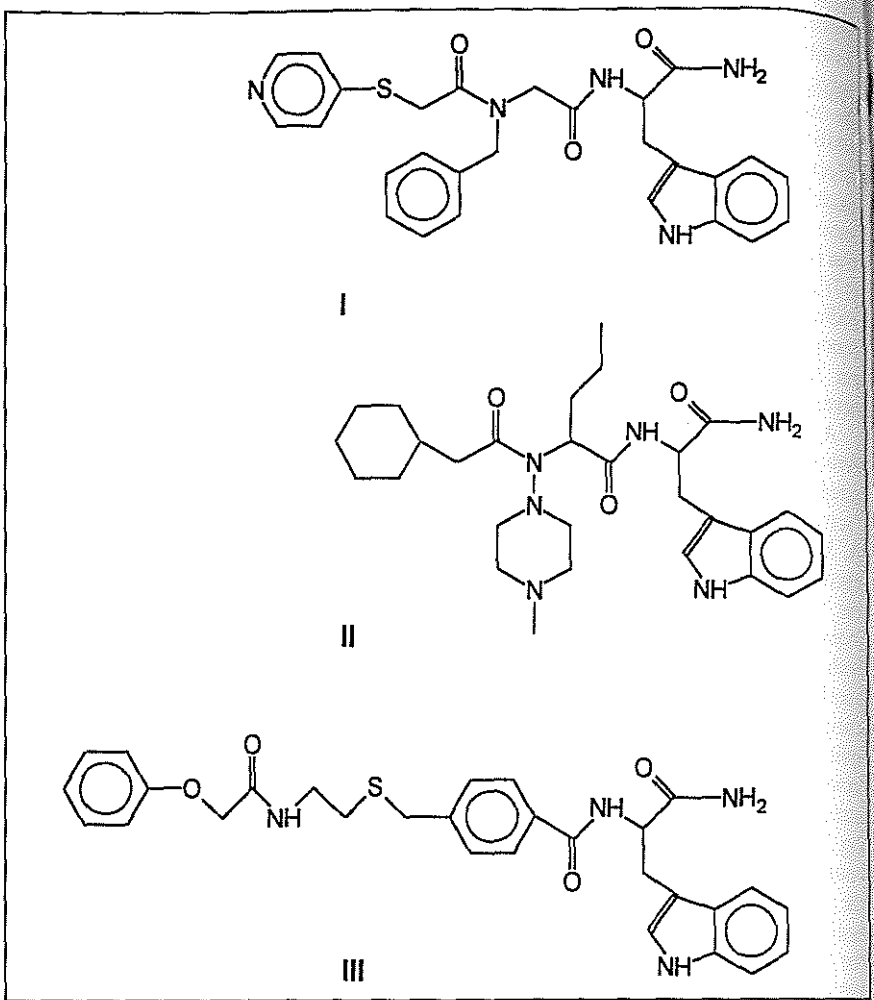
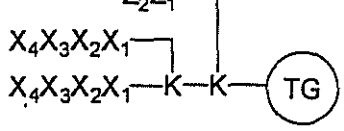


Figure 5. Structure of representative molecules from the nonpeptide library.



X = YGFLHPQRE

Z = AD IKMNSTV

Sequencing results:

1st cycle	2nd cycle	Deduced sequence	Target
Y, D	G, I	YGGF	anti-β-endorphin
Y, N	G, I	YGPF	anti-β-endorphin
Y, D	G, K	YGGL (3x)	anti-β-endorphin
H, S	P, I	HPQF (5x)	streptavidin

Figure 6. Structure of library with coding sequence and results of decoding of positively identified beads

An alternative analysis was performed on randomly selected beads from the second library (Figure 4B). Individual beads were treated with a reducing agent to labilize the SCAL linker, and the nonpeptide structure was cleaved by a TFA/water mixture. After this treatment, the beads were successfully sequenced (see Table 2) and the structure of the nonpeptide compound could be deduced. The cleaved compounds were analyzed by a micro-HPLC system.

The principle of encoding a "screening" structure by "coding" sequence can also be used for the determination of the structure of peptides containing a nonsequenceable component within the peptide chain. In this case it is necessary to code only for the nonsequenceable residue and any amino acid residues located on the carboxyl terminus of the molecule, after the nonsequenceable part. We have constructed a library mimicking this situation, although not containing a nonsequenceable component. The structure of the library is given in Figure 6. Amino acid residues X_4 and X_3 in the "screening" arm are not encoded by any counterpart in the "coding" arm. Amino acids Z_1 and Z_2 encode residues X_1 and X_2 and are present in one half the concentration versus amino acids in the "screening" sequence. Two cycles of Edman degradation can reveal the structure of the peptide of interest. The amino acid detected in the higher amount is the residue from position 1 or 2 of the "screening" sequence, and the amino acid detected in lower quantity is the residue encoding position 4 or 3 of this sequence. The coding amino acid may be the same as the one it encodes, or it can be another one. The coding and screening set of amino acids used in this example is given in Figure 6.

This library was synthesized using a combination of three amino-protecting groups. As a temporary protection of the α-amino group in the "screening" sequence, we have used the Fmoc group, cleavable by piperidine in dimethylformamide. Amino groups in the "coding" sequence were protected by the Ddz group (1), cleavable with diluted trifluoroacetic acid (2%). Side-chain functional groups were protected by *tert*-butyl type protecting groups, cleavable by trifluoroacetic acid of higher concentration (50%). One cycle of randomization with sequence tagging consisted of (i) distribution of the resin into a number of reaction vessels corresponding to the number of amino acids randomized in this step; (ii) coupling of Fmoc-protected amino acids (Y, G, F, L, H, P, Q, R, E); (iii) washing, cleavage of Ddz group and neutralization; (iv) coupling of the corresponding Ddz-protected amino acids (A, D, I, K, M, N, S, T, V); (v) mixing the solid support and deprotection of the Fmoc group.

This library was used in the screening against two model targets: anti-β-endorphin monoclonal antibody and streptavidin. Positive beads were identified by the standard staining technique (14,15). The beads (5 for each target) identified in this screen were subjected to two cycles of Edman degradation, the results of which are given in Figure 6. As can be seen, streptavidin-positive beads gave, in all cases, H and S (coding for Q) in the first cycle, and P and I (coding for F) in the second cycle. Thus, the sequence of the screening arm HPQF could be decoded without difficulties. Beads identified in the anti-β-endorphin screening gave more diverse results. Besides Y and D (coding for G), also N (coding for P) was found in the first cycle; and G and I (coding for F), and K (coding for L), were found in the second cycle. Therefore, sequences YGGL (3×), YGGF and YGPF could be constructed from these data. The sequences found are in agreement with the data obtained earlier (14–16).

SUMMARY

These experimental proofs have confirmed our thesis that the structure of non-Edman degradable compounds can be determined by a peptide-coding technique. This technique should enable the synthesis and screening for biological activity of a great diversity of nonpeptidic compounds. As a result, the technique broadens the potential application of combinatorial library approaches, such as the Selectide Process, in the search for new drug candidates.

ACKNOWLEDGMENTS

Authors are indebted to Dr. Farid Abdul-Latif for the peptide sequencing, Douglas Zoucha, Stephen Baum and Shelly Wade for technical assistance, and to Dr. Jutta Eichler for the valuable comments during the preparation of the manuscript. The work was supported in part by cooperative agreement UO1-CA57723-01 from the National Institutes of Health. KSL is the recipient of a Scholar's award from the Leukemia Society of America. FAB MS data were provided by Czechoslovak Peptide Service, Prague.

REFERENCES

1. **Birr, C., W. Lochinger, G. Stahnke and P. Lang.** 1972. The alpha, alpha dimethyl-3,5-dimethoxybenzyloxycarbonyl residue – an N-protecting group labile toward weak acids and irradiation. Liebig's Ann. Chem. *763*:162-173.
2. **Brenner, S. and R.A. Lerner.** 1992. Encoded combinatorial chemistry. Proc. Natl. Acad. Sci. USA *89*:5381-5383.
3. **Bunin, B.A. and J.A. Ellman.** 1992. A general and expedient method for the solid phase synthesis of 1,4-benzodiazepine derivatives. J. Am. Chem. Soc. *114*:10997-10998.
4. **Fodor, S.P.A., J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu and D. Solas.** 1991. Light-directed, spatially addressable parallel chemical synthesis. Science *251*:767-773.
5. **Furka, A., F. Sebestyen, M. Asgedom and G. Dibo.** 1991. General method for rapid synthesis of multicomponent peptide mixtures. Int. J. Pept. Protein Res. *37*:487-493.
6. **Geysen, M.H., H.M. Rodda and T.J. Mason.** 1986. A priori delineation of a peptide which mimics a discontinuous antigenic determinant. Mol. Immunol. *23*:709-715.
7. **Hirschmann, R., K.C. Nicolaou, S. Pietranico, J. Salvino, E.M. Leahy, P.A. Sprengler, G. Furst, A.B. Smith III, C.D. Strader, M.A. Cascieri, M.R. Candelore, C. Donaldson, W. Vale and L. Maechler.** 1992. Nonpeptidal peptidomimetics with β-D-glucose scaffolding. A partial somatostatin agonist bearing a close structural relationship to a potent, selective substance P antagonist. J. Am. Chem. Soc. *114*:9217-9218.
8. **Hirschmann, R., P.A. Sprengeler, T. Kawasaki, J.W. Leahy, W.C. Shakespeare and A.B. Smith III.** 1992. The first design and synthesis of a steroidal peptidomimetic. The potential value of peptidomimetics in elucidating the bioactive conformation of peptide ligands. J. Am. Chem. Soc. *114*:9699-9701.
9. **Houghten, R.A., 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 *357*:84-86.
10. **Jung, G. and A.G. Beck-Sickinger.** 1992. Multiple peptide synthesis methods and their applications. Angew. Chem. Int. Ed. Engl. *31*:367-383.
11. **Kerr, J.M., S.C. Banville and R.N. Zuckermann.** 1993. Encoded combinatorial peptide libraries containing non-natural amino acids. J. Am. Chem. Soc. *115*:2529-2531.
12. **King, D.S., C.G. Fields and G.B. Fields.** 1990. A cleavage method which minimizes

side reactions following Fmoc solid phase peptide synthesis. Int. J. Pept. Protein Res. *36*:255-266.

13. **Krchňák, V. and J. Vágner.** 1990. Color-monitored solid-phase multiple peptide synthesis under low-pressure continuous flow conditions. Pept. Res. *3*:182-193.
14. **Lam, K.S. and M. Lebl.** 1992. Streptavidin and avidin recognize peptide ligands with different motifs. Immunomethods *1*:11-15.
15. **Lam, K.S., 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.
16. **Lam, K.S., V.J. Hruby, M. Lebl, R.J. Knapp, W.M. Kazmierski, E.M. Hersh and S.E. Salmon.** 1993. The chemical synthesis of large random peptide libraries and their use for the discovery of ligands for macromolecular acceptors. Bioorg. Med. Chem. Lett. *3*:419-424.
17. **Lebl, M., M. Pátek, P. Kočiš, V. Krchňák, V.J. Hruby, S.E. Salmon and K.S. Lam.** 1993. Multiple release of equimolar amounts of peptides from a polymeric carrier using orthogonal linkage-cleavage chemistry. Int. J. Pept. Protein Res. *41*:201-203.
18. **Merrifield, R.B.** 1963. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. J. Am. Chem. Soc. *85*:2149-2154.
19. **Oldenburg, K.R., D. Loganathan, I.J. Goldstein, P.G.Schultz and M.A. Gallop.** 1992. Peptide ligands for a sugar-binding protein isolated from a random peptide library. Proc. Natl. Acad. Sci. USA *89*:5393-5397.
20. **Parmley, S.F. and G.P. Smith.**1988. Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. Gene *73*:305-318.
21. **Pátek, M. and M. Lebl.** 1991. Safety-catch anchoring linkage for synthesis of peptide amides by Boc/Fmoc strategy. Tetrahedron Lett. *32*:3891-3894.
22. **Pavia, M.R., T.K. Sawyer and W.H. Moos.** 1993. The generation of molecular diversity. Bioorg. Med. Chem. Lett. *3*:387-396.
23. **Pinilla, C., J.R. Appel, P. Blanc and R.A. Houghten.** 1992. Rapid identification of high affinity ligands using positional scanning synthetic peptide combinatorial libraries. BioTechniques 13:901-905.
24. **Scott, J.K.** 1992. Discovering peptide ligands using epitope libraries. TIBS *17*:241-245.
25. **Simon, R.J., R.S. Kaina, R.N. Zuckermann, V.D. Hueber, D.A. Jewell, S. Banville, N.G. Simon, L. Wang, S. Rosenberg, C.K. Marlowe, D.C. Spellmeyer, R. Tan, A.D. Frankel, D.V. Santi, F.E. Cohen and P.A. Bartlett.**1992. Peptoids: A modular approach to drug discovery. Proc. Natl. Acad. Sci. USA *89*:9367-9371.
26. **Smith III, A.B., T.P. Keenan, R.C. Holcomb, P.A. Sprengeler, M.C. Guzman, J.L. Wood, P.J. Carroll and R. Hirschmann.** 1992. Design, synthesis, and crystal structure of a pyrrolinone-based peptidomimetic possessing the conformation of a β-strand: Potential application to the design of novel inhibitors of proteolytic enzymes. J. Am. Chem. Soc. *114*: 10672-10674.
27. **Wang, K.Y., S. McCurdy, R.G. Shea, Regan, S. Swaminathan and P.H. Bolton.** 1993. A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. Biochemistry *32*:1899-1904.
28. **Zuckermann, R.N., J.M. Kerr, S.B.H. Kent and W.H. Moos.** 1992. Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis. J. Am. Chem. Soc. *114*:10646-10647.

Address correspondence to:

Michal Lebl
Selectide Corporation
1580 E. Hanley Boulevard
Tucson, AZ 85737