Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry

John Nielsen, Sydney Brenner,* and Kim D. Janda*,†

The Scripps Research Institute
Departments of Molecular Biology and Chemistry
10666 North Torrey Pines Road
La Jolla, California 92037

Received June 2, 1993

Screening of natural and synthetic compounds for activity as potential therapeutic agents has for decades been the main source of drug discoveries in the pharmaceutical industry. With advances in molecular biology, protein crystallography, and computational chemistry, "rational drug design" has found many advocates, but this approach is still difficult and slow. In recent years there has been a renaissance in drug screening with the advent of new technologies based on "combinatorial libraries".[1] These methods expose a large number of compounds to a target, and they allow compounds that bind to a target with the highest affinity to be filtered out from a pool of statistical sequences. Present technologies have some limitations; in chemically synthesized libraries, mostly of peptides to date, there can be intrinsic difficulties in the identification, selection, and enrichment of active compounds. In bacteriophage and nucleic acid libraries, the active compounds can be amplified by replication, but chemical diversity is inherently limited. Recently, a conceptual scheme for "encoded combinatorial chemistry" has been proposed as a possible solution to a number of these potential restrictions.[2,3] Such libraries would combine the potentially limitless diversity of synthetic libraries with the advantages of natural libraries based on gene technologies. *It was proposed to link each molecule of a chemically synthesized entity to a particular oligonucleotide sequence constructed in parallel and to use this encoding genetic tag to identify and enrich active compounds.* In this paper, we detail the chemistry necessary to synthesize encoded libraries, choosing peptides as an example.

In essence, such libraries require alternating stepwise syntheses of a peptide and an oligonucleotide sequence on a common linker (Scheme I). Thus a linker had to be constructed and attached to a suitable solid support capable of housing both the oligonucleotides and the peptides synthesized. Such a spacer had to be compatible with both solid-phase peptide synthesis (SPPS) using Fmoc chemistry and oligonucleotide synthesis with phosphoramidites, and it needs to be functionalized in such a way that it is possible to deprotect and detach the oligonucleotide-tagged peptides in a regioselective, nondestructive manner. *This latter process allows the encoding of molecules rather than their carriers.* Furthermore, we wanted the linker unit to be versatile enough to allow a "dendritic" display of the chemical library for controlled multivalent ligand display.[4]

Controlled-pore glass (CPG) was used as the solid matrix, because this material has a long-standing reputation for efficient oligonucleotide synthesis and has also proved efficient in SPPS. Scheme II delineates the synthesis of four new supports, CPG-**3**, -**4**, -**5**, and -**6**. There are several features to be noted: (A) *N*-Fmoc- and *O*-DMT-protected L-serine/lysine-branched monomers serve as differential orthogonal attachment points for the oligonucleotides and the peptides (singular and bivalent). (B) The high optically purity of these CPG supports is necessary since a stereochemically impure linker complicates testing, analysis, and interpretation of results from such biopolymers. (C) Employment of the succinyl aminohexanol–sarcosine appendage provides a method for nonacidic detachment of the oligonucleotide-tagged peptides and serves to stabilize the succinyl ester from cleavage by nonaqueous bases (i.e., piperidine and 1,8-diazabicyclo[5.4.0]-undec-7-ene (DBU)).[5] (D) Enlistment of variable-length tethers between the encoding genetic tag and the chemical units, combined with the overall general elaboration of the CPG support, provides a variety of "steric space", a functionality generally agreed to be an important factor in efficient oligonucleotide and peptide synthesis.

Initially we probed the coupling efficiencies and general overall capabilities of our new supports, using the simpler of the four matrices, CPG-**3**. Its ability to support oligonucleotide synthesis was tested by the preparation of a number of oligodeoxyribonucleotides using standard 2-cyanoethyl-protected nucleoside phosphoramidite chemistry. The sequences ranged from 24- to 45-mers, and repetitive yields were typically upwards of 98%, as determined by dimethoxytrityl cation release. The ability of CPG-**3** to serve as a support for SPPS was investigated by the synthesis of the test peptide CPG-**3**-Ser'(aho)-Val-Phe-Gln-Pro-His-H [Ser'(aho) indicates the CPG-linked serine]. The peptide was released from the support and analyzed by reversed-phase HPLC, and its identity was confirmed by ion-spray mass spectrometry (data not shown).

Several additional model studies were undertaken to prepare for the construction of oligonucleotide-tagged peptide libraries on our CPG supports by alternating parallel synthesis. Appropriate Fmoc-amino acid derivatives were subjected to the reagents used in oligonucleotide synthesis,[6] and all were found to be stable. A study on the stability of the required protecting groups for the exocyclic amino moieties contained within the nucleobases (adenine, cytosine, and guanine) was also performed. The conditions used for the Fmoc deprotection do not significantly affect the standard exocyclic benzoyl and isobutyryl protecting groups, eliminating the possibility that these nucleobases could generate erroneous branch points. With regard to phosphate protection, we chose to use the methyl phosphate moiety as it has robust stability toward our Fmoc deprotection reagent.[7]

To investigate the biochemical properties of these oligonucleotide-tagged peptides, the syntheses of several peptide ligand sequences containing the β-endorphin epitope were performed on CPG-**3** and -**4**, Chart I.[8] The sequences of both the monomeric and the dendritic peptides were confirmed by Edman degradation. All nucleotide sequences were shown to be efficiently radiolabeled with T4-polynucleotide kinase, and the nucleotide tags could be amplified using standard PCR techniques.[9] Binding assays were initiated using both the mono- and the bivalent ligands. Our preliminary findings show that antibody 3-E7[10] binds only the relevant leucine enkephalin-encoded members (**7–10**; K_d of **7** and **8** = 24 nM). *Notably, the potentially sterically impeding oligonucleotide tag does not appear to affect peptide binding.*

* To whom reprint requests and correspondence should be addressed.
† A. P. Sloan Fellow, 1993–1995.

(1) For some excellent reviews, see: (a) Jung, G.; Beck-Sickinger, A. G. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 367–383. (b) Pavia, M. R.; Sawyer, T. K.; Moos, W. H. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 387–396.

(2) (a) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381–5383. (b) Amato, I. *Science* **1992**, *257*, 330–331.

(3) Recently, two papers describing the encoding of combinatorial libraries have appeared: (a) Kerr, J. M.; Banville, S. C.; Zuckermann, R. N. *J. Am. Chem. Soc.* **1993**, *115*, 2529–2531. (b) Nikolaiev, V.; Stierandova, A.; Krchnak, V.; Seligmann, B.; Lam, K. S.; Salmon, S. E.; Lebl, M. *Peptide Res.* **1993**, *6*, 161–170. However, these methods use peptides as the coding elements, and enrichment by amplification is not possible.

(4) For a discussion of multivalent peptide display, see: (a) Cull, M. G.; Miller, J. F.; Schatz, P. J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1865–1869. (b) Lowman, H. B.; Bass, S. H.; Simpson, N.; Wells, J. A. *Biochemistry* **1992**, *30*, 10832–10838.

(5) Brown, T.; Pritchard, C. E.; Turner, G.; Salisbury, S. A. *J. Chem. Soc., Chem. Commun.* **1989**, 891–893.

(6) 3% TCA/CH_2Cl_2 and I_2/H_2O/pyridine/THF.

(7) Palom, Y.; Alazzouzi, E.; Gordillo, F.; Grandas, A.; Pedroso, E. *Tetrahedron Lett.* **1993**, *34*, 2195–2198.

(8) Syntheses here were performed in a nonalternating manner.

(9) PCR of the oligonucleotide-tagged β-endorphin peptides were successfully accomplished either in solution or when bound to 3-E7 coated DynaBeads.

(10) Meo, T.; Gramsch, C.; Inan, R.; Höllt, V.; Weber, E.; Herz, A.; Riethmüller, G. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4084–4088.

0002-7863/93/1515-9812$04.00/0

Scheme I. General Methodology for Obtaining Encoded Combinatorial Chemical Libraries

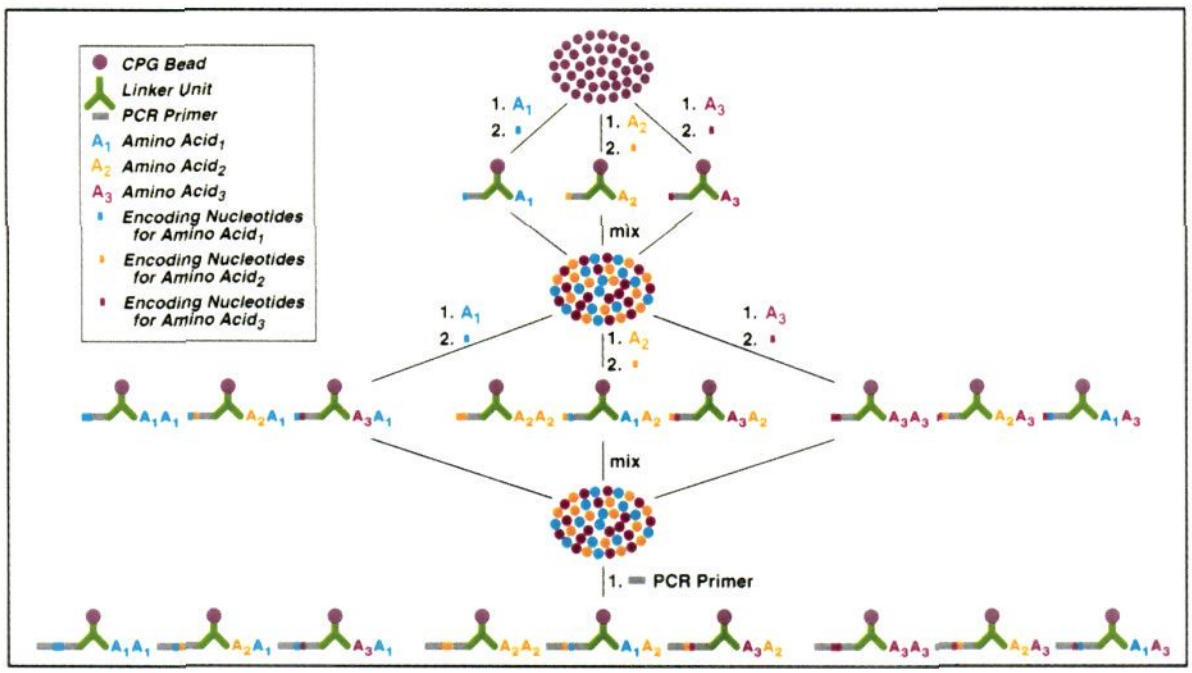
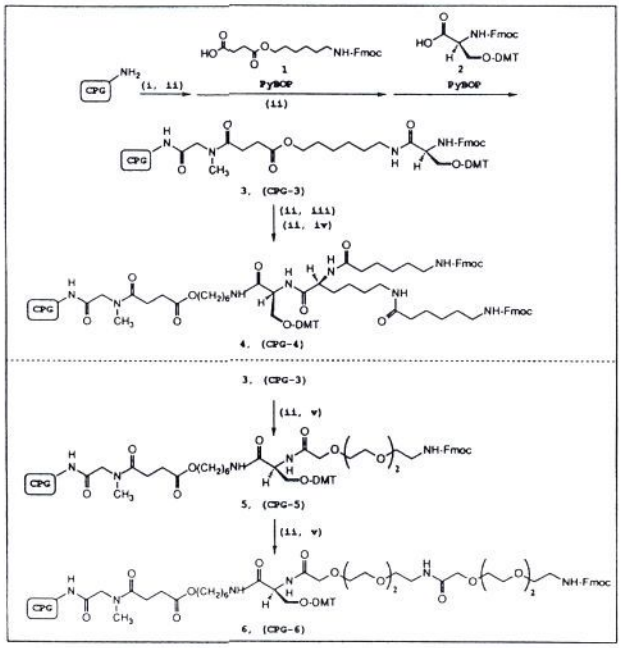


Chart I. Sequences (One-Letter Codes) of Compounds Used in Binding Study[a]

7	5′-AGCTACTTCCCAAGGGAGCTGCTGCTAGTCGGGCCCTATTCTTAG-3′-*S*-[**LFGGY**-H]
8	5′-AGCTACTTCCCAAGGGAGCTGCTGCTAGTCGGGCCCTATTCTTAG-3′-*EtG*-*S*-[**LFGGY**-H]
9	5′-AGCTACTTCCCAAGGGAGCTGCTGCTAGTCGGGCCCTATTCTTAG-3′-*S*-[***K***(***ε*ahx**-**LFGGY**-H)$_2$]
10	5′-AGCTACTTCCCAAGGGAGCTGCTGCTAGTCGGGCCCTATTCTTAG-3′-*EtG*-*S*-[***K***(***ε*ahx**-**LFGGY**-H)$_2$]
11	5′-AGCTACTTCCCAAGGATCACCACACTAGCGGGGCCCTATTCTTAG-3′-*S*-[**VFQPH**-H]
12	5′-AGCTACTTCCCAAGGATCACCACACTAGCGGGGCCCTATTCTTAG-3′

[a] ***ε*ahx** is 6-aminohexanoic acid, *EtG* is a hexaethylene glycol spacer, and *S* represents the detached serine–aminohexanol appendage.

Scheme II[a]



[a] Reagents: (i) Fmoc-sarcosine, PyBOP (benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate), DIPEA; (ii) piperidine/DMF (2/8, v/v); (iii) Fmoc-Lys(Fmoc)-OH, PyBOP, DIPEA; (iv) Fmoc-ε-ahx-OH (Fmoc-6-aminohexanoic acid), PyBOP, DIPEA; (v) 11-Fmoc-amido-3,6,9-trioxaundecanoic acid, PyBOP, DIPEA.

Strikingly, ligand affinity is dependent on the valency of the encoded peptide. Thus the bivalent units **9** and **10** have a greater affinity for antibody 3-E7 (K_d of **9** and **10** = 7 nM) than their singular counterparts **7** and **8**.

Based on the above findings, we proceeded to make encoded peptide members on CPG-**3** in the required alternating pattern depicted in Scheme I. We discovered that when CPG-**3** was functionalized with the required primer sequence 1, deprotection of the Fmoc was slowed significantly, and, most importantly, amino acid addition was not efficient (<10%). Suspecting steric hindrance, we tested primer 1-functionalized CPG-**5** and -**6**. CPG-**5** provided a significant improvement (35–60%), while CPG-**6** gave virtually quantitative couplings. *Hence, CPG-**6** provides the elements necessary to efficiently make genetically encoded peptide libraries.*

In conclusion, we have designed and synthesized a set of four chemically pure, base-labile, orthogonally protected branched matrices, of which CPG-**6** is capable of supporting alternating oligonucleotide–peptide synthesis in the required manner. This novel support has led to the first successful alternating bidirectional synthesis of encoded peptide entities. The simplicity involved in creating this matrix for encoded libraries, coupled with our controlled "dendritic" methodology, should lend itself to the rapid expansion of the technology. Undertakings to create tagged libraries based on natural and nonnatural building blocks are currently under investigation.

Acknowledgment. This work was supported in part through a fellowship from the A. P. Sloan Foundation (K.D.J.) and by a Markey Foundation grant (S.B.).

Supplementary Material Available: Details of the preparation and characterization of encoded chemical members (7 pages). Ordering information is given on any current masthead page.