J|A|C|S
COMMUNICATIONS

Published on Web 06/06/2002

A Novel Peptide-Based Encoding System for "One-Bead One-Compound" Peptidomimetic and Small Molecule Combinatorial Libraries

Ruiwu Liu, Jan Marik, and Kit S. Lam*

Division of Hematology & Oncology, Department of Internal Medicine, UC Davis Cancer Center, University of California, Davis, 4501 X Street, Sacramento, California 95817

Received April 4, 2002

The "one-bead one-compound" (OBOC) combinatorial library method has been successfully applied to the identification of ligands for a large number of biological targets.[1] The OBOC method is highly efficient, especially when used with well-established on-bead binding or functional assays. Literally, millions of compounds can be screened concurrently within 1 to 2 days. The positive beads are then isolated for structural determination. For peptides, the beads can be routinely microsequenced with Edman chemistry, using an automatic protein sequencer. However, structure determination of peptidomimetic and small molecule compounds on a single bead is not trivial. Various encoding methods have been reported by several groups and the subject has been recently reviewed.[1,2] A successful encoding tag must survive organic synthesis conditions, not interfere with screening assays, and be readily decoded with no ambiguity. Preferably, the encoding process is highly sensitive and reliable, and the tag can encode a large number of compounds, e.g. 100 000. Furthermore, to fully exploit the OBOC combinatorial method, the testing compound and the encoding tag must be able to be "packed" into a very small volume such as a 90 μm diameter bead.

Current chemical encoding methods have two major limitations. First, the coding structure may interfere with the binding of the target protein to the testing compound. Second, the current chemical encoding method requires that the synthetic chemistry of the coding tag and the testing compound be orthogonal, which doubles the number of required synthetic steps. To address the first problem, we have developed a new and simple method that enables us to topologically segregate the testing compound from the coding molecule. In this method, the testing compounds are present on the exterior of the bead, and the coding molecules reside in the interior of the bead. Thus, only the testing compounds are exposed to the target proteins. To address the second problem, we have developed a novel strategy by which we can combine the synthetic steps of both coding molecules and testing compounds, thus eliminating the extra synthetic steps and minimizing the amount of undesirable side products. We report here a novel, highly efficient and robust peptide-based encoding system for OBOC peptidomimetic and small molecule combinatorial libraries with testing molecules on the outer layer and the coding tags in the interior of the bead.

We have previously reported the use of polyglutamic acid[3] or proteases,[3,4] to modify the bead surface of the TentaGel bead (Rapp Polymere) so that topologically segregated bifunctional beads can be generated. These methods are successful to a certain degree but far from ideal because they are technically difficult to control and not reproducible from batch to batch. Recently, we have developed a much simpler, inexpensive, and highly reproducible method to topologically bifunctionalize resin beads. In this physical approach, we have created a biphasic solvent environment, such that the surface of the TentaGel bead is exposed to organic solvent that contains the derivatizing reagent (e.g. Fmoc-OSu), while the interior of the bead remains in water without any derivatizing reagent. The end result is that only the outer layer of the bead is derivatized. By varying the ratio of diethyl ether and dichloromethane in the organic phase, and adjusting the amount of Fmoc-OSu used, the thickness of the outer layer can be controlled.

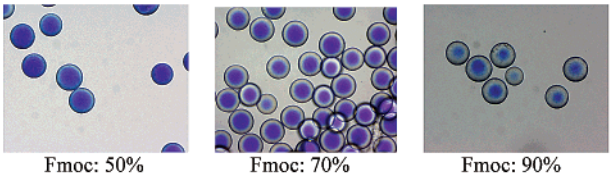


Figure 1. Photomicrograph of topologically segregated bifunctional TentaGel beads (with N-Fmoc on the outer layer and free amino group in the interior) that have been derivatized with bromophenol blue. The blue core reflects the location of the free amino groups prior to reacting with bromophenol blue, and the outer colorless shell reflects the location of the Fmoc-protecting group. Percent Fmoc substitution is determined by UV spectrophotometric analysis of released dibenzofulvene−piperidine adduct.

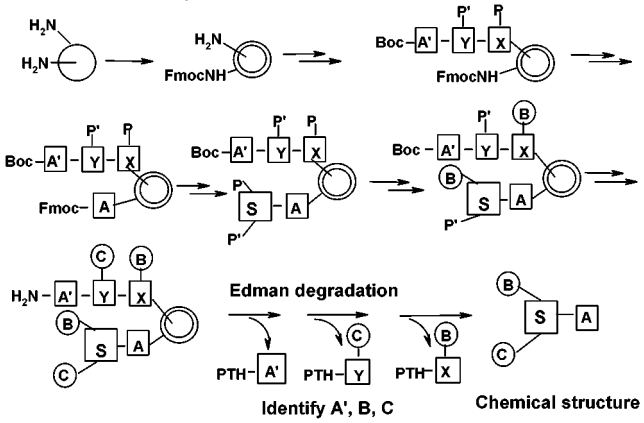
The location of the free amino groups on the topologically segregated bifunctional beads can be visualized microscopically by derivatizing these functional groups with bromophenol blue (3′,3″,5′,5″-tetrabromophenol sulfone-phthalein).[5] The photomicrographs in Figure 1 clearly demonstrate the varying thickness of the colorless outer layer of the beads that corresponds to the amount of Fmoc-protected amino groups. We have prepared a batch of bifunctional beads with 60% Fmoc substitution as the outer layer and have used three different methods to quantify the amount of Fmoc coupled to the bead. We have used a UV spectrophotometer to quantitate the amount of released dibenzofulvene−piperidine adduct after treatment with piperidine;[6] it is determined to be 59%. We have derivatized a portion of the same batch of beads with Leu on the outer layer and Ile in the interior. Sequence analysis with Edman degradation using a protein microsequencer (ABI Procise 494) reveals a Leu:Ile ratio of 57:43. Quantitative amino acid analysis of the same batch of beads yields a Leu:Ile ratio of 64:36.

To efficiently generate an encoded chemical library, the appropriate orthogonal protecting groups and synthetic schemes must be chosen so that the number of necessary chemical reactions can be minimized, e.g. by coupling the building blocks to the testing and coding arms simultaneously. In our encoding system, the testing compound is a peptidomimetic or small molecule, and the coding molecule is a peptide containing a large number of unnatural α-amino acids derived from different building blocks that are used for generating the peptidomimetics and small molecules. The

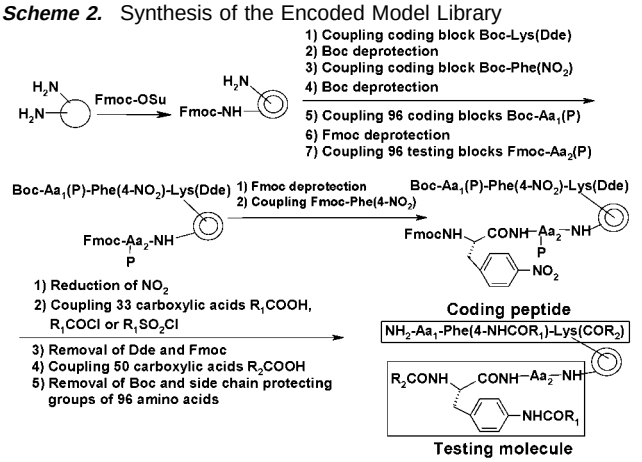
* To whom correspondence should be addressed. E-mail: kit.lam@ucdmc.ucdavis.edu.

10.1021/ja026421t CCC: $22.00

Scheme 1. General Synthetic and Encoding/Decoding Scheme of the Encoded Library



Scheme 2. Synthesis of the Encoded Model Library



retention time of these unnatural α-amino acids must be predetermined by a protein microsequencer.

Scheme 1 illustrates the general synthetic and decoding scheme of our encoded library. The bead is first topologically derivatized with Fmoc on the outer layer (it can also be other protecting groups, e.g. Alloc) and a free amino group in the interior. A coding arm precursor, **Boc-A′-(P′)Y-(P)X**, is then built in the interior of the bead. **(P)X** and **(P′)Y** represent trifunctional amino acids, which are preprotected by orthogonal groups, **P** and **P′**, respectively. **A** represents a bifunctional building block that links the scaffold to the bead, and it can be coded by amino acid **A′**. **S** represents the scaffold of the testing arm. Beads with the coding precursor (**Boc-A′-(P′)Y-(P)X-Bead-Fmoc**) can be bulk prepared, and be ready for future use in the preparation of any peptidomimetic and small molecule libraries that are compatible with the chemistry and encoding scheme of the coding arm.

By combining two α-amino acids to code for one building block ***A*** (Scheme 1), the number of building blocks one can choose for the library synthesis can be greatly increased. For example, if each building block is coded by a combination of 2 of 22 selected amino acids, a total number of 231 ([22 × 21]/2 = 231) structurally different building blocks can be encoded. The 22 selected amino acids are composed of the 19 natural amino acids except cysteine plus three unnatural amino acids: norvaline (Nva), norleucine (Nle), and cyclohexylglycine (Chg). In practice, we can also include 22 single-coding subunits to encode 22 additional building blocks. Therefore, the total number of building blocks that can be encoded by the above scheme is 253, which should meet the requirement of most peptidomimetic and small molecule libraries. However, it is extremely important that a signal from each of the doublet amino acids be detectable, preferably in "equi-peak height" ratio. This can be achieved by adjusting the ratio of the two encoding Boc-amino acids according to their relative reactivities. Briefly, equal molar coding Boc-amino acids (3 equiv in total to resin) can be premixed before coupling to the resin. Approximately 20−30 amino acid beads are inserted into the cartridge for microsequencing, and one cycle of Edman degradation is performed. The relative reactivity of the 22 Boc-protected amino acids is obtained by comparing the peak height of each amino acid with the peak height of the selected amino acid, Boc-His(Trityl)-OH.

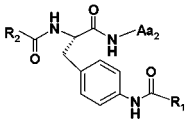
During the peptidomimetic and small molecule library synthesis, the corresponding protecting groups on both the coding arm and testing arm are deprotected prior to the coupling of each building block (**B** or **C**), so that the building blocks can be coupled to the testing arm and coding arm simultaneously. Since the protecting group on the scaffold is identical or of the same type (e.g. Fmoc and Dde both can be removed by 2% hydrazine) as the one on the coding arm, each step of the peptidomimetic or small molecule synthesis can proceed as if there is no coding arm. Therefore, no extra deprotection steps will be needed. As shown in Scheme 1, the decoding process is straightforward since the testing molecule is stable to the Edman condition, and the coding molecule consists of peptides with α-amino acids that can be readily sequenced by an automatic protein microsequencer. Alternatively, the coding tag can be released (if a cleavable linker is used) and decoded unambiguously by mass spectrometry, if each of the coding subunit(s) has a unique molecular mass.

To demonstrate the efficiency and simplicity of our encoding system, a simple, encoded peptidomimetic model library has been generated (see Scheme 2). We have predetermined the elution profiles of many PTH derivatives of lysine and Phe(4-NH_2) using an improved sequencing gradient program.[7] The side chain amino groups of these α-amino acids are acylated by a large number of carboxylic acids, acyl chlorides, or sulfonyl chlorides. With our previous experience,[7] it is not difficult distinguishing two amino acids with a retention time difference greater than 0.10 min. Therefore, we can easily choose the appropriate acylating reagents as building blocks for both the testing and coding arms. The model library has been synthesized with Phe(4-NH_2) as a trifunctional scaffold for the testing compound and also as a peptide backbone for the coding tag. We have selected its precursor Phe(4-NO_2) for the attachment of the building blocks because the NO_2 group can be treated like a protected amino group, and it can be easily converted to amino groups by reduction with $SnCl_2$.

In our encoding strategy, the reduction of the NO_2 group, the removal of the protecting groups Fmoc and Dde of NH_2 (with 2% hydrazine in DMF), and the acylation of both coding and testing arms are achieved simultaneously. This greatly reduces the number of synthetic steps. For decoding, positive beads are isolated and subjected to automatic microsequencing. Since the testing molecule is unsequenceable, only the coding signal is observed. The signal of the amino acid combination (**Aa_1**) in the first cycle corresponds to **Aa_2** of the testing compound. In the second cycle, the signal of Phe(4-NHCO**R_1**) corresponds to the aniline acylation derivative (**R_1**). The signal of the lysine derivative in the third cycle corresponds to the last coupling moiety (**R_2**). By combining all the decoded structural information, the complete chemical structure of the testing molecule is easily determined.

We use an enzyme-linked colorimetric assay[8] to screen this 158 400-member library (96 × 50 × 33) for ligands that bind to streptavidin at an extremely dilute streptavidin−alkaline phosphatase conjugate concentration (1:100 000). Ten of the darkest beads are picked and sequenced. Novel ligands with clear motifs have been identified (see Table 1). All ligands can be catalogued

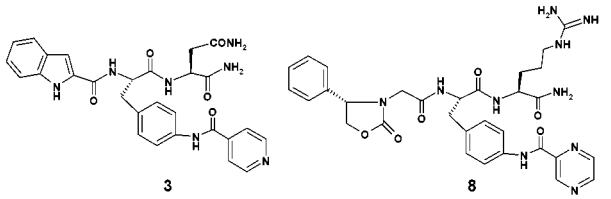
Table 1. Streptavidin Binding Assay Results[a]



Aa_2: 96 amino acids including 20 L-natural amino acids and 19 D-isomers, 15 beta-amino acids and 42 other amino acids
R_1COOH: 33 carboxylic acids, acyl chlorides or sulfonyl chlorides
R_2COOH: 50 carboxylic acids

entry	Aa_2	R_1COOH	R_2COOH
1	Gly	isonicotinic acid	indole-2-carboxylic acid
2	Gly	isonicotinic acid	indole-2-carboxylic acid
3	Asn	isonicotinic acid	indole-2-carboxylic acid
4	Asn	isonicotinic acid	indole-2-carboxylic acid
5	Ser	isonicotinic acid	indole-2-carboxylic acid
6	Leu	isonicotinic acid	indole-2-carboxylic acid
7	D-Arg	2-pyrazine carboxylic acid	(S)-(+)-2-oxo-4-phenyl-3-oxazolidine acetic acid
8	Arg	2-pyrazine carboxylic acid	(S)-(+)-2-oxo-4-phenyl-3-oxazolidine acetic acid
9	Tyr	2-pyrazine carboxylic acid	(S)-(+)-2-oxo-4-phenyl-3-oxazolidine acetic acid
10	Cptd	2-pyrazine carboxylic acid	(S)-(+)-2-oxo-4-phenyl-3-oxazolidine acetic acid

[a] Cptd = 3-carboxymethyl-1-phenyl-1,3,8-triazaspiro[4,5]decan-4-one.



3 8

Figure 2. Chemical structure of compounds **3** and **8**.

into two groups with indole-2-carboxylic acid or (*S*)-(+)-2-oxo-4-phenyl-3-oxazolidine-acetic acid as the acylating reagents in the last step ($\mathbf{R_2}$). All the ligands have an aromatic *N*-heterocyclic structure ($\mathbf{R_1}$) attached to the amino group of aniline. In contrast, there is significant variation in the first building units ($\mathbf{Aa_2}$). These ligands are then re-synthesized on beads and re-tested for their binding to streptavidin, using the enzyme-linked colorimetric assay. The beads containing compounds **3** and **8** (Figure 2) stain much darker (higher binding affinity) than the other beads. This result validates the encoding system for the OBOC peptidomimetic and small molecule combinatorial libraries. It is rapid, reliable, and direct.

In principle, many commercially available trifunctional amino acids with orthogonal protecting groups such as lysine, 4-aminophenylalanine, 4-bromophenylalanine, glutamic acid, serine, tyrosine, cysteine, 1-amino-4-ketocyclohexylcarboxylic acid, and 2-amino-2-propynylacetic acid can be used as encoding units. The side chains of these α-amino acids are sites to which building blocks such as carboxylic acids, acyl halides, sulfonyl halides, alkyl halides, amines, alcohols, isocyanates/isothiocyanates, aldehydes (aliphatic, aromatic, heterocyclic), or aryl boronic acids can be attached via a large number of organic reactions, including C−C bond-forming reactions (e.g. Suzuki coupling). Other unnatural α-amino acids such as 2-aminopent-4-enoic acid, 2-amino-3-acryloyloxypropinonic acid, and 2-amino-3-(4-formyl-benzoylamino)propionic acid can also be prepared and used as the site of attachment of building blocks via 1, 3-dipolar cycloaddition reaction, Michael addition, and Aldol condensation, respectively.

Sample processing for decoding is easy and direct as there is no need for cleavage and retrieval of the coding tag, and the positive beads can be sequenced directly with Edman degradation. The encoding method takes full advantage of (1) the chemical stability of peptide tag under many organic reaction conditions, (2) the ease of simultaneous coupling of common building blocks to the scaffold of the testing compounds and to the side chains of the α-amino acids on the coding peptide, and (3) straightforward and unambiguous decoding with automatic microsequencing or mass spectrometry. Besides eliminating a number of synthetic steps by combining the synthesis on the testing and encoding arms in many of the coupling steps, this method also enables the coding arm precursor to be prepared in bulk prior to the synthesis of the libraries. Consequently, encoded peptidomimetic and small molecule libraries may be prepared as if there is no coding arm. The biphasic approach to derivatize a resin bead topologically and bifunctionally is highly efficient, inexpensive, and easy to control. Limitations of this encoding method are the following: (1) synthetic conditions of the small molecule, peptide tag, and scaffolding have to be compatible with each other and (2) building blocks need to be carefully chosen so that the retention times of their amino acid derivatives do not overlap. Work is currently underway in our laboratory to decode this library with mass spectrometry from a coding peptide extracted from one single bead. With use of this latter approach, hundreds of beads, instead of only eight, can be decoded in 1 day, and the selection of building blocks is not limited by overlapping retention times of some of the PTH-amino acid derivatives.

Acknowledgment. This work was supported by NIH R33CA-86364. Ruiwu Liu is supported in part by the University of California Systemwide Biotechnology Research Program, grant No. 2001-07.

Supporting Information Available: Experimental procedures for bead derivatization, library synthesis, and screening; retention time of 50 PTH-derivatives of lysine and 33 PTH-derivatives of Phe(4-NH_2); and relative activities of Boc-amino acids (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lam, K. S.; Krchnak. V.; Lebl. M. *Chem. Rev.* **1997**, *97*, 411−448.
(2) (a) Czarnik, A. W. *Curr. Opin. Chem. Biol.* **1997**, *1*, 60−66. (b) Xiao, X. Y. *Front. Biotechnol. Pharm.* **2000**, *1*, 114−149. (c) Barnes, C.; Balasubramanian, S. *Curr. Opin. Chem. Biol.* **2000**, *4*, 346−350. (d) Affleck, R. L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 257−263.
(3) Lebl, M.; Lam, K. S.; Salmon, S. E.; Krchnak, V.; Sepetov, N.; Kocis, P. U.S. Patent 5840485, 1998.
(4) Vagner, J.; Barany, G.; Lam, K. S.; Krchnak, V.; Sepetov, N. F.; Ostrem, J. A.; Strop, P.; Lebl, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8194−8199.
(5) Krchnak, V.; Vagner, J.; Safar, P.; Lebl, M. *Collect. Czech. Chem. Commun.* **1998**, *53*, 2542−2548.
(6) Bennett, W. D.; Christensen, J. W.; Hamaker, L. K.; Peterson, M. L.; Rhodes, M. R.; Saneii, H. H., Eds. In *Advanced ChemTech Handbook of combinatorial and solid-phase organic chemistry? A guide to principles, products and protocols*; Advanced ChemTech Inc.: Louisville, KY, 1998; p 330.
(7) (a) Liu, R.; Lam, K. S. Automatic Edman microsequencing of peptides containing multiple unnatural amino acids. *Anal. Biochem.* **2001**, *295*, 9−16. (b) Liu, R.; Lam, K. S. In *Peptides: The wave of the future. Proceedings of the Second International and the Seventeen American Peptide Symposium;* Lebl, M., Houghten, R. A., Eds.; American Peptide Society and Kluwer Academic Publishers: San Diego, CA, 2001; pp 299−300.
(8) (a) Liu, G.; Lam, K. S. In *Combinatorial Chemistry, A Practical Approach Series*; Fenniri, H., Ed.; Oxford University Press: New York, 2000; pp 33−49. (b) Lam, K. S.; Lebl, M. *ImmunoMethods* **1992**, *1*, 11−15.

JA026421T