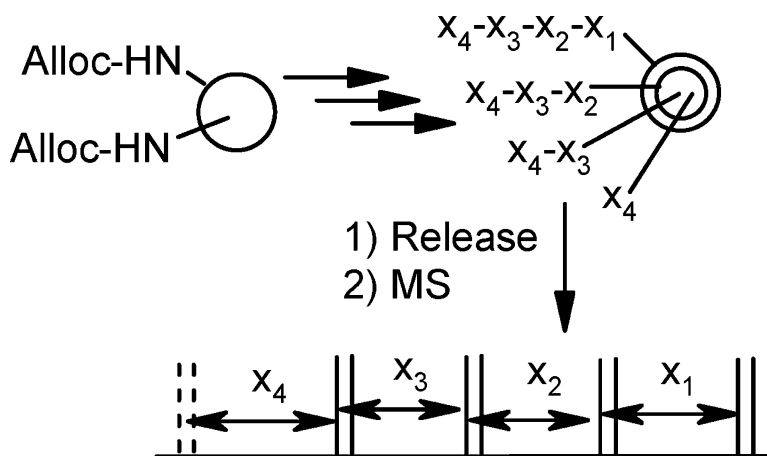Article

Partial Alloc-Deprotection Approach for Ladder Synthesis of "One-Bead One-Compound" Combinatorial Libraries

Xiaobing Wang, Li Peng, Ruiwu Liu, Sukhjinder S. Gill, and Kit S. Lam

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 8 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Partial Alloc-Deprotection Approach for Ladder Synthesis of "One-Bead One-Compound" Combinatorial Libraries

Xiaobing Wang, Li Peng, Ruiwu Liu, Sukhjinder S. Gill, and Kit S. Lam*

Division of Hematology and Oncology, Department of Internal Medicine, UC Davis Medical Center, 4501 X Street, Sacramento, California 95817

The topologically segregated bilayer-bead concept has been applied to encoded "one-bead one-compound" (OBOC) combinatorial libraries to avoid the interference of coding tags with biological screening. In this paper, we report on the development of a novel partial Alloc-deprotection (PAD) approach and the use of this approach to establish a new ladder-synthesis method for OBOC combinatorial libraries to further exploit the concept. In the PAD approach, Alloc-protected beads are partially deprotected, sequentially layer by layer, starting from the outer layer toward the bead interior. The degree of deprotection (or thickness of each layer) is controlled by the time of exposure to the deprotecting agent, palladium. By repetitive use of the PAD approach, a small portion of Alloc-protected N termini in the bead interior is liberated in each synthetic cycle for generation of an additional ladder member such that each library bead will carry a full-length library compound on the bead surface and a series of truncated ladder members in the bead interior. For the libraries containing isobaric residues, a simple encoding strategy is introduced in the ladder-synthesis method so that the isobaric residues can be differentiated by the coding tags. One advantage of this encoding strategy is that the coding tags are confined together with the truncated ladder members in the bead interior, thus maintaining the arrangement that only the library compounds are displayed on the bead surface. The PAD approach of forming multiple concentric functional layers inside a bead is simple, reliable, and may have other applications in addition to OBOC combinatorial library bead encoding, such as the development of novel optically encoded beads for multiplex immunodiagnostics or even information recording.

Introduction

In the "one-bead one-compound" (OBOC) combinatorial library method,[1a] thousands to millions of compound beads are rapidly generated using the "split−mix synthesis" approach,[1a−c] in such a manner that each bead displays only a single compound entity.[1a,d] After biological screening, the positive beads are physically isolated for structural analysis. For peptide beads, we routinely use an automatic protein microsequencer in conjunction with Edman chemistry to determine the peptide sequence on an individual bead; however, this technique is time-consuming and expensive. To facilitate the structural analysis of peptide beads, two approaches based on mass spectrometry as the sequencing technique have been developed for OBOC combinatorial libraries. These two approaches, namely, the ladder-sequencing approach and the ladder-synthesis approach, are summarized in Figure 1A and B, respectively.[2] The "ladder-sequencing" approach, first introduced by Chait et al.[3a] and subsequently modified by others,[3b,c] used a mixture of phenylisothiocyanate (PITC, 95%, as Edman degradation reagent) and phenylisocyanate (PIC, 5%, as terminating reagent) in each step of sequential Edman degradation to generate a peptide ladder on each bead. The ladder members from a positive bead were released and subsequently analyzed by mass spectrometry to elucidate the sequences

of the original peptides by calculating the mass differences between adjacent peaks. The utility of this approach has been confirmed. However, it is limited to Edman degradative libraries, such as α-peptide or peptoid libraries (consist of α-amino acids), with a free N terminus. It cannot be applied to other diverse libraries, such as N-terminal blocked libraries, β-peptide libraries (consisting of β-amino acids), inverted peptide libraries (from C terminus to N terminus),[4] and peptidomimetic or small molecule libraries. The "ladder-synthesis" approach originally described by Sepetov et al.[5a] and Youngquist et al.[5b] can overcome some of these limitations. In this approach, the bead-bound peptides were encoded with a series of sequence-specific, partially terminated products by capping a small portion of the peptides at each coupling cycle of the library synthesis. Thus, a ladder for each compound had been generated prior to biological screening. Other variations of this "ladder-synthesis" method include the use of the same amino acid but a different protecting group as capping reagent (e.g., use Boc-Ala as the capping reagent when coupling Fmoc-Ala)[6] or the use of partial incorporation of methionine at each coupling step (e.g., use 5% of methionine and 95% of amino acid as coupling reagents in each step such that a ladder can be obtained upon cyanogen bromide cleavage).[4] Unlike the ladder-sequencing approach, the ladder-synthesis approach can be applied to N-terminally blocked peptides, peptides with β-amino acids or other nonsequencable building blocks.

* Corresponding author. Phone: (916) 734-8012. Fax: (916) 734-7946. E-mail: Kit.lam@ucdmc.ucdavis.edu.
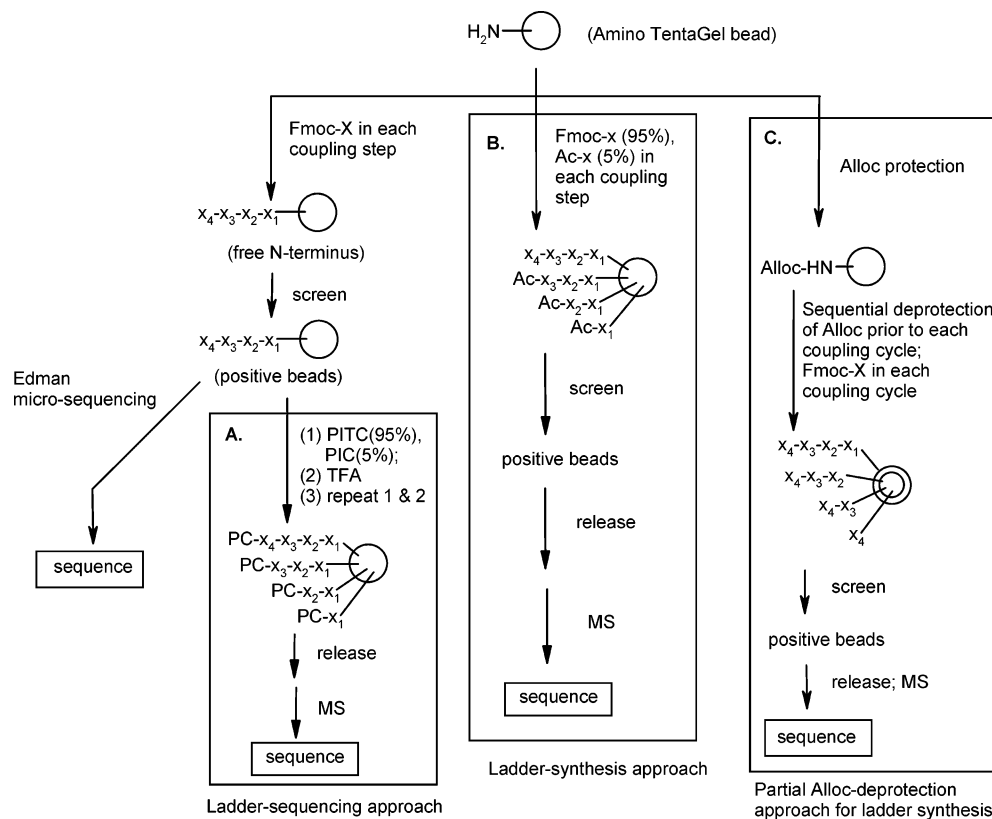
Figure 1. Methods for determining peptide sequence on a single OBOC combinatorial library bead. The "partial Alloc-deprotection approach for ladder-synthesis" method described in this report is shown on the right (X = amino acid, Ac = acetyl, PITC = phenyl isothiocyanate, PIC = phenyl isocyanate, PC = phenyl carbamyl, and Alloc = allyloxycarbonyl. Adapted from Wang, et al. *J. Am. Chem. Soc.* **2004**, *126,* 5740−5749).

However, the major disadvantage of the conventional ladder-synthesis method is that all the ladder members are displayed together with the full-length library compound on the bead surface. These ladder members may interact with the screening probes, thus complicating the interpretation of the screening result. The ideal ladder-synthesis method for OBOC libraries is to confine the truncated ladder members in the bead interior and only allow the outer layer of beads to construct the complete library compounds. This arrangement can eliminate the interference of the truncated tags with biological screening.

We previously reported a bilayer-bead concept that beads could be topologically segregated into double layers by proteases.[7] On such bilayer beads, the outer layer of beads could be used for library compound synthesis and the bead interior for coding tag synthesis. However, this segregation method was tedious and not very flexible. To further exploit the utility of this concept, we subsequently developed a much more simple and robust biphasic approach to prepare the bilayer beads.[8] In this biphasic approach, TentaGel beads were first thoroughly swollen in water. After the excess water was drained, a limiting amount of amino-protecting reagent, such as Fmoc-OSu, dissolved in organic solvent was added to the swollen beads. Under this condition, only the outer layer of the TentaGel beads was derivatized. We have successfully used the biphasic approach to develop three different encoding methods for OBOC peptidomimetic and small molecule libraries, with library compounds on the bead surface and one to three different coding tags in the bead

interior.[2,8,9] However, this biphasic approach is practical for only one or two bilayer partitions. For multistep ladder synthesis of OBOC combinatorial libraries (e.g., more than five steps), a more versatile approach to partition the bead, layer by layer, is needed. We herein report a novel partial Alloc-deprotection (PAD) approach that enables us to achieve this goal. In the PAD approach, Alloc-protected TentaGel beads can be partially deprotected, sequentially layer by layer, starting from the bead surface toward the bead interior. The degree of deprotection or thickness of each layer can be controlled by varying the exposure time of the beads to the deprotecting agent, palladium. By taking advantage of this new development, we have designed a new ladder-synthesis method for OBOC combinatorial libraries. This method differs from the conventional ladder-synthesis method in the following manners: (i) only one building block, instead of a mixture of two building blocks, is used in each coupling step; therefore, the problems resulting from the different coupling rate of two building blocks can be avoided; (ii) the ladder members of library compounds are topologically confined (bead surface, the complete library compounds; bead interior, truncated ladder members), thus eliminating the potential interference of truncated ladder members with biological screening; and (iii) each library bead carries a reverse ladder, and for α-peptide and peptoid libraries, the reverse ladder, unlike the forward ladder generated by the conventional ladder-synthesis method, does not interfere with Edman microsequencing. For the libraries containing isobaric residues, a simple encoding strategy has been introduced so

Scheme 1. Generation of Bilayer Beads by "Partial Alloc-Deprotection (PAD)" Approach[a]
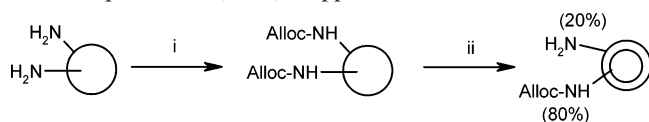
[a] Reagents and conditions: (i) Alloc-OSu (3 equiv), DIPEA (6 equiv), 1 h; (ii) (ii-a) swell in water, 24 h; (ii-b) Pd(PPh$_3$)$_4$ (0.24 equiv), PhSiH$_3$ (20 equiv), DCM, 7 min.

that the isobaric residues can be differentiated by the coding tags. The remarkable advantage of this encoding strategy is that the coding tags can be constructed together with the truncated ladder members in the bead interior so that only the library compounds are displayed on the bead surface.

To simplify the interpretation of the mass spectra, bromine-containing β-amino acid was incorporated into the cleavable linker, as reported,[2,5a,10] so that each cleavage product generates a characteristic doublet. To demonstrate the new ladder-synthesis method, a model bead-bound tetrapeptide and a model encoded pentapeptide library were synthesized. Some beads were individually sliced into two pieces. One piece was submitted for microsequencing; the remaining piece was chemically treated with cyanogen bromide, and the releasate was analyzed by MALDI-TOF MS. Consistent results obtained by these two different sequencing techniques have validated this method.

Results and Discussion

Partial Alloc-Deprotection (PAD) Approach. Because we have successfully utilized the bilayer beads in the encoding methods for small molecule and peptidomimetic libraries,[2,8,9] our current effort is to explore the applicability of the bilayer-bead concept for a multistep ladder-synthesis method of OBOC combinatorial libraries. We seek to develop such an approach that meets two requirements: (i) it enables us to topologically segregate beads so that the outer layer of beads can be constructed with the full-length library compounds, and (ii) it permits stepwise liberation of some free amino groups in the interior of the same bead so that a series of ladder members can be conveniently prepared.

With these requirements in mind, we have developed a novel approach termed "partial Alloc-deprotection" (PAD) by using a deprotection reaction under the biphasic condition. Scheme 1 illustrates the generation of bilayer beads by the PAD approach. Typically, the amino groups of the TentaGel beads (diameter, 90 μm) are first protected with an Alloc group and then thoroughly swollen in water, followed by deprotection using palladium chemistry[11] in an organic solvent for a predefined limited time, resulting in beads with a deprotected outer layer (free N termini) and a protected inner core. The thickness of the outer layer (Alloc-deprotection percentage) of beads is dependent on the duration of deprotection. The remaining Alloc groups in the bead interior can be successively exploited by repeating the same biphasic deprotection approach. Alloc is crucial to this approach because it can be selectively removed in the presence of other protecting groups, such as Fmoc and Boc, and its deprotection has been demonstrated to be time-dependent under the biphasic condition.
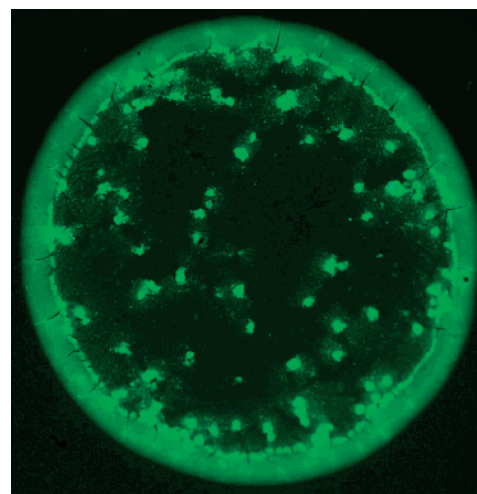

Figure 2. Fluorescent photomicrograph of a sliced FITC-labeled bilayer bead (generated by the PAD approach) under a fluorescent microscope. The deprotected outer layer containing free amino groups was labeled with FITC and, therefore, fluoresced. The Alloc-protected bead interior remained dark. Some fluorescent spots scattered inside the bead are probably due to incomplete swelling of the bead in water prior to Alloc derivatization.

To test the efficacy of bilayer segregation generated by the PAD approach, water-swollen Alloc beads were treated with Pd(PPh$_3$)$_4$/PhSiH$_3$ in DCM for 7 min, and the exposed N termini were labeled with FITC (fluorescein-5-isothiocyanate, fluorescent probe). Some beads were cross sectioned with two scalpels under a dissecting microscope, and a section of one bead was visualized under a fluorescent microscope (Figure 2). Since only the deprotected sites (free N termini) would allow FITC-labeling, the deprotected layer fluoresced and could be visualized microscopically under a fluorescent microscope as a bright green fluorescent outer ring and a dark inner core. A similar result was also observed with intact beads using confocal microscopy (data not shown). Some fluorescent spots were found scattered inside the bead interior, as well. We are not sure what causes this. It could be due to incomplete swelling of the bead during the initial Alloc-protection step.

To determine the time dependence of Alloc deprotection with palladium, several aliquots of water-swollen Alloc beads were treated with palladium in two different solvent systems (DCM/ether (50/50, v/v)[8] and DCM alone) over time. The treated beads were then reprotected with Fmoc and later released for UV quantitation.[12] Figure 3 illustrated the Alloc-deprotection plots under the two different solvent systems. When DCM/ether was used as the solvent, Alloc deprotection proceeded rapidly to 80% completion within 10 min. In contrast, deprotection proceeded gradually in the presence of DCM alone, making it an ideal organic solvent for PAD. Thus, using DCM as the solvent and by limiting the exposure time of the beads to palladium, Alloc groups on the beads can be partially and sequentially removed starting from the outer layer toward the bead interior in a highly controlled manner. It should be noted that the Alloc deprotection percentages achieved in the subsequent PAD processes could be smaller than those indicated by the plot in Figure 3 because the deprotection reagents need to diffuse through the outer layers first and then into the interior of the bead.
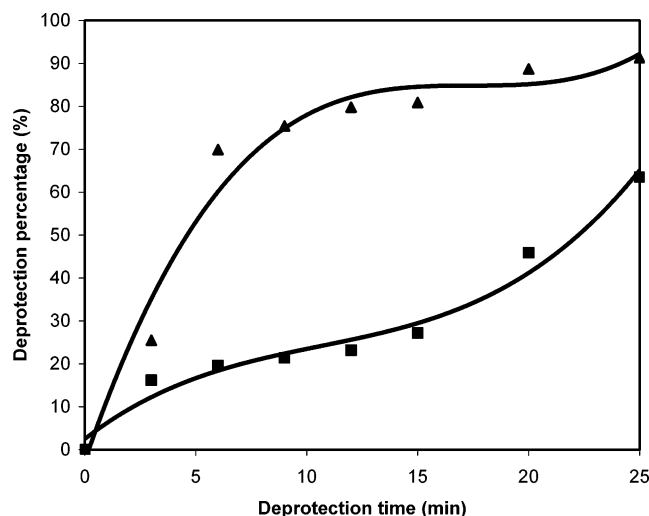
Figure 3. Degree of deprotection (determined by Fmoc quantitation) as a function of exposure time to palladium (deprotection time) in two different organic solvents: DCM/ether (50:50, v/v) (▲) and DCM (■). See text for experimental details.

Nevertheless, current deprotection protocol enables us to generate an adequate amount of five distinct ladder sequences from each bead for unambiguous decoding.

New Ladder Synthesis Method. Since the PAD approach allows time-controlled partial removal of Alloc groups on beads, as demonstrated above, it creates the feasibility of repetitive Alloc deprotection on the same bead during library synthesis. Deprotection begins sequentially from the bead surface, layer by layer, toward the bead center. On the basis of this approach, we developed a new ladder synthesis method for OBOC libraries (Scheme 2). To simplify the drawings in the Scheme, we combine the multiple inner layers as one and only show the outer layer as a separate layer. In this method, the beads preprotected with an Alloc group are used for library synthesis. Initially, the Alloc beads are thoroughly swollen in water for a long time (24 h) to ensure complete swelling and then topologically derivatized with a deprotected outer layer (free N termini, $\sim$20% of total substitution) and a protected inner core ($\sim$80% of total substitution) by the PAD approach (deprotection time, 7 min). The exposed N termini on the outer layer are coupled with a Fmoc-amino acid (Fmoc-X_1) to establish the first library residue (X_1). The remaining Alloc groups in the bead

interior are then partially liberated using the same approach (swelling time in water, 3 h; deprotection time, 4 min) to produce $\sim$10% free amino-group as a second layer inside the beads. After Fmoc is removed from the outer layer with piperidine, all free N termini on both first and second layers are simultaneously coupled with the second residue Fmoc-X_2, resulting in two distinct peptides in the two layers (outer layer, Fmoc-X_2X_1-; second layer, Fmoc-X_2-). These three steps of partial Alloc deprotection (PAD), Fmoc deprotection, and coupling with a Fmoc-amino acid represent one synthetic cycle. Every time this synthetic cycle is repeated, a new layer is formed inside the bead interior, and an additional residue is added to the N termini of each ladder peptide on each preceding layer. As indicated in Scheme 2, the beads are split into bead aliquots prior to coupling with Fmoc-amino acid, and the bead aliquots are mixed together prior to the PAD step. This split–mix synthesis approach is essential to produce the OBOC combinatorial library.[1a] In the last (e.g., fourth) cycle, the remaining Alloc groups on each bead are completely removed according to standard Alloc-deprotection protocol,[11] and the last Fmoc residue (X_4) is incorporated. After Fmoc and side-chain deprotection, each library bead carries a complete library compound ($X_4X_3X_2X_1$-) on the outer layer, and three truncated ladder members ($X_4X_3X_2$-, X_4X_3- and X_4-) in the bead interior.

In contrast to the conventional ladder-synthesis method (Figure 1B) that uses a mixture of two building blocks (e.g., acetyl-amino acid and Fmoc-amino acid) for coupling,[5] this method (Figure 1C) only uses one single building block for coupling during each coupling step, thereby avoiding the problems caused by the differential coupling rates of two different building blocks. Moreover, all the truncated ladder members are confined to the bead interior, and only the full-length library compound is displayed on the bead surface. As a result, the undesirable interference of ladder tags with the biological screening can be avoided. In addition, this method generates a reverse ladder that allows one to determine the peptide sequences by calculating mass differences between each two adjacent peaks in mass spectrometry. A unique feature of the "reverse ladder" generated by this method, as opposed to the "forward ladder" generated by the conventional ladder-synthesis method or ladder-sequencing method (Figure 1A, B), is that the reverse ladder (Figure

Scheme 2. General Synthetic Route of New Ladder-Synthesis Method for a Tetrapeptide Library[a]
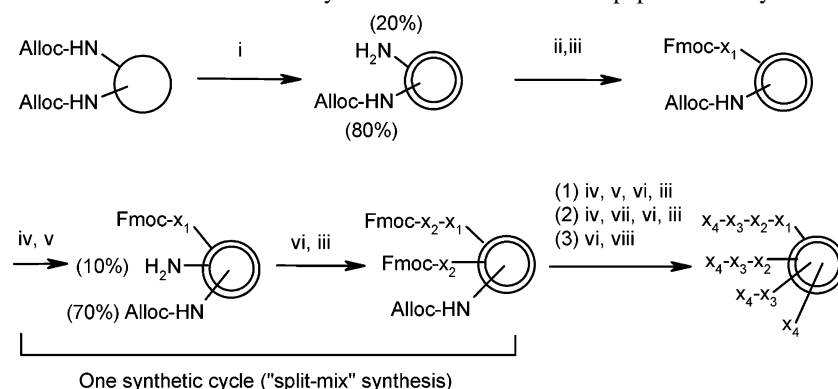
^{a} Reagents and conditions: (i) (a) H_2O, 24 h; (b) Pd(PPh$_3$)$_4$ (0.24 equiv), PhSiH$_3$ (20 equiv), DCM, 7 min; (ii) split beads into aliquots; (iii) Fmoc-amino acid (Fmoc-X, 3 equiv), HOBt (3 equiv), DIC (3 equiv); (iv) mix beads; (v) (a) H_2O, 3 h; (b) Pd(PPh$_3$)$_4$ (0.24 equiv), PhSiH$_3$ (20 equiv), DCM, 4 min; (vi) 25% piperidine, 10 min, twice; (vii) Pd(PPh$_3$)$_4$ (0.24 equiv), PhSiH$_3$ (20 equiv), DCM, 1 h; (viii) TFA/TIS/H_2O (95:2.5:2.5, v/v/v), 2.5 h.
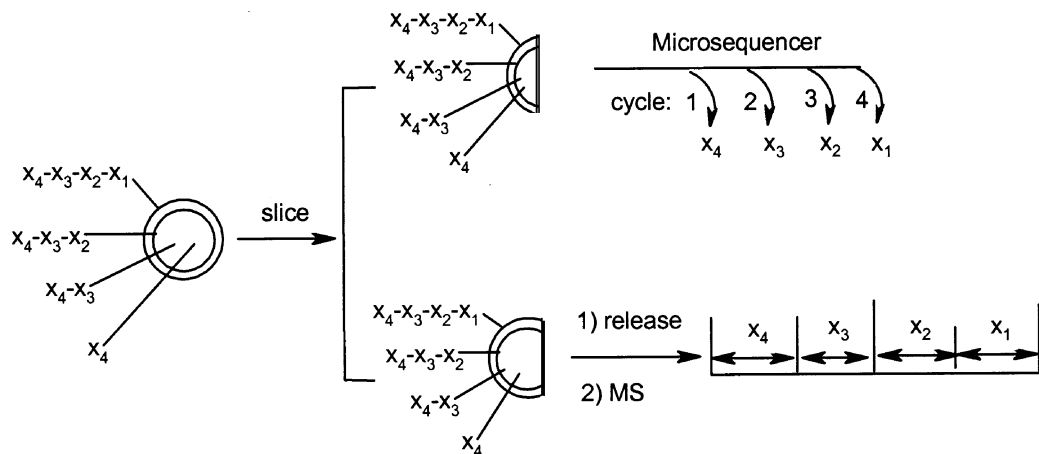
Figure 4. Schematic of two sequencing techniques (protein microsequencer and MALDI-TOF MS) carried out on two different sections of the same compound bead prepared by the new ladder-synthesis method. The smaller piece is sequenced directly, without releasing of the ladders, with the automatic protein microsequencer using Edman chemistry. The larger piece is treated with CNBr and the releasate was analyzed by MALDI-TOF MS.

1C) will not interfere with Edman degradation. This is because the same PTC-amino acid created by the reaction of PITC with the N terminus of the amino acid (PTC, phenyl thiocarbamyl) is released from both the full-length peptide and its truncated ladders during each Edman degradation cycle. One obvious application of this unique feature is that each positive bead isolated from screening can be sectioned into two pieces. If the MALDI-TOF MS result obtained from one piece is ambiguous or incomplete, for example, due to isobaric amino acids or other technical difficulties, the other piece can be subjected to automatic Edman degradation (Figure 4).

To validate the new ladder-synthesis method, a model tetrapeptide, Pro-Leu-Gly-Ile (P-L-G—I) together with its three truncated ladder members (P-L-G, P-L, and P) was prepared on Tentagel beads according to Scheme 2 (where $X_1 = I$, $X_2 = G$, $X_3 = L$, and $X_4 = P$). Prior to the peptide synthesis, a cleavable linker consisting of three components (methionine, 3-(4-bromophenyl)-β-alanine, and 2,2′-ethyl-enedioxy-bis(ethylamine) monosuccinamide[13]), was assembled on beads. As recently reported,[2] methionine acts as a chemoselective cleavage site by cyanogen bromide[14] to ensure quantitative release of compounds from beads; to simplify interpretation, the bromine-containing β-amino acid can generate a characteristic isotopic doublet[2,5a,10] on the mass spectra for each compound; the last hydrophilic component can improve the solubility and extractability of the released compounds from the beads.[2,9] Moreover, this final component will produce, together with the other two components, a big mass shift of 555 Da, thus pushing all releasates away from the background noise region in MALDI-TOF MS.[15] The beads with the preassembled cleavable linker were then protected with Alloc. According to Scheme 2, the beads were treated by the PAD approach for 7 min in the first step to achieve ∼20% of free N termini (for the outer layer) and 4 min in the subsequent two steps to achieve ∼10% of free N termini in each of the two other layers. In the last step, the remaining Alloc groups were thoroughly removed. The free N termini liberated by the PAD approach in each step were monitored by the chloranil test.[16] By the end of the library synthesis, each bead carried a complete peptide (P-L-G-I-

linker) on the outer layer (20% of the total substitution) and three ladder members (P-L-G-linker, 10%; P-L-linker, 10%; and P-linker, 60%) in the bead interior. Five beads were randomly isolated. One of them was sliced into two pieces. The smaller piece was sequenced by the protein microsequencer, and the larger piece plus four other intact beads were individually cleaved by cyanogen bromide, and the releasates were analyzed by MALDI-TOF MS separately (Figure 4). As expected, the microsequencer unambiguously confirmed the peptide sequence of PLGI as originally designed. The releasates from each bead and bead piece generated an identical mass spectrum (Figure 5). The first mass peak beyond m/e of 557 (mass of the protonated cleavage linker which does not appear in the mass spectrum) had a molecular mass of 654.1498. Because there was a mass loss of 18 Da (H_2O) when the peptide bond was formed, the molecular mass for residue X_4 could be determined by the following formula: (mass of the first peak − mass of the cleavage linker + 18) or (654 − 557 + 18). The calculated molecular mass of 115 confirms the identity of X_4 as Pro. The molecular mass of the next residue (X_3) is simply the mass difference between the two peaks (767 − 654 = 113), and it could be either isoleucine or leucine because they have the same molecular weight. The third residue, X_2, was identified as glycine (824 − 767 = 57). Likewise, the last residue (X_1) could be either leucine or isoleucine (937 − 874 = 113). Therefore, from the MS analysis, the peptide sequence was P-L(/I)-G-I(/L), which is totally consistent with the Edman degradation result, thus validating our methodology. However, isobaric residues (e.g., isoleucine and leucine, or glutamine and lysine) cannot be differentiated by MS; therefore, for OBOC libraries with isobaric residues, we may use either Edman sequencing of the second half of the bead as a backup or use a simple encoding strategy to differentiate isobaric residues so that the identity of the isobaric residue can be determined by MS.

Encoding Strategy for Isobaric Residues. As shown above, the isobaric residues, such as leucine and isoleucine, cannot be differentiated by MS. An encoding strategy is therefore needed to differentiate the isobaric residues in libraries. In the conventional ladder-synthesis method,[5] two
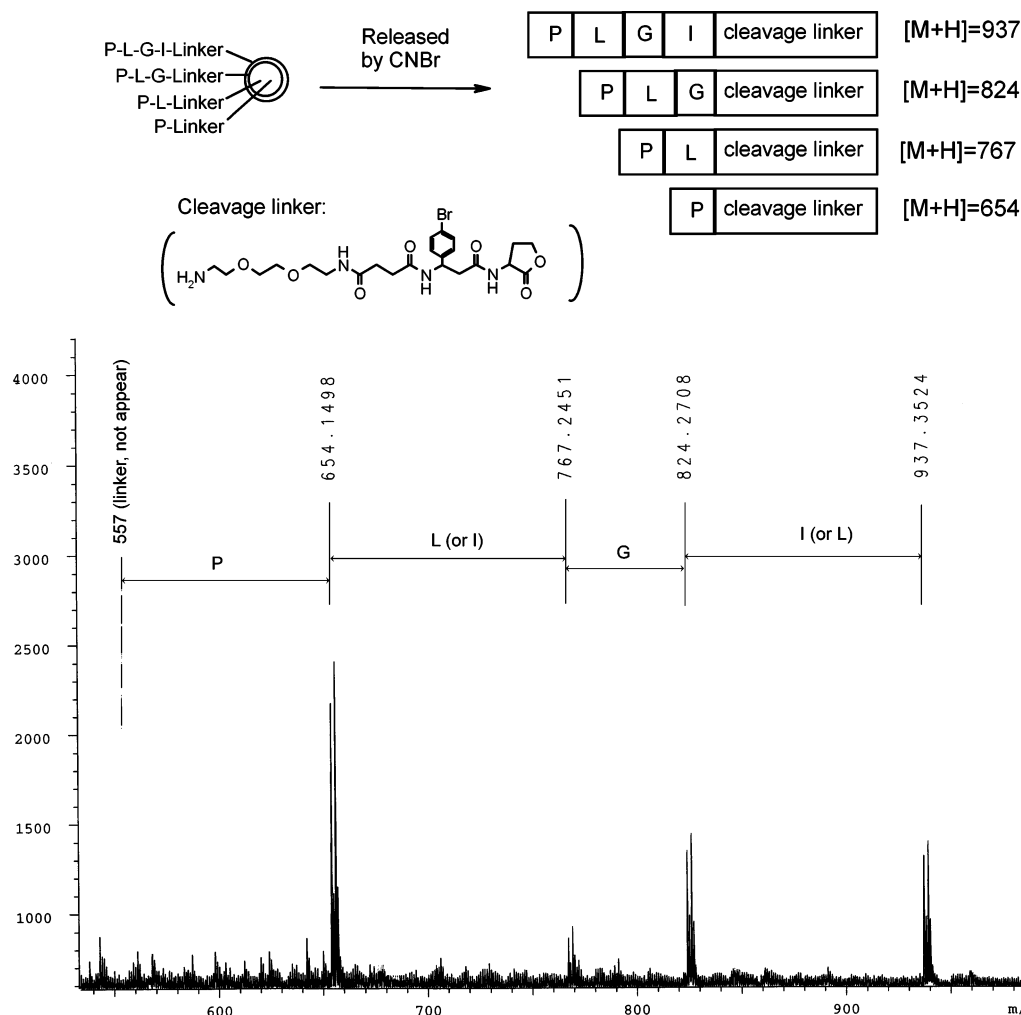
Figure 5. The mass spectrum of the model peptide ladder released from one single bead prepared by the new ladder-synthesis method. The peptide sequence (starting from the N terminus) is identified by calculating the mass differences between adjacent peaks from low mass to high mass. The first residue was determined as P by deducting 557 (mass of protonated cleavage linker, does not appear in the mass spectrum) from the mass of the first peak. The full-length peptide sequence was identified as P-L(/I)-G-I(/L).

different terminating reagents (e.g., *N*-acetylalanine and *N*-propionylalanine) together with an isobaric Fmoc-amino acid were simultaneously coupled to beads in each step to generate an additional coding tag. They produced doublet peaks with a unique mass difference so that the identity of the isobaric residue can be determined. However, this strategy further compromised the library quality because additional coding tags were displayed together with the complete library compound and all the other truncated coding tags on the bead surface. To avoid this problem, we developed a new encoding strategy that can confine all the coding tags together with the truncated ladder members in the bead interior. To illustrate how this encoding strategy works, the synthesis of a hypothetical tripeptide library is shown in Scheme 3. Among the three building blocks (A, B, and E), A and B are a pair of isobaric residues. Building block A is encoded by a coding tag, but building block B is not encoded. As a result, the presence of "a" in the final mass spectrum signifies A and not B, even though both A and B have identical mass.

The method for synthesizing a nonencoded OBOC combinatorial peptide "ladder-synthesis" library (Scheme 2) is very similar to that of standard OBOC peptide libraries, as originally reported,[1a] except that (i) cleavable linkers, fol-

lowed by Alloc, are incorporated onto the beads prior to library synthesis, and (ii) a PAD step is added prior to each Fmoc-deprotection step. The synthesis of an encoded OBOC combinatorial peptide "ladder-synthesis" library, although it looks a little complicated in Scheme 3, is indeed rather simple once the general principle is followed. The synthetic scheme of the encoded library is essentially the same as that of the nonencoded library (Scheme 2), except for an extra encoding step applied to the bead aliquot that has just reacted with the designated encoded isobaric building block. This bead aliquot will undergo PAD and react with a coding tag (e.g., Gly) prior to mixing with the rest of the bead pool that has just undergone PAD. This extra step is applied to any designated isobaric building block (e.g., isoleucine for the isoleucine/leucine pair and glutamine for the glutamine/lysine pair) at any coupling cycle except the last. At the last coupling cycle, the bead aliquot that has just reacted with the designated encoded isobaric building block will undergo full Alloc deprotection followed by coupling with another coding tag (e.g., Ac) rather than Gly.

According to Scheme 3, the synthesis of the encoded tripeptide "ladder-synthesis" library is as follows: In each synthetic cycle, beads are first partially deprotected by the

Scheme 3. General Encoding Ladder Synthesis for a Trimer Combinatorial Library Containing Isobaric Residues[a]
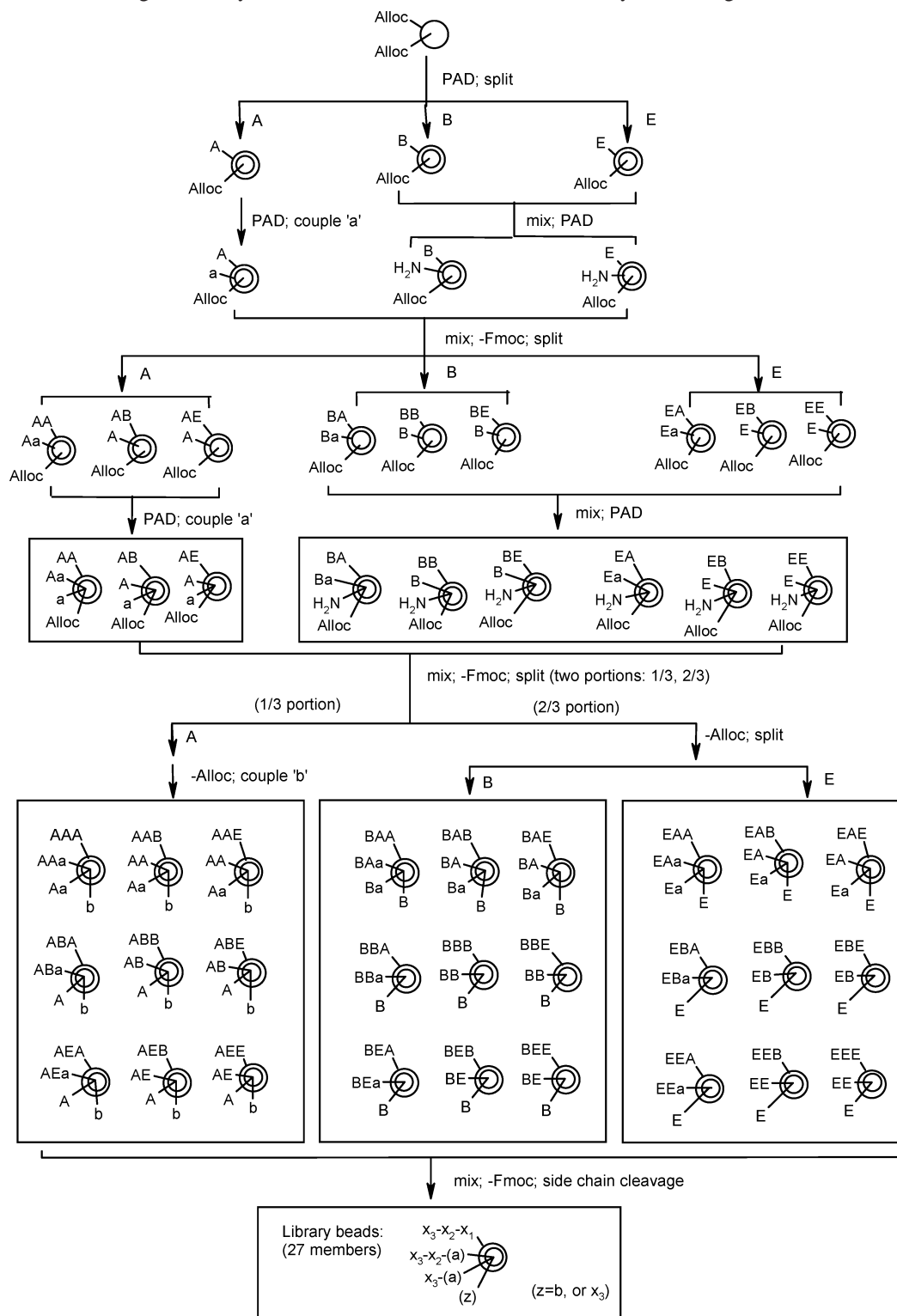
[a] The residues in the brackets of the ladder members on the resulting library beads may or may not be present in the corresponding ladder members. In this library, A and B are a pair of isobaric residues.

PAD approach, followed by a Fmoc-amino acid coupling. Then the aliquot of beads coupled with A are partially deprotected, followed by coupling with a coding amino acid "a" (e.g., Gly), the other aliquots of beads are pooled and partially deprotected. Since the terminus of A is still Fmoc-protected after coupling (not shown in the scheme), the subsequent coding block ("a") can be anchored only to the exposed free N termini. All beads are then combined together, Fmoc-deprotected and split for the next synthetic cycle. By repeating the synthetic cycle, the ladder segments generated in the preceding cycles are simultaneously extended with the current coupling of a residue. However, in the last cycle of library synthesis, the encoding is different. In this cycle, all beads are first split into two aliquots. The small aliquot with 1/3 portion of the beads are coupled with A and then thoroughly Alloc-deprotected, followed by
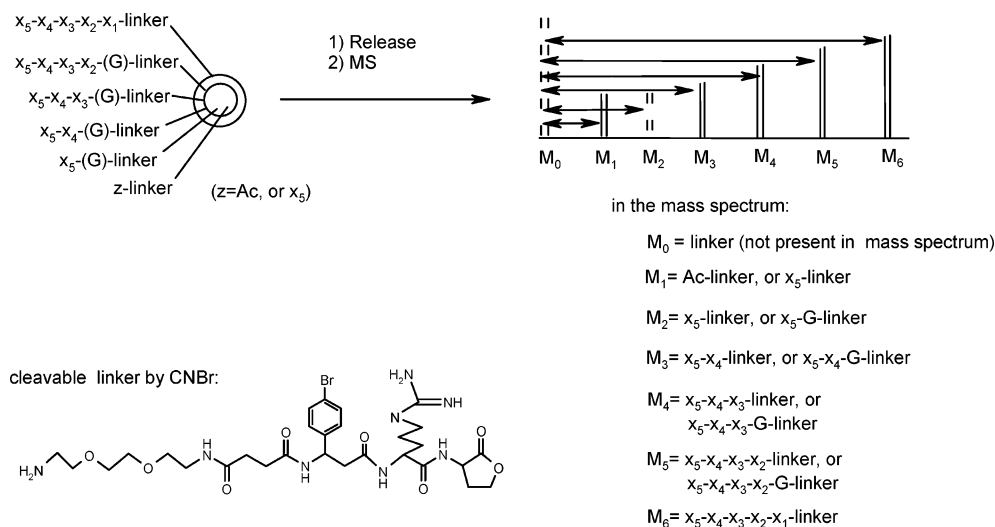
Figure 6. General structure of ladder compounds on an encoded pentapeptide library bead and hypothetical mass spectrum of releasate with six mass peaks (M_1-M_6) from one bead. The chemical structure of the cleavable linker is shown.

acylation with a different coding block "b" (e.g., Ac); the large aliquot with 2/3 portion of the beads is thoroughly deprotected to remove all remaining Alloc groups, then split into two aliquots, followed by coupling with a residue (B or E) to each aliquot. As a result, each of the encoded OBOC library beads carries four ladder products: full-length library compound ($X_3X_2X_1$-) on the outer layer and three truncated ladder products containing the coding tags in the bead interior ($X_3X_2(a)$-, $X_3(a)$- and z- (z = b when X_3 is the encoded residue, or z = X_3 when X_3 is not the encoded residue); "a" in the brackets of the ladder members means that it may or may not be present in the corresponding ladder segments). Obviously, when z = X_3 and "a" is not present in the second ladder product, $X_3(a)$- and z- will be degenerate as a single product (X_3-). In this case, the library compound actually has only three ladder members.

Fmoc-glycine is the preferred coding block (Fmoc-"a") because of its lowest molecular weight among all amino acids, making all ladder products on a single bead always appear in a fixed order in the mass spectrum (from low mass to high mass). The reason to use a different coding block ("b" ≠ G) in the last cycle is to allow glycine as one of the library residues; "b" can be a small organic acid such as acetic acid or propionic acid. For a library containing *n* residues, the synthetic cycle can be reiterated for *n* times to achieve the desired length of library sequence. In this case, the Alloc-deprotection time in each PAD step can be properly shortened to reserve sufficient Alloc groups in the bead interior for subsequent multistep liberations. If a library contains several pairs of isobaric residues, an identical coding block (e.g., glycine) can still be used to encode one of each pair because the coding tag is used only to reveal the presence of an encoded residue. But for libraries containing a residue with more than two isobaric isomers, one may use additional small coding blocks (e.g., Ala) to encode additional isomers.

Using the above encoding strategy, one can readily generate an encoding peptide library with a desired sequence length. For example, an encoded pentapeptide library ($X_5X_4X_3X_2X_1$-bead) including two pairs of isobaric residues (isoleucine/leucine, and glutamine/lysine) can be synthesized in one week. In the library, glutamine and isoleucine are encoded by glycine (G), but lysine and leucine are not. The general structure of the ladder family on the pentapeptide library beads and hypothetical mass spectrum are illustrated in Figure 6. Each bead carries six ladders: M_1-M_6. In each ladder segment, the coding tag (G) in the brackets means that it may or may not be present. It is important to keep in mind that (i) an extra encoding step is applied to the bead aliquot that has just reacted with the designated encoded isobaric building block. This bead aliquot undergoes PAD, reacts with a coding tag Fmoc-G prior to mixing with the rest of the bead pool that has just undergone PAD, and (ii) at the last coupling cycle, the bead aliquot that has just reacted with the designated encoded isobaric building block will undergo Alloc deprotection, followed by coupling with another coding tag, acetic acid (Ac) rather than glycine. As a result, the presence of a coding tag (G) in a ladder member signifies that an encoded residue (either isoleucine or glutamine) is present in the preceding ladder member corresponding to that position. However, when z = X_5- and G is absent in the ladder member of X_5-(G)-, M_1 and M_2 will be degenerate as one single peak in the mass spectrum. In this case, $M_2 = M_1$, and there will be a total of only five, instead of six, MS peaks. All ladder members are released from a bead and analyzed by MALDI-TOF MS to determine the peptide sequence. To illustrate how the encoding system works, the expected ladders generated from four different hypothetical beads isolated from the $X_5X_4X_3X_2X_1$-bead library, with isoleucine and glutamine as the designated encoded isobaric building block, is shown in Table 1. Bead I has no isobaric building block; therefore $M_2 = M_1$, and five standard reverse ladders are expected. For bead II, the presence of ladder X_5-a-linker and X_5-I-linker indicates that X_4 is an encoded isobaric building block, in this case, isoleucine. For bead III, the presence of b-linker and I-linker signifies that the last residue incorporated (i.e., the amino terminus) is an encoded isobaric building block, in this case, isoleucine. For bead IV, the presence of X_5-a-linker, X_5-I-linker, X_5-I-X_3-X_2-a-linker, and X_5-I-X_3-X_2-Q-linker indicates that

Table 1. Expected Ladder Sequences Released from Five Hypothetical Beads, Some with Designated Encoded Isobaric Building Blocks (I for I/L; Q for Q/K) at Specific Position, Isolated from an OBOC Combinatorial Peptide Library ($X_5X_4X_3X_2X_1$-Bead) Generated by the New Ladder-Synthesis Method Shown in Scheme 3[a]

MS peak	$X_5X_4X_3X_2X_1$-linker-bead bead (I)	$X_5IX_3X_2X_1$-linker-bead bead (II)	$IX_4X_3X_2X_1$-linker-bead bead (III)	$X_5IX_3X_2Q$-linker-bead bead (IV)	$I\text{-}IX_3QX_1$-linker-bead bead (V)
M_6	$X_5X_4X_3X_2X_1$-linker	$X_5IX_3X_2X_1$-linker	$IX_4X_3X_2X_1$-linker	$X_5IX_3X_2Q$-linker	$I\text{-}IX_3QX_1$-linker
M_5	$X_5X_4X_3X_2$-linker	$X_5IX_3X_2$-linker	$IX_4X_3X_2$-linker	$X_5IX_3X_2a$-linker	$I\text{-}IX_3Q$-linker
M_4	$X_5X_4X_3$-linker	X_5IX_3-linker	IX_4X_3-linker	X_5IX_3-linker	$I\text{-}IX_3a$-linker
M_3	X_5X_4-linker	X_5I-linker	IX_4-linker	X_5I-linker	$I\text{-}I$-linker
M_2	(same as M_1)	X_5a-linker	I-linker	X_5a-linker	Ia-linker
M_1	X_5-linker	X_5-linker	b-linker	X_5-linker	b-linker
M_0	linker	linker	linker	linker	linker

[a] Encoding blocks: "a" = G (glycine); "b" = Ac (acetyl group). In the library, Q and I are encoded but their corresponding isomers K and L are not.

that both X_4 and X_1 are the encoded isobaric building blocks, isoleucine and glutamine, respectively. For bead V, the presence of b-linker signifies that X_5 is an encoded isobaric building block, and the presence of I-a-linker, I-I-linker, I-I-X_3-a-linker, and I-I-X_3-Q-linker indicates that both X_4 and X_2 are the encoded isobaric building blocks, isoleucine and glutamine, respectively.

General Decoding Strategy. Although most MS spectra can be readily decoded by inspection using the above principles and examples, it is important to develop a general decoding algorithm that works for all beads. Such an algorithm is shown in Scheme 4. On the basis of this algorithm, we have written software (Perl script included in the Supporting Information) to facilitate the decoding process. The algorithm is based on the following decoding principles using the encoded pentapeptide library shown in Figure 6 as an example: (i) Identity of each residue in the sequence is determined from low mass to high mass by calculating the mass differences between the mass of each cleavage ladder member and the cleavage linker (do not appear in mass spectra). (ii) Molecular masses of all building blocks and the sum of building block and G (building block mass + 57) are unique; therefore, each mass peak except M_1 can reveal two facts: the molecular mass of the residue and the presence or absence of the coding block (G). (iii) M_1 peak reveals the presence or absence of the coding block (Ac), and in some cases, it reveals the identity of X_5 (when $M_1 = X_5$-linker). For example, the presence of Ac indicates that X_5 is an encoded residue. (iv) The absence of the coding tags G or Ac indicates that the corresponding residue is not an encoded residue. (v) It should be noted that the identity of the residue cannot be determined solely by the coding tag because there may be several pairs of isobaric residues in the library. In this case, both the molecular weight of the residue (determined from the preceding larger ladder) and information on the presence or absence of the coding tag are needed. As illustrated by the correlation between the coding tags and library residues shown in Table 1, the presence (or absence) of the coding tag (Ac or G) in one ladder member reveals that the corresponding residue in the preceding larger ladder member is (or is not) an encoded residue. (vi) When only five peaks appear in the mass spectrum, $M_2 = M_1$.

On the basis of the algorithm (Scheme 4) developed from these principles, we can use the six mass peaks, M_1, M_2,

M_3, M_4, M_5, and M_6, from low mass to high mass, as shown in Figure 6, to identify the full-length peptide sequence. The first step of this algorithm is to determine the mass difference between M_1 and M_0. If the value matches the mass of Ac (= 42), it confirms that X_5 is an encoding residue. If the value is different from the mass of Ac, X_5 is not an encoded residue. In the second step, if $M_2 - M_0$ matches the mass of any residue (= molecular weight of any residue − 18, due to mass loss from coupling) in the library, it means that (i) the coding tag G is absent; therefore, X_4 is not an encoding residue; and (ii) the calculated molecular weight of $M_2 - M_0$ is the molecular mass of residue X_5. If $M_2 - M_0$ does not match any mass of building blocks used in the library synthesis, that suggests that the coding tag (G) is involved. In this case, $X_5 = M_2 - M_0 - G$, and X_4 is an encoded residue because of the presence of G. Once X_5 is determined, it will be used in the subsequent steps. Likewise, in the next step, if $M_3 - M_0 - X_5$ matches the mass of any library residue, then $X_4 = M_3 - M_0 - X_5$, and X_3 is not an encoded residue; otherwise, $X_4 = M_3 - M_0 - X_5 - G$, and X_3 is an encoded residue. Once X_4 is identified, it will be used together with X_5 in the subsequent steps. Using a similar strategy, X_3, X_2 can be successively identified. In the last step, since M_6 corresponds to the full-length sequence $X_5X_4X_3X_2X_1$-linker, it is easy to identify X_1 by the equation: $X_1 = M_6 - M_0 - X_5 - X_4 - X_3 - X_2$. Thus, the full-length library sequence (from N terminus) can be determined. This strategy can readily be automated and expanded to the libraries with longer sequence. A software program, Perl script, which greatly facilitates the interpretation of the mass spectra, is included in the Supporting Information.

To demonstrate the encoding and decoding strategies, a real model pentapeptide library was synthesized by the general procedure shown in Scheme 3. Seventeen eukaryotic amino acids, excluding methionine, cysteine, and glutamine, were used as building blocks in the library in which isoleucine (I) and leucine (L) were purposely used as the only pair of isobaric isomers. To ensure uniform protonated response of all cleavage products, arginine was incorporated in the cleavable linker as previously reported.[2,5b] The mass of the protonated cleavage linker shown in Figure 6 is 713 ($M_0 = 713$, does not appear in the mass spectrum). Glycine (G) was used to encode isoleucine in each synthetic cycle except in the last step, where acetic acid (Ac) was used

Scheme 4. General Decoding Strategy for an Encoded Pentapeptide Library[a]
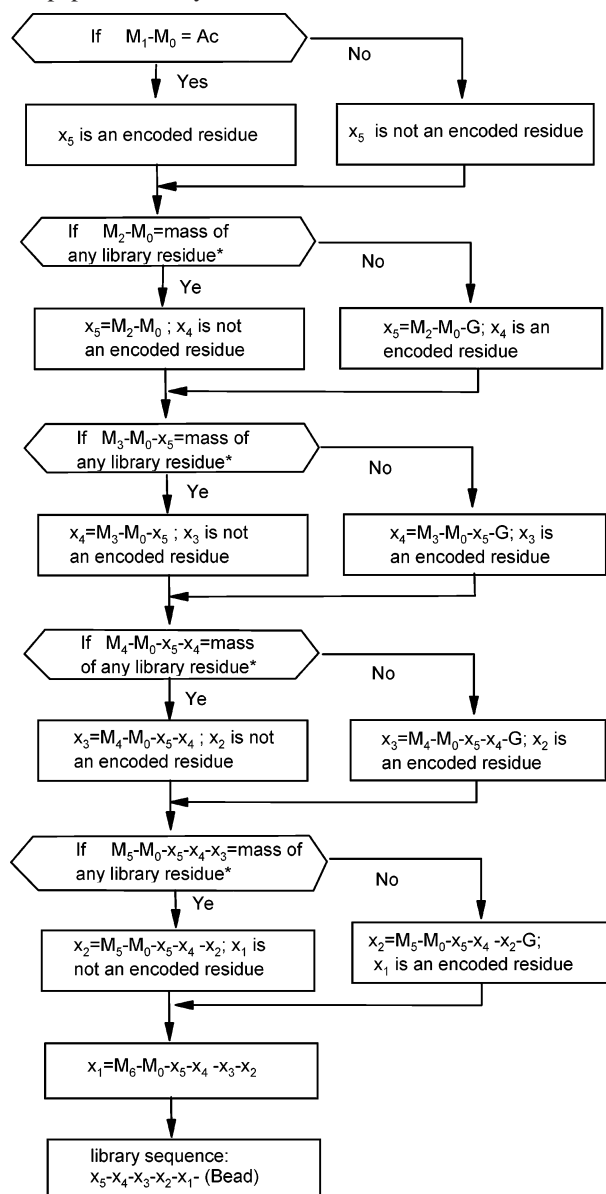
[a] The mass peaks of M_1–M_6 correspond to the ladder sequences shown in Figure 6. Each mass peak (except M_1) reveals two facts: the molecular weight of a library residue and the presence or absence of an coding tag (G). M_1 is only used to reveal the presence or absence of Ac. For libraries containing several residues with two isobaric isomers each, glycine (G, M = 57) is used to encode one of each isomer pair. In the last synthetic cycle, acetic acid (Ac, M = 42) is used for encoding instead. Mass of any library residue* = (molecular weight of any residue − 18), because of mass loss of 18 during coupling. In the model library, the mass of the protonated cleavage linker, M_0 = 713.

instead. Ten library beads were randomly isolated. Four of them were sliced into two pieces. The ladders on six intact beads and four sliced bead pieces were released individually by cyanogen bromide for MALDI-TOF MS analysis. Figure 7 illustrates two typical mass spectra of library beads (7a, coding tag observed; 7b, no coding tag observed). In Figure 7a, six distinct protonated peaks were detected: M_1 = 826, M_2 = 883, M_3 = 939, M_4 = 1010, M_5 = 1111, and M_6 = 1224. Using the above decoding strategy, the full-length pentapeptide sequence was determined as L-I-A-T-L, which was confirmed by microsequencing of the remaining piece

of the same bead. In Figure 7b, five mass peaks appeared, where M_1 = 814, (M_2 = M_1 = 814), M_3 = 927, M_4 = 984, M_5 = 1121, and M_6 = 1222. The compound sequence was identified as T-L-G-H-T, which was also confirmed by Edman microsequencer. However, since no coding tag was present in all ladder members on this bead, the analysis of the mass spectrum can be simplified as that of a nonencoded library, which used the mass differences between adjacent peaks to identify each library residue (shown in Figure 7b). It should be noted that, in this model, an encoded α-peptide library, in the four cases when the calculated mass difference in a step equals 114, 128, 156, or 186 that respectively matches N (or G + G), K (or A + G), R (or V + G) or W (or E + G), since these molecular masses have two possibilities each (e.g., 128 = K or A + 57), they may make it a little more complicated to determine the whole sequence. However, since the mass peak (M_6) of the full-length sequence does not contain any coding tag, the misleading results (X_5, X_4, X_3, X_2) in the preceding steps by the decoding strategy that only follows one choice (e.g., the mass difference of 114, 128, 156, or 186 is identified as N, K, R, or W, respectively) will no longer match M_6. In these circumstances, the other choice will reveal the correct sequence.

To test the reliability of the encoded ladder-synthesis method, we further sequenced 40 library beads by MALDI-TOF MS upon cyanogen bromide cleavage. The compound sequences of 37 beads were unambiguously identified. Only three beads did not generate sufficient peaks for analysis, of which one lost all signals, probably due to loss of the bead while transferring. For the other two beads, we observed four peaks each (losing the second peak according to the large mass differences). However, the four peaks still enabled us to identify the compound sequence because the pair of amino acids that summed to the measured mass could be easily identified, although the order of the two residues could not be determined due to the loss of the second peak. Thus, the overall MS-based sequencing success rate of over 92% has demonstrated the reliability of the method.

Using the new ladder synthesis method described in this Article, all ladder members inside the bead, as well as the full-length library compound on the bead surface, will have the cleavable linker (consisting of 2,2′-ethylenedioxy-bis-(ethylamine) monosuccinamide, 3-(4-bromophenyl)-β-ala-nine, arginine, and methionine). This long linker may sometimes have undesirable effects on biological screening. Therefore, in some applications, one may want to link the library compound to the bead surface directly via the amino terminus of the PEG on TentaGel, that is, without the long cleavable linker. This can easily be achieved by first using the PDA approach to deprotect the outermost layer of the Alloc bead (e.g., 10%), block the deprotected amino groups with Fmoc, then deprotect all the remaining Alloc group with palladium. The cleavable linker can then be assembled on the entire bead interior and N-terminally protected with Alloc. Library synthesis can then proceed as described in this report. Since the library compound on the bead surface does not have a cleavable linker, it will never be released and will not affect the MS analysis.
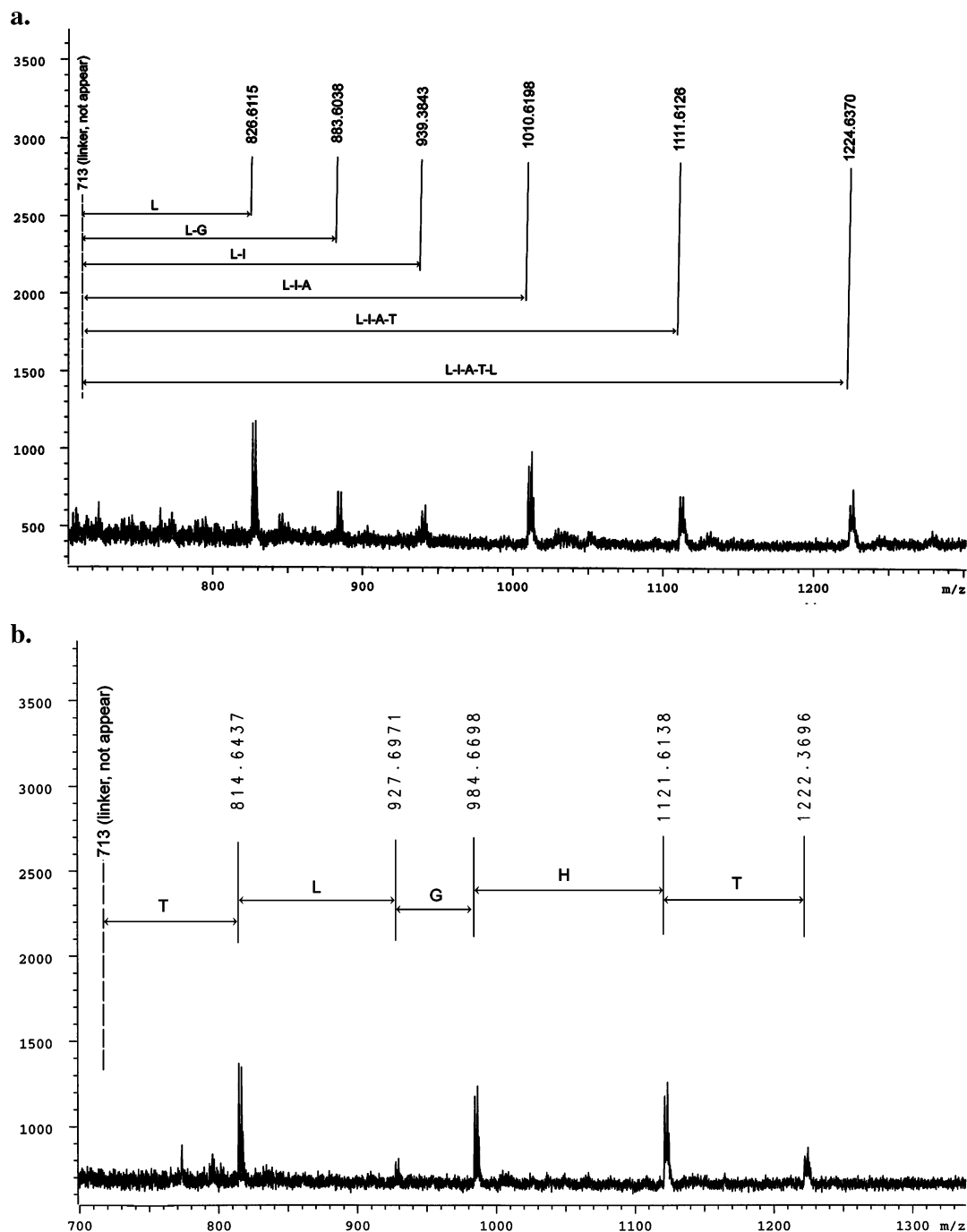
a.

b.

Figure 7. Typical mass spectra of releasate obtained from a single pentapeptide library bead prepared by the new ladder-synthesis method. (a) The full-length compound sequence was determined as L-I-A-T-L according to the decoding strategy (coding tag present); (b) the compound sequence was identified as T-L-G-H-T (coding tag absent). The mass of the protonated cleavage linker (M_0) is 713 (it does not appear in the mass spectra).

Synthesis and Applications of Beads with Multiple Concentric Layers. In 2002, Farrer et al. reported the use of a limiting amount of Fmoc-chloride and other protecting agents to sequentially generate TentaGel bead with multilayer functionalities.[17] The method is based on the fact that the reaction rate of amino protection is significantly faster than the diffusion rate of the protecting agent inside the bead. However, as indicated by the authors in the article, as the synthesis of each layer progresses, the production of each successive layer becomes more difficult. About the same time, we introduced the biphasic approach[8] to prepare multilayer beads and applied it to encode OBOC small-

molecule combinatorial libraries. Since TentaGel bead is fully compatible with water and organic solvents, a biphasic system can be set up so that the diffusion of the protecting reagent through the bead becomes even slower; as a result, the thickness of each layer can be better controlled. However, to use the above differential protecting methods to generate a large number of concentric layers on beads is problematic. The PAD approach described in this report, however, is much more robust, versatile, and easy to control. Unlike the previous methods described above, the PAD approach is based on controlling the deprotection reaction rather than the protection reaction. As shown in Figure 3, the depro-

tection reaction is slow when DCM is used as the solvent for palladium. Therefore, the extent of deprotection, or thickness of each layer, can be easily controlled with time of exposure to palladium. In principle, beads with well-defined multi concentric layers of different fluorescent dyes (e.g., up to 10 layers) may be prepared with precision and used for optical encoding. The number of permutations generated by such optical encoding system can be enormous, and it can greatly surpass the 100−1000 different dye combinations afforded by the Luminex technology.[18] These multilayer optically encoded beads can potentially be applied to the development of multiplex immunodiagnostics. The challenge would be to develop a detector (similar to a two-photon fluorescent or confocal microscope) that can rapidly decode beads with multi concentric fluorescent layers.

Conclusions

In this paper, we have developed a novel partial Alloc-deprotection (PAD) approach that can be used to generate topologically segregated bilayer beads and partially deprotect Alloc beads, layer by layer, starting from the outer layer toward the bead interior. The degree of deprotection and, therefore, the thickness of each layer can be easily controlled by adjusting the exposure time of the beads to palladium. This approach is simple and robust. It does not require any specialized equipment. Since this approach allows repetitive Alloc deprotection in multistep solid-phase synthesis, it enables us to develop a new ladder-synthesis method for the generation of OBOC combinatorial libraries in which the full-length testing library compounds are displayed on the outer layer of beads and the truncated ladder members reside in the bead interior. Since the ladders are generated during the library synthesis, this method is not limited to Edman degradative libraries. The library sequences (from N terminus) can be rapidly identified by calculating the mass differences between adjacent peaks from low mass to high mass in the mass spectra. The introduction of bromine in the cleavable linker simplifies the interpretation of the mass peaks of ladder members due to the generation of a characteristic doublet. We have further developed and incorporate a new encoding strategy into the new ladder-synthesis method for OBOC combinatorial libraries containing isobaric residues. The advantages of the encoding strategy are that it is simple and the coding tags can be topologically confined together with the truncated ladder members, thereby maintaining the presence of only full-length library compounds on the bead surface. Compared to the conventional ladder-synthesis method, the present method has three additional advantages: (i) It avoids the problems that result from the differential coupling rates of two different building blocks because only one building block is used in each coupling step. (ii) Only a full-length library compound is displayed on the bead surface. As a result, the potential interaction of the ladder tags with screening probes can be eliminated. (iii) A reverse ladder is generated so that the library beads can be sequenced not only by MS but also by the automatic microsequencer when the library is Edman degradative. MALDI-TOF MS analysis is extremely fast and, in principle, could be automated. The new ladder-synthesis method is

efficient and reliable. In addition, the PAD approach can be expanded for diverse encoding applications because the Alloc-protected amino groups reserved in the bead interior can be stepwise liberated in a desired percentage at any point of synthesis while the outer layer is assembled with library compounds. In addition to using it to encode OBOC libraries, the PAD approach could also lead to the development of novel optically encoded beads for multiplex immunodiagnostics or even information recording.[19]

Supporting Information Available. Experimental section and software for decoding encoded 5-mer OBOC combinatorial libraries. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84. (b) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res*. **1991**, *37*, 487−493. (c) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84−86. (d) Lam, K. S.; Krchnak, V.; Lebl, M. *Chem. Rev*. **1997**, *97*, 411−448.

(2) Wang, X.; Zhang, J.; Song, A.; Lebrilla, C. B.; Lam, K. S. *J. Am. Chem. Soc*. **2004**, *126*, 5740−5749.

(3) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. H. *Science* **1993**, *262*, 89−92. (b) Wang, P.; Arabaci, G.; Pei, D. *J. Comb. Chem*. **2001**, *3*, 251−254. (c) Sweeney, M. C.; Pei, D. *J. Comb. Chem*. **2003**, *5*, 218−222.

(4) Davies, M.; Bradley, M. *Tetrahedron Lett*. **1997**, *38*, 8565−8568.

(5) Sepetov, N.; Issakova, O.; Krchnak, V.; Lebl, M. U.S. Patent 5470753, 1995. (b) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc*. **1995**, *117*, 3900−3906.

(6) St. Hilaire, P. M.; Lowary, T. L.; Meldal, M.; Bock, K. *J. Am. Chem. Soc*. **1998**, *120*, 13312−13320.

(7) Lebl, M.; Lam, K. S.; Salmon, S. E.; Krchnak, V.; Sepetov, N.; Kocis, P. U.S. Patent 5840485, 1998. (b) 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.

(8) Liu, R.; Marik, J.; Lam, K. S. *J. Am. Chem. Soc*. **2002**, *124*, 7678−7680.

(9) Song, A.; Zhang, J.; Lebrilla, C. B.; Lam, K. S. *J. Am. Chem. Soc*. **2003**, *125*, 6180−6188.

(10) Hollemeyer, K.; Heinzle, E.; Tholey, A. *Proteomics* **2002**, *2*, 1524−1531.

(11) Thieriet, N.; Alsina, J.; Giralt, E.; Guibe, F.; Albericio, F. *Tetrahedron Lett*. **1997**, *38*, 7275−7278. (b) Orain, D.; Ellard, J.; Bradley, M. *J. Comb. Chem*. **2002**, *4*, 1−16.

(12) Fields, G. B.; Noble, R. L. *Int. J. Pept. Protein Res*. **1990**, *35*, 161−214.

(13) Song, A.; Wang, X.; Zhang, J.; Mařík, J.; Lebrilla, C. B.; Lam, K. S. *Bioorg. Med. Chem. Lett*. **2004**, *14*, 161−165.

(14) Franz, A. H.; Liu, R.; Song, A.; Lam, K. S.; Lebrilla, C. B. *J. Comb. Chem*. **2003**, *5*, 125−137.

(15) Duewel, H. S.; Honek, J. F. *J. Protein Chem*. **1998**, *17*, 337−350.

(16) Vojkovsky, T. *Pept. Res.* **1995**, *8*, 236−237. (b) Mařík, J.; Song, A.; Lam, K. S. *Tetrahedron Lett.* **2003**, *44*, 4319−4320.

(17) Farrer, R. A.; Copeland, G. T.; Previte, M. J. R.; Okamoto, M. M.; Miller, S. J.; Fourkas, J. T. *J. Am. Chem. Soc*. **2002**, *124*, 1994−2003.

(18) Fulton, R. J.; McDade, R. L.; Smith, P. L.; Kienker, L. J.; Kettman, J. R. *Clin. Chem.* **1997**, *43*, 1749−1756.

(19) Pham, H. H.; Gourevich, I.; Oh, J. K.; Jonkman, J. E. N.; Kumacheva, E. *Adv. Mater*. **2004**, *16*, 516−520.