ACS Combinatorial Science

Technology Note

pubs.acs.org/acscombsci

Introducing an Asp-Pro Linker in the Synthesis of Random One-Bead-One-Compound Hexapeptide Libraries Compatible with ESI-MS Analysis

Yordanka Masforrol,[†] Jeovanis Gil,[‡] Luis Javier González,[‡] Yasset Pérez-Riverol,[§] Jorge Fernández-de-Cossío,[∥] Aniel Sánchez,[‡] Lázaro Hiram Betancourt,[‡] Hilda Elisa Garay,[†] Ania Cabrales,[†] Fernando Albericio,[⊥,#,▽] Hongqian Yang,[▼] Roman A. Zubarev,[▼] Vladimir Besada,[‡] and Osvaldo Reyes Acosta*[,†]

[†]Departments of Chemical Synthesis, [‡]Proteomics, and [§]Bioinformatics, Division of Physical-Chemistry and [∥]Division of Informatics, Center for Genetic Engineering and Biotechnology, PO Box 6162, Havana, Cuba

[⊥]Institute for Research in Biomedicine, Barcelona Science Park, Baldiri Reixac 10, 08028, Barcelona, Spain

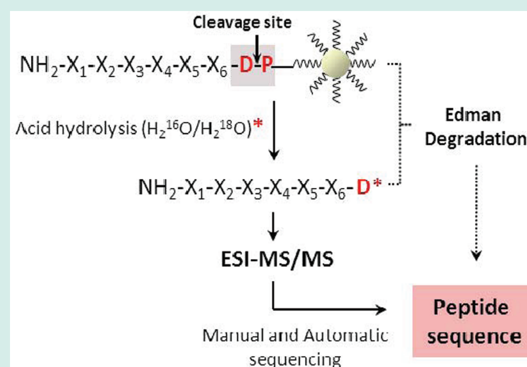
[#]CIBER-BBN, Networking Centre on Bioengineering, Biomaterials and Nanomedicine, Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain

[▽]Department of Organic Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028 Barcelona, Spain

[▼]Division of Molecular Biometry, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17 177 Stockholm, Sweden

Supporting Information

ABSTRACT: A random hexapeptide library (one-bead-one-compound), containing sixteen amino acids (16^6 different sequences) was synthesized on a Tentagel resin previously modified with a dipeptide linker (Asp-Pro). This peptide bond is highly susceptible to cleavage under mild acidic conditions in a salt-free solution prepared with $H_2{}^{16}O/H_2{}^{18}O$ (60/40% v/v). In the hydrolysis, hexapeptides are released with an additional Asp residue partially labeled with ^{18}O at the C-terminus. These conditions are fully compatible with ESI-MS analysis and facilitate sequencing by MS, as N- and C-terminal ions can be easily differentiated in MS/MS spectra. The peptides were sequenced manually and also with *de novo* sequencing programs, and identifying them in a database containing all possible heptapeptide sequences or in a filtered database. The proposed strategy is also compatible with stepwise Edman degradation using either intact beads or the released free peptides.



KEYWORDS: *peptide library, one-bead-one-compound, tentagel, mass spectrometry, electrospray ionization, de novo sequencing*

Combinatorial chemistry has a direct impact on drug discovery, allowing the synthesis and screening of a large number of compounds with potential therapeutic or diagnostic uses. Phage display peptide libraries, together with one-bead-one-compound (OBOC) peptide libraries,[1] have become important tools for understanding the biological basis of molecular recognition. OBOC libraries, in particular, allow for further optimization of lead compounds by simply including non-natural, or D-amino acids and other organic building blocks, thus accelerating the drug discovery process.[2−5]

After library screening, the structural elucidation of the selected compounds is a critical step. Edman sequencing has been widely used for this purpose because the beads can be directly analyzed without cleaving off and retrieving the peptide. However, this procedure is time-consuming, expensive, requires a free N-terminus on the peptide and is not compatible with modified amino acids and other organic blocks commonly used during lead optimization.

Mass spectrometry has been successfully used for sequencing peptides isolated from OBOC libraries by the combination of "ladder-synthesis",[6] "ladder-sequencing",[7] and MALDI-MS analysis. The major disadvantage of ladder-synthesis is that peptide ladders and the full length sequence coexist on the bead surface, which might interfere with biological screening.[8] To overcome this disadvantage, Son et al.[9] reported the application of "ladder synthesis" and bilayer bead concepts to segregate full-length peptides and their truncated variants. The outer layer carries the full-length library compound synthesized at low

Received: July 4, 2011
Revised: January 25, 2012
Published: January 25, 2012

ACS Publications © 2012 American Chemical Society

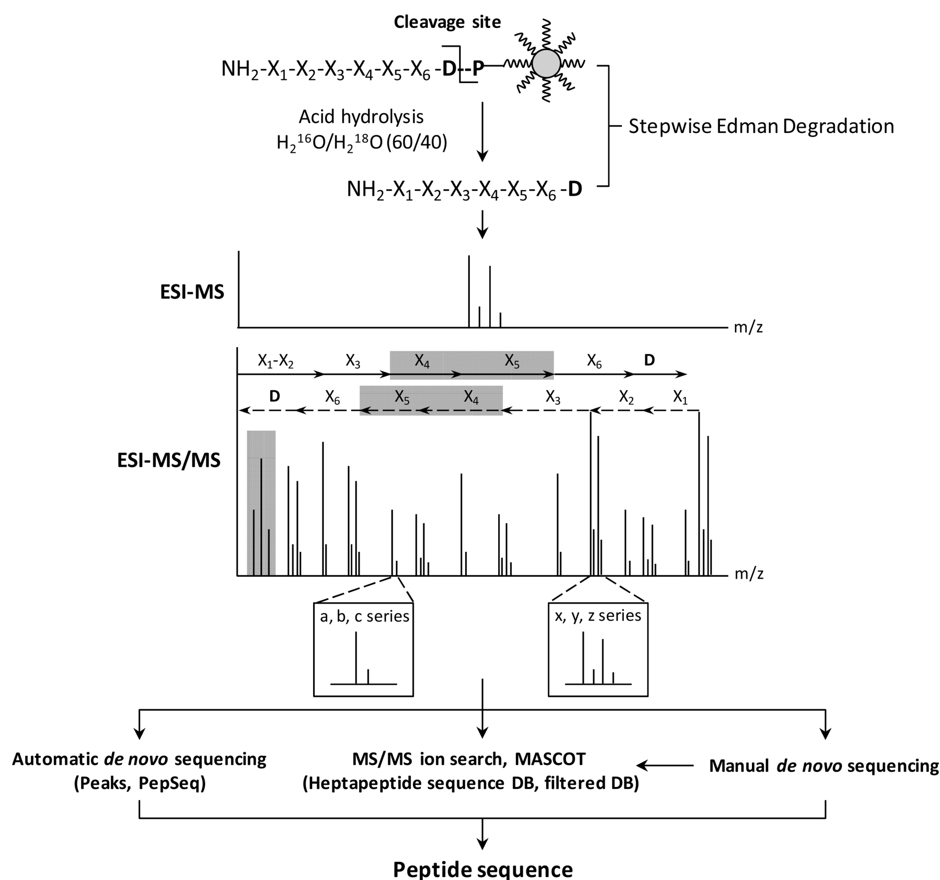


Figure 1. General strategy proposed for identifying biologically active peptides isolated from the hexapeptide library containing an Asp-Pro linker. Shadowed areas in the ESI-MS/MS spectra represent the immonium ions and a sequence tag used to filter sequence database.

concentration, tuned for the selection of high affinity interactions during biological screening. The inner layer is fully substituted with truncated ladder members to ensure that the bead can be successfully decoded by MALDI-MS.[10] In this library the peptides of interest are excised from the library by CNBr treatment and carry an alanine bromoderivative to facilitate their sequencing. Other authors[11] have used similar approaches, but introducing a bromide atom at the aldehyde group formed at the C-terminal end during the hydrolysis reaction.

Lorthioir et al.[12] have reported an analytical construct for a OBOC library consisting of two orthogonal linkers that allow releasing the peptide either in the free form or attached to a MS sensitizer with a linker having a mixture of H2/D2 to facilitate its sequencing by MS/MS. Each cleavage uses a different treatment: 90% TFA during 4 h for the first, and consecutive treatments with oxone, methanol, and 1-methylpiperazine, (overnight) for the second. In both cases peptides are recovered in DMSO; therefore, a drying step is required.

To our knowledge, partial ^{18}O-labeling,[13] has never been used in peptide libraries to facilitate the identification of hits and their sequencing by MS. Peptide sequencing by MS/MS is fundamental for identifying hits during the biological screening of peptides libraries. However, the interpretation of certain mass spectra can be difficult even for specialists, so its automation would require optimization. Furthermore, collision-induced dissociation is not efficient enough to fragment all peptides present in a library and provide full sequence coverage. ^{18}O-Labeling can be useful in these types of situations as well.

ESI-MS analysis has been also used to analyze peptides isolated from libraries, although to a much lesser extent than MALDI-MS. ESI-MS is more sensitive to the presence of salts, detergents, and other additives, imposing stricter demands on sample processing. Hoffmann and Frank[14] proposed a linker based on the well-known glycolic acid anchor, with structural modifications allowing the recovery of peptides by treatment with phosphate buffers. However, full compatibility with ESI-MS analysis requires a desalting step. The authors argued that the use of a volatile buffer (triethylammonium acetate, pH 7.3) would circumvent this problem; however, the rate of hydrolysis is much slower, and a drying step would be required anyway.

Photocleavable linkers can overcome this limitation because peptides are excised from the beads by irradiation with direct UV light, and in a second step, they are recovered in an ESI-MS-compatible solution.[15,16] Unfortunately, these linkers are expensive, and storage and handling of such libraries require special care.

As a convenient solution to these problems, we explored the use of the Asp-Pro dipeptide since it contains the peptide bond most susceptible to cleavage by controlled acid hydrolysis.[17] To our knowledge, this dipeptide has never been used in peptide libraries as a linker. We synthesized an OBOC hexapeptide library on a Tentagel S NH_2 resin previously modified by a dipeptide Asp-Pro linker allowing the release of heptapeptides, (hexapeptides + Asp at their C-terminus) for analysis by either ESI-MS and/or Edman sequencing (Figure 1). This library contains 16 amino acids at each position. To ensure a longer stability of the synthesized library and to allow its usage in several screening experiments, amino acids that can undergo

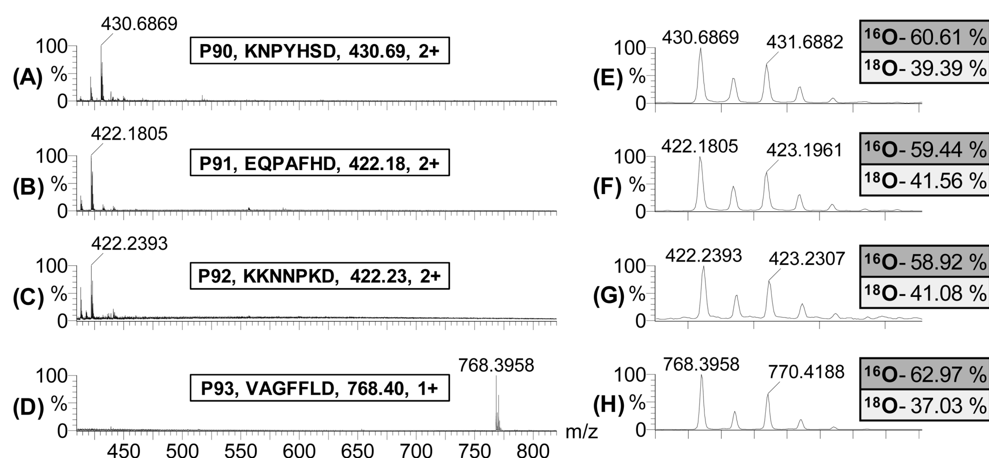


Figure 2. ESI-MS spectra shown in (A−D) correspond to model peptides (P90−P93) released from individuals beads. Peptide code, sequence, expected molecular mass, and charge are boxed in each ESI-MS spectrum. The expanded regions (E−H) show the isotopic ion distribution of individual peptides.

Table 1. Summary of the Identification of Four Model Peptides and Two Peptides Randomly Selected from the Library by Using *De Novo* Sequencing and a Database Searching

					rank[b]		
code	sequence	*m/z* (obs/calcd)	*z*	error (ppm)	*PepSeq*	*Peaks*	*MASCOT*[c]
P90	KNPYHSD[a]	430.6869/430.6985	2	26.9	1	1	1/(1)
P91	EQPAFHD[a]	422.1805/422.1852	2	11.1	1	2 (QE**PAFHD**)[e]	2/(1)
P92	KKNNPKD[a]	422.2393/422.2378	2	3.6	1	1	2/(1)
P93	VAGFFLD[a]	768.3958/768.4000	1	5.5	>10	1	1/(1)
R01	APQTKYD[d]	411.7152/411.7032	2	29.1	1	1	1/(1)
R02	FPQTRKD[d]	297.8286/297.8276	3	3.3	-	>10 (1)[f]	1/(1)

[a]Sequences correspond to the synthesized model peptides attached to Tentagel S NH_2 resin. [b]Ranking of the correct sequence in the program output. [c]The values within parentheses correspond to the ranks of the correct sequence after performing the search in a subdatabase. [d]Randomly selected beads. The sequences were obtained by manual sequencing. [e]Top-ranked sequence. [f]The value in parentheses indicates the rank of the correct sequence after performing a hybrid search (*de novo* sequencing + database search).

oxidation or that promote aggregation (Trp, Cys, and Met) were excluded from the synthesis. Asp was also excluded from the combinatorial approach to eliminate the possibility that some sequences selected at random might contain internal Asp-Pro bonds that would be destroyed during hydrolysis. In principle, this peptide library could contain about 16.8 million different sequences. It should be noted, however, that if Asp is included the synthesized library gains 6.9 millions new sequences, losing only 417 605 sequences to the hydrolysis of internal Asp-Pro bonds (only 1.73% of the total). After biological screening, positive beads can be isolated, placed in an Eppendorff tube and treated under mild acidic conditions in 10 μL of 1% formic acid solution, 50% acetonitrile, at 100 °C for 25 min for peptide release and analysis.

This salt-free hydrolysis solution was selected to allow the release of heptapeptides attached to a single bead in sufficient amounts for a sensitive analysis by ESI-MS without further processing. This is particularly advantageous for an efficient recovery of either hydrophilic or hydrophobic peptides present in the wide diversity of the library. Several concentrations of formic acid (0.1%, 1%, 2%, and 5%) were also evaluated in the hydrolysis solution, but no significant improvement in the relative recovery of P90 model peptide was observed from 1% to 5% (see Supporting Information). If the hydrolysis solution is prepared in 40% (v/v) $H_2{}^{18}O$, the released heptapeptides will be partially labeled with ^{18}O at their C-terminal end. The resulting isotopic ion distribution facilitates their identification among the background signals that might be generated during hydrolysis. Additionally, partial ^{18}O-labeling makes manual sequencing of the target peptides easier and more reliable[13] since N- and C-terminal ions are unambiguously differentiated in their ESI-MS/MS spectra (Figure 1).

In our strategy the identification of hits relies on the integration of several complementary approaches. In addition to manual interpretation, the ESI-MS/MS spectra are analyzed with the aid of *de novo* sequencing software (*PepSeq* and *Peaks*[18]). *MASCOT*[19] software commonly used in proteomics for protein identification can be also used with peptide libraries, identifying the released peptides in a database with all possible sequences or in a filtered database restricted to sequences complying with specific criteria (molecular mass and accuracy, qualitative amino acid composition, or a sequence tag). The most-likely sequences are proposed after analyzing the output of all of these approaches. Beside ESI-MS analysis, the peptides are also amenable to sequencing by Edman degradation because they are released as free N-terminal species and only 1/3 of the total amount is consumed by MS analysis. If MS is not available, the intact bead can be also loaded inside the N-terminal sequencer, and the sequence can be readily obtained.

Four model hexapeptides: KNPYHS (P90), EQPAFH (P91), KKNNPK (P92), and VAGFFL (P93) were synthesized to evaluate if the chosen hydrolysis conditions permit a sufficiently sensitive MS analysis and to explore the occurrence of some potential side reactions. The resultant ESI-MS spectra (Figure

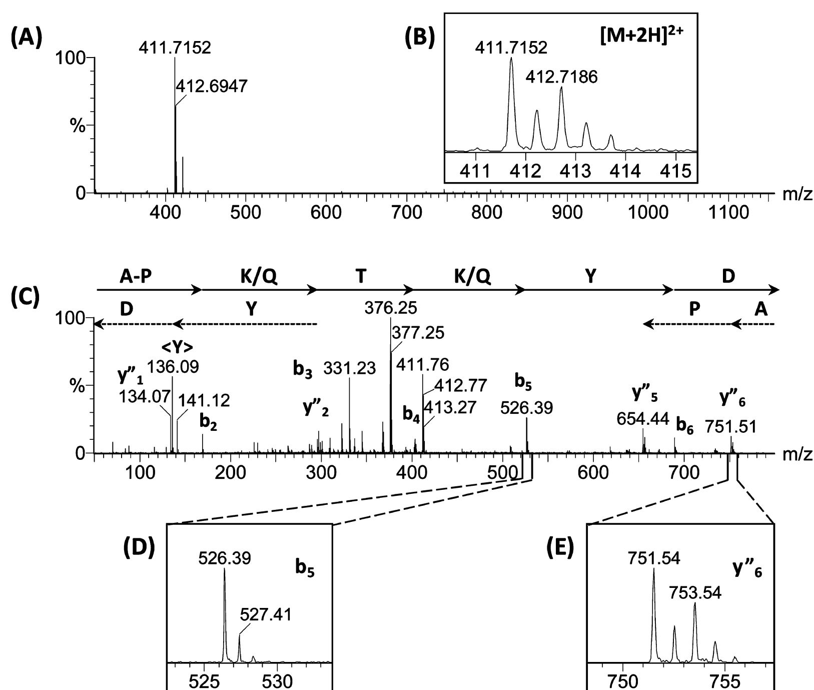


Figure 3. (A) ESI-MS spectrum of peptide AP-(K/Q)-T-(K/Q)-YD released from a bead (R01) randomly selected from the library. The expanded region in (B) shows its isotopic ion distribution. (C) ESI-MS/MS spectrum of this peptide partially labeled with ^{18}O at its C-terminus. Continuous and dashed lines indicate the sequences deduced from the N- and C-terminal ions, respectively. The expanded regions in D and E show the isotopic ion distributions of the N- and C-terminal ions, respectively.[13]

2A–D) showed intense signals at m/z that agreed with the expected values with a mass error lower than 50 ppm (Table 1).

No evidence of deamidation, peptide degradation or pyroglutamic acid formation were detected (Figure 2A–D). Analysis of the isotopic ion distribution of model peptides (Figure 2E–H) revealed that approximately 40% of ^{18}O was incorporated in all cases, confirming that additional ^{18}O was not incorporated by acid-catalyzed exchange. The fact that all isotopic labeling was located at the C-terminus makes peptide sequencing easier and more reliable.[13] Stable isotope labeling has been chemically encoded in the synthesis of OBOC libraries[10–12] to release the peptides as partially labeled species for MS analysis. In our case, the user can choose to employ partial ^{18}O-labeling or not, since only 2 μL of $H_2{}^{18}O$ is needed in the hydrolysis solution (10 μL) required for a single bead. While we typically analyzed the released peptides by MS just after hydrolysis, freezing partially ^{18}O-labeled peptides at −20 °C for a week in the hydrolysis solution did not incorporate ^{18}O beyond 40%. Table 1 summarizes the identification of model peptides as well as peptides isolated from beads randomly selected from the library using the strategy described in Figure 1.

Best results during *de novo* sequencing of model peptides were obtained by using *Peaks* software. However, none of these applications were able to take full advantage of partial ^{18}O-labeling, as a skilled practitioner would do in analyzing individual spectra. Moreover, they cannot be customized to restrict their output to 7 amino acid-long sequences, favoring instead longer sequences where heavier amino acids such as Trp, Lys/Gln or Arg have been replaced by pairs of lighter amino acids. This later limitation, however, can be overcome in the case of *Peaks,* since it is possible to perform a hybrid search by combining *de novo* sequencing and database search approaches (see Table 1). For model peptides (P90–P93, Table 1), *MASCOT* always ranked the correct sequence among the top two. However, when the database was filtered taking into account additional information obtained from manual sequencing (immonium ions and a sequence tag) the correct sequence was always ranked at the top. To our knowledge this is the first time that MASCOT has been applied successfully to decode peptide libraries. Therefore, the results (manual, automatic sequencing and database search) should always be analyzed as a whole before selecting the correct sequence among a number of possibilities.

Two randomly selected library beads (R01 and R02, Table 1) were also analyzed by hydrolysis and MS sequencing. The peptide contained in bead R01 was fragmented by conventional CID in a QTOF-1, and the resulting ESI-MS spectrum (Figure 3A) showed an intense doubly charged signal at m/z 411.71 with the typical isotopic ion distribution of a partially ^{18}O-labeled peptide (Figure 3B). The ESI-MS/MS spectrum (Figure 3C) clearly showed different isotopic distributions[13] for the N- and C-terminal ions (Figure 3D and 3E) leading us to propose the sequence: AP-(K/Q)-T-(K/Q)-YD. On the basis of the mass accuracy of the consecutive b_n ions, Q and K could be tentatively assigned at the positions third and fifth of the sequence. *De novo* sequencing and identification of this peptide in a sequence database is shown in Table 1.

It was impossible to obtain a full sequence for the peptide in bead R02 when fragmented by HCD in an Orbitrap MS, due to gaps in the resulting ESI-MS/MS spectra. Only when fragmented by ETD (see Supporting Information) was it possible to sequence the peptide in its entirety. Since the proposed method is compatible with ESI-MS analysis, the use of state of the art MS (such as Orbitrap or FT-MS) is also advisible because their highest accuracy (<5 ppm) allows discarding sequences with very close molecular masses that otherwise would be a source of false identifications, also

permitting the unambiguous differentiation of isobaric amino acids (Lys/Gln). The end result is a reduction in the number of peptides to be synthesized in validation stages.

Taking into account that the cleavage sequence (Asp-Pro) is linked directly to the hexapeptide library, in principle it should also be synthesized during the validation stages of positive hits. If it turned out not to be relevant for biological activity it can be ignored but only after careful experimental validation. If libraries based on Asp-Pro cleavage do not contain unstable amino acids, they are compatible with on-bead assays over long periods of time and can be reused for multiple screening experiments. Designing and synthesizing Asp-Pro linkage libraries is simple and affordable in any laboratory dedicated to peptide synthesis or combinatorial chemistry.

■ ASSOCIATED CONTENT

Supporting Information

Experimental conditions used to release the peptide from a single bead and the optimization of some parameters (acid concentration, hydrolysis time, temperature, acetonitrile concentration), sequencing of hexapeptides released from a single bead sequenced by electro transfer dissociation in an orbitrap and illustrates the advantage over the conventional Collision induced dissociation and all the materials and procedures used in this manuscript. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Address: Chemical Synthesis Department, Center for Genetic Engineering and Biotechnology, PO Box 6162, Havana, Cuba. E-mail: osvaldo.reyes@cigb.edu.cu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank to Alejandro Martín for the critical reading and English correction of the manuscript. The authors would like to thank to Dr.Toshifumi Takao (Osaka University) for developing the partial ^{18}O-labeling to make easier the sequencing of peptides by mass spectrometry.

■ ABBREVIATIONS

ACN, acetonitrile; CID, collision induced dissociation; ECD, electron capture dissociation; ESI, electrospray ionization; ETD, electron transfer dissociation; FA, formic acid; FTMS, Fourier transform mass spectrometer; HCD, higher energy collision dissociation; MS, mass spectrometry; OBOC, one-bead-one-compound

■ REFERENCES

(1) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **1991**, *354*, 82−84.

(2) Hoesl, C. E.; Nefzi, A.; Ostresh, J. M.; Yu, Y.; Houghten, R. A. Mixture-based combinatorial libraries: from peptides and peptidomimetics to small molecule acyclic and heterocyclic compounds. *Methods Enzymol.* **2003**, *369*, 496−517.

(3) Liu, R.; Marik, J.; Lam, K. S. A novel peptide-based encoding system for "one-bead one-compound" peptidomimetic and small molecule combinatorial libraries. *J. Am. Chem. Soc.* **2002**, *124*, 7678−7680.

(4) Yao, N.; Xiao, W.; Wang, X.; Marik, J.; Hyoung, S. P.; Takada, Y.; Lam, K. S. Discovery of targeting ligands for breast cancer cells using the one-bead one-compound combinatorial method. *J. Med. Chem.* **2009**, *52*, 126−133.

(5) Peng, L.; Liu, R.; Marik, J.; Wang, X.; Takada, Y.; Lam, K. S. Combinatorial chemistry identifies high-affinity peptidomimetics against alpha4beta1 integrin for in vivo tumor imaging. *Nat. Chem. Biol.* **2006**, *2*, 381−389.

(6) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keoug, T. Generation and screening of combinatorial peptide libraries designed for rapid sequencing by mass spectrometry. *J. Am. Chem. Soc.* **1995**, *117*, 3900−3906.

(7) Chait, B. T.; Wang, R.; Beavis, R. C.; Kent, S. B. Protein ladder sequencing. *Science* **1993**, *262*, 89−92.

(8) Aina, O. H.; Liu, R.; Sutcliffe, J. L.; Marik, J.; Pan, C. X.; Lam, K. S. From combinatorial chemistry to cancer-targeting peptides. *Mol. Pharmaceutics* **2007**, *4*, 631−651.

(9) Song, A.; Zhang, J.; Lebrilla, C. B.; Lam, K. S. A novel and rapid encoding method based on mass spectrometry for one-bead-one-compound small molecule combinatorial libraries. *J. Am. Chem. Soc.* **2003**, *125*, 6180−6188.

(10) Wang, X.; Peng, L.; Liu, R.; Xu, B.; Lam, K. S. Applications of topologically segregated bilayer beads in "one-bead one-compound" combinatorial libraries. *J. Pept. Res.* **2005**, *65*, 130−138.

(11) Paulick, M. G.; Hart, K. M.; Brinner, K. M.; Tjandra, M.; Charych, D. H.; Zuckermann, R. N. Cleavable hydrophilic linker for one-bead-one-compound sequencing of oligomer libraries by tandem mass spectrometry. *J. Comb. Chem.* **2006**, *8*, 417−426.

(12) Lorthioir, O; Carr, R. A. E.; Congreve, M. S.; Geysen, M. H.; Kay, C.; Marshall, P.; McKeown, S. C.; Parr, N. J.; Scicinski, J. J.; Watson., S. P. Single bead characterization using analytical constructs: Application to quality control of libraries. *Anal. Chem.* **2001**, *73*, 963−970.

(13) Takao, T.; Hori, H.; Okamoto, K.; Harada, A.; Kamachi, M.; Shimonishi, Y. Facile assignment of sequence ions of a peptide labeled with ^{18}O at the carboxyl terminus. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 312−315.

(14) Hoffmann, S.; Frank, R. A new safety-catch peptide-resin linkage for the direct release of peptides into aqueous buffers. *Tetrahedron Lett.* **1994**, *35*, 7763−7766.

(15) Brown, B. B.; Wagner, D. S.; Geysen, H. M. A single-bead decode strategy using electrospray ionization mass spectrometry and a new photolabile linker: 3-Amino-3-(2-nitrophenyl)propionic acid. *Mol. Diversity* **1995**, *1*, 4−12.

(16) Madder, A.; Li, L.; De Muynck, H.; Farcy, N.; Van Haver, D.; Fant, F.; Vanhoenacker, G.; Sandra, P.; Davis, A. P.; De Clercq, P. J. Evaluation of a two-stage screening procedure in the combinatorial search for serine protease-like activity. *J. Comb. Chem.* **2002**, *4*, 552−562.

(17) Piszkiewicz, D.; Landon, M.; Smith, E. L. Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1173−1178.

(18) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337−2342.

(19) Perkins, D. N.; Pappin, D. J.; Creasy, D. N.; Cotrell, J. S. Probability-based protein identification by searching sequence using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551−3567.